MIT Press • Open Encyclopedia of Cognitive Science

Distributional Semantics

Gemma Boleda[1,2]

[1]Universitat Pompeu Fabra, [2]ICREA

MIT Press

Published on: Jul 01, 2026

DOI: https://doi.org/10.21428/e2759450.50749b98

Distributional semantics is an approach to meaning that takes the use of words and other linguistic units as the basis for semantic representation. It is based on the observation that words that are similar or related in meaning (such as "cat" and "dog") appear in similar linguistic contexts ("I just took my __ to the vet"). This systematic relationship enabled the development of a methodological framework to obtain useful semantic representations at scale thanks to the availability of large amounts of digitized texts (*corpora*), computing power, and learning algorithms. These algorithms take a corpus as input and produce a distributional model (also known as *semantic space* or *vector space*) as output. Distributional models treat words as *vectors*, that is, points in a high-dimensional space. Vector similarity, or closeness in the space, mirrors semantic relatedness because two expressions that appear in similar textual contexts in the corpus are assigned similar vectors in the model. Distributional semantics is influential in cognitive science and artificial intelligence and is one of the foundations of modern deep learning approaches to language.

History

There are three main ingredients in distributional semantics, which were developed along sometimes synergistic, sometimes parallel and noninteracting paths in philosophy, linguistics, psychology, and computer science. The first ingredient is the theory connecting meaning and use, whose foundations were laid in philosophy (Wittgenstein, 1953/1958) and further developed in linguistics (Firth, 1957; Harris, 1954). The second is the representational framework, the vector space, which was first proposed in psychology by Osgood et al. (1957). And the third is the method for inducing semantic representations from data, which is what gave distributional semantics its bite.

Initial distributional methods were manual and applied to linguistic analysis (Firth, 1957; Harris, 1954). Later computational and algorithmic advances, together with the availability of digital texts, enabled their full automation. This happened first in the field of information retrieval (Salton et al., 1975; Sparck Jones, 1986) and then in cognitive science (Deerwester et al., 1990) and computational linguistics (Schütze, 1992). These earlier methods extracted and processed statistics for word co-occurrences from texts; in the 2010s, more powerful machine learning–based algorithms were developed (Mikolov et al., 2013). Modern deep learning models, such as large language models (LLMs; Devlin et al., 2019), can be seen as a generalization of distributional semantics that also incorporate grammatical and reasoning capabilities [see Large Language Models]. In particular, LLMs are known for processing input hierarchically; at the bottom of the hierarchy are *token embeddings*, which are akin to word embeddings or distributional models—effectively the lexicon of the LLM.

Core concepts

Meaning as use

The view that "the meaning of a word is its use in the language" (Wittgenstein, 1953/1958, p. 43) is the central tenet in distributional semantics. Zellig Harris (1954) made it explicit in the *distributional hypothesis*, which states that similarity in meaning results in similarity of linguistic distribution (*distribution* refers to contexts of use in the language; see the "cat/dog" example above). This causal relationship enables scholars to reverse engineer the process, inducing semantic representations from contexts of use (Figure 1).

