# Topics in the Lexical Semantics–Morphosyntax Interface

Louise McNally

Universitat Pompeu Fabra Barcelona

Université Paris Cité, June 2024

Interlude: Context and “extra” meaning, and why we shouldn’t abstract away from them anymore

# Outline for Part 6

- Vector-based representations as a tool for getting a better grip on lexical meaning in composition.
- How to integrate them or at least make allowances for them starting from a formal syntactic/semantic framework.

# Preliminary comments

- Formal semantics:
  - Grounded in reference
  - Discrete, symbolic
  - Restricted idea of what linguistic meaning is (vs. “world knowledge”) that matches what the approach does well.
- Has left the lexicon largely to cognitive linguists, psychologists, computational semanticists.

# Preliminary comments

- Parts 1-5 have been an argument that a closer look at the lexicon is important for understanding the morphosyntax-semantics interface.
- Classic tools of formal semantics are of limited help.
- Way forward: Bring in additional tools
  - Does **not** mean rejecting what works!
  - New tools can give us new ways to think about some aspects of language $\rightarrow$ new opportunities for new research.

# Vector-based representations

- **Idea:** You can learn a lot about a word's meaning by looking at the other words it tends to (or tends not to) occur with.
- Main origin in linguistics: "Distributional Hypothesis" (Harris 1954; Firth 1957)
- Influential in lexicography (Hanks 2013)

# Vector-based representations

- **Idea:** You can learn a lot about a word's meaning by looking at the other words it tends to (or tends not to) occur with.
- Main origin in linguistics: "Distributional Hypothesis" (Harris 1954; Firth 1957)
- Influential in lexicography (Hanks 2013)
- Can be conceived of as (partial) models of concepts, as some kind of memory, or as classifiers.
- Long history in psychology ("Latent semantic analysis", Landauer and Dumais 1997; see also Gärdenfors' 2000 Conceptual Spaces).

# Vector-based representations

- **Idea:** You can learn a lot about a word's meaning by looking at the other words it tends to (or tends not to) occur with.
- Main origin in linguistics: "Distributional Hypothesis" (Harris 1954; Firth 1957)
- Influential in lexicography (Hanks 2013)
- Can be conceived of as (partial) models of concepts, as some kind of memory, or as classifiers.
- Long history in psychology ("Latent semantic analysis", Landauer and Dumais 1997; see also Gärdenfors' 2000 Conceptual Spaces).
- Wildly successful in natural language processing/AI – basis for (predictive or *generative*) Large Language Models like ChatGPT.

# Vector-based representations: short history

- 1970s-2000s: Make a **distributional** vector (ordered list) of numbers of occurrences of words with all other (relevant) words in some linguistic context (see Lenci 2018 for review).

# Vector-based representations: short history

- 1970s-2000s: Make a **distributional** vector (ordered list) of numbers of occurrences of words with all other (relevant) words in some linguistic context (see Lenci 2018 for review).
- Early 2010s: **(Word) embeddings**, more abstract vectors that allow a neural network to predict a word given a context or a context given a word (Mikolov *et al.* 2013; Baroni *et al.* 2014).

# Vector-based representations: short history

- 1970s-2000s: Make a **distributional** vector (ordered list) of numbers of occurrences of words with all other (relevant) words in some linguistic context (see Lenci 2018 for review).
- Early 2010s: **(Word) embeddings**, more abstract vectors that allow a neural network to predict a word given a context or a context given a word (Mikolov *et al.* 2013; Baroni *et al.* 2014).
- Mid 2010s: Embeddings **for subwords**, allow method to apply to morphology (Bojanowski *et al.* 2017).

# Vector-based representations: short history

- 1970s-2000s: Make a **distributional** vector (ordered list) of numbers of occurrences of words with all other (relevant) words in some linguistic context (see Lenci 2018 for review).
- Early 2010s: **(Word) embeddings**, more abstract vectors that allow a neural network to predict a word given a context or a context given a word (Mikolov *et al.* 2013; Baroni *et al.* 2014).
- Mid 2010s: Embeddings **for subwords**, allow method to apply to morphology (Bojanowski *et al.* 2017).
- Late 2010s: **Contextualized** embeddings, take dynamics of local context into account (e.g., Peters *et al.* 2018; Devlin *et al.* 2019).

# Vector-based representations: short history

- 1970s-2000s: Make a **distributional** vector (ordered list) of numbers of occurrences of words with all other (relevant) words in some linguistic context (see Lenci 2018 for review).
- Early 2010s: **(Word) embeddings**, more abstract vectors that allow a neural network to predict a word given a context or a context given a word (Mikolov *et al.* 2013; Baroni *et al.* 2014).
- Mid 2010s: Embeddings **for subwords**, allow method to apply to morphology (Bojanowski *et al.* 2017).
- Late 2010s: **Contextualized** embeddings, take dynamics of local context into account (e.g., Peters *et al.* 2018; Devlin *et al.* 2019).
- Have been extended to include **visual**, other information (from at least Bruni *et al.* 2012).

# Example, simple distributional version

```
on demand by storekeepers, who cut, folded, and pasted sheets of
roll of the required width, cut off from the roll and
literature. They also had to cut back time given to electives,
unique path by taking a cut in salary and risking our
blade or something sharp to cut into one’s own skin. The
the ‘en face’ method are cut very thin to enable maximal
years ago, the Texas legislature cut its funding for schools by
describe the central themes that cut across a vast array of
scissors can allow children to cut out different pictures or shapes
reconstructing models of killing centers cut into valuable instructional
```

# Example, simple distributional version

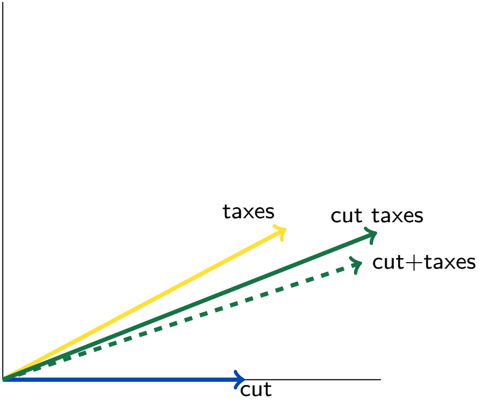
```
a lasting boost. Cutting taxes ensures better after-tax returns.
say, accompanied by higher corporate taxes. Governments might commit to
law Monday will hike gasoline taxes, add new nightly fees on
information on how to file taxes, the Internet was used for solving
younger days? Why not cut income tax and roll back the state
schools? Why should they pay taxes to educate other people's kids?
words of Oliver Wendell Holmes: taxes are the price we pay
the item's availability. Furthermore, Primo taxes our link resolver, sir
votes to voluntarily increase property taxes, a public school can not
sales, property, and state income taxes. Overall, investing in community
```

# Example, simple distributional version

```
we've cut taxes by $5 billion, the
chief architect. And -- and we cut taxes. The economy was growing and
taxes on the rich. We cut taxes on the middle class. All
We turned it around. We cut taxes by $2 billion. Property
you Republican candidates wants to cut taxes for the wealthy and hope
great president. I thought he cut taxes dramatically. # But not Hillary?
in taxes while trying to cut taxes for those in lower brackets.
in both chambers want to cut taxes by well over $4 billion
regulations. They're trying to cut taxes. They've kind of made
size of government. You have cut taxes. You have strong economic growth.
```

- Vectors for words can be composed by addition, multiplication
  $\approx$ Vectors extracted for phrases.

# Comparing vector-based representations



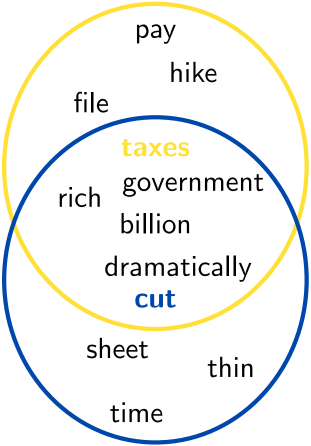

- The similarity of vectors can be compared.
- Vector similarity reflects semantic similarity of words/phrases.
- Can be used to evaluate composition operations.

# Formal vs. vector-based composition

**Formal:**
**– "Additive"**
**– Entailment-based**

$x$ **cut** $y$
$x$ causes $e$
$e$ = a property of $y$ changes to some degree
$y$ is **taxes**
$y$ is money
$y$ imposed by governing body
⋮

**Vector-based:**
**– "(Soft) Filtering"**
**– Soft inference-based**



# What vector-based semantics does well

- Directly captures "co-composition" (Pustejovsky 1995, Asher 2011).
- Does not require making hard decisions about individual word senses.
- Good model for kind/generic/stereotypical information, some aspects of metaphor (e.g., Kintsch 2000), idioms (e.g., Senaldi *et al.* 2016, Part 7).
- Promising for hard problems derivational morphology (e.g., blends like *mockumentary*) (Part 8).

# What vector-based semantics does poorly

Once upon a time there were three little pigs. One pig built a house of straw while the second pig built his house with sticks. They built their houses very quickly and then sang and danced all day because they were lazy. The third little pig worked hard all day and built his house with bricks.

A big bad wolf saw the two little pigs while they danced and played and thought, ''What juicy tender meals they will make!'' He chased the two pigs and they ran and hid in their houses. The big bad wolf went to the first house and huffed and puffed and blew the house down in minutes. The frightened little pig ran to the second pig's house that was made of sticks. The big bad wolf now came to this house and huffed and puffed and blew the house down in hardly any time. Now, the two little pigs were terrified and ran to the third pig's house that was made of bricks.

The big bad wolf tried to huff and puff and blow the house down, but he could not. He kept trying for hours but the house was very stron and the little pigs were safe inside.

(*http://shortstoriesshort.com/*)

# What vector-based semantics does poorly

- No notion that language is connected to token entities/events in the world (Bender and Koller 2020).
  - Though they *can* extract token entity/event information through linguistic markers (e.g. anaphoric relations, mention of dates, adverbials).
- No evidence of modeling interlocutors as intentional communicative agents.
- Limited ability to adapt to sudden changes of context (e.g., McNally and Boleda 2017; Eisenschlos *et al.* 2023; next slide).

# Example: difficulties with context

- Kayne (1984): "Argument saturating" adjectives generally interpreted as agents.
- Arsenijević *et al.* (2014): With appropriate context, other interpretations possible.
- McNally and Boleda (2017): Sort of context dependence not well-captured by vector-based semantics unless reference is incorporated.

(1) a. Yeltsin met the prospective Democratic presidential candidate Bill Clinton on June 18. His itinerary also included an official visit to Canada/??an official **Canadian visit**. (BNC)

b. Prince Edward and wife begin **Canadian visit** (http://metronews.ca/news/canada/365325/prince-edward-and-wife-begin-canadian-visit/

# Combining formal and vector-based semantics

- Various architectures, used now for some time (Garrette *et al.* 2011; Lewis and Steedman 2013; Asher *et al.* 2016; Emerson 2018; Lapesa *et al.* 2018, a.o.)
- Explicitly or implicitly recognize a division of labor:
  - Vector-based semantics for **descriptive content** phenomena
  - Formal semantics for **reference-related** phenomena
- Even if one does not implement vector models, one can take them into account doing linguistic analysis.
  - Assuming division of labor can reduce complexity in other parts of the one's analysis (see Part 7).

## Summary: The bigger picture

- Language is complex.
  - Why should a model of it be simple?

# Summary: The bigger picture

- Language is complex.
  - Why should a model of it be simple?
- **Models** of language $\neq$ **theories** or **hypotheses** about language.
  - Models are descriptions; theories/hypotheses offer explanations.
  - Need not be stated in the same theoretical vocabulary.

# Summary: The bigger picture

- Language is complex.
  - Why should a model of it be simple?
- **Models** of language $\neq$ **theories** or **hypotheses** about language.
  - Models are descriptions; theories/hypotheses offer explanations.
  - Need not be stated in the same theoretical vocabulary.
- Resist the inclination to think that the fewer types of tools we use to model language, the more theoretical insight we will *necessarily* gain.
  - Use just the tools needed to make as good a model as possible.

## References

Arsenijević, B., Boleda, G., Gehrke, B., and McNally, L. (2014). Ethnic adjectives are proper adjectives. In *Proceedings of the 46th Annual Meeting of the Chicago Linguistic Society* (ed. R. Baglini, A. Baker, T. Grinsell, J. Keane, and J. Thomas), Chicago, IL, pp. 17–30.

Asher, Nicholas (2011). *Lexical Meaning in Context*. Cambridge University Press, Cambridge.

Asher, Nicholas, Van de Cruys, Tim, Bride, Antoine, and Abrusán, Márta (2016, December). Integrating type theory and distributional semantics: A case study on adjective–noun compositions. *Computational Linguistics*, **42**(4), 703–725.

Baroni, Marco, Dinu, Georgiana, , and Kruszewski, Germán (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of ACL*, Baltimore, MD, pp. 238–247.

Bender, Emily M. and Koller, Alexander (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (ed. D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault), Online, pp. 5185–5198. Association for Computational Linguistics.

Bojanowski, Piotr, Grave, Edouard, Joulin, Armand, and Mikolov, Tomas (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, **5**, 135–146.

Bruni, Elia, Boleda, Gemma, Baroni, Marco, and Tran, Nam Khanh (2012, July). Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Jeju Island, Korea, pp. 136–145. Association for Computational Linguistics.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, and Toutanova, Kristina (2019, 6). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota, pp. 4171–4186. Association for Computational Linguistics.

Eisenschlos, Julian Martin, Cole, Jeremy R., Liu, Fangyu, and Cohen, William W. (2023, May). WinoDict: Probing language models for in-context word acquisition. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (ed. A. Vlachos and I. Augenstein), Dubrovnik, Croatia, pp. 94–102. Association for Computational Linguistics.

Emerson, Guy (2018). *Functional Distributional Semantics: Learning*

*Linguistically Informed Representations from a Precisely Annotated Corpus*. Ph.D. thesis, University of Cambridge.

Firth, John R. (1957). *Papers in Linguistics, 1934-1951*. Oxford University Press, Oxford.

Gärdenfors, Peter (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press, Cambridge, MA.

Garrette, Dan, Erk, Katrin, and Mooney, Raymond (2011). Integrating logical representations with probabilistic information using Markov logic. In *Proceedings of IWCS 2011*.

Hanks, Patrick W. (2013). *Lexical Analysis: Norms and Exploitations*. MIT Press, Cambridge, MA.

Harris, Zellig S. (1954). Distributional structure. *Word*, **10**(2-3), 146–162.

Kayne, Richard (1984). *Connectedness and Binary Branching*. Foris, Dordrecht.

Kintsch, Walter (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin and Review*, **7**, 257–266.

Landauer, Thomas and Dumais, Susan (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**(2), 211–240.

Lapesa, Gabriella, Kawaletz, Lea, Plag, Ingo, Andreou, Marios, Kisselew, Max, and Padó, Sebastian (2018). Disambiguation of newly derived nominalizations in context: A distributional semantics approach.

Lenci, Alessandro (2018). Distributional models of word meaning. *Annual Review of Linguistics*, **4**, 151–171.

Lewis, Mike and Steedman, Mark (2013). Combined distributional and logical semantics. *Transactions of the Association for Computational Linguistics*, **1**, 179–192.

McNally, Louise and Boleda, Gemma (2017). Conceptual vs. referential affordance in concept composition. In *Compositionality and concepts in linguistics and philosophy* (ed. Y. Winter and J. Hampton). Springer, Berlin.

Mikolov, Tomas, Yih, Wen-Tau, and Zweig, Geoffrey (2013). Linguistic regularities in continuous space word representations. In *HLT-NAACL*, pp. 746–751.

Peters, Matthew E., Neumann, Mark, Iyyer, Mohit, Gardner, Matt, Clark, Christopher, Lee, Kenton, and Zettlemoyer, Luke (2018, June). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (ed.

M. Walker, H. Ji, and A. Stent), New Orleans, Louisiana, pp. 2227–2237. Association for Computational Linguistics.

Pustejovsky, James (1995). *The Generative Lexicon*. MIT Press, Cambridge.

Senaldi, Marco S. G., Lebani, Gianluca E., and Lenci, Alessandro (2016). Lexical variability and compositionality: Investigating idiomaticity with distributional semantic models. In *Proceedings of the 12th Workshop on Multiword Expressions*. ACL.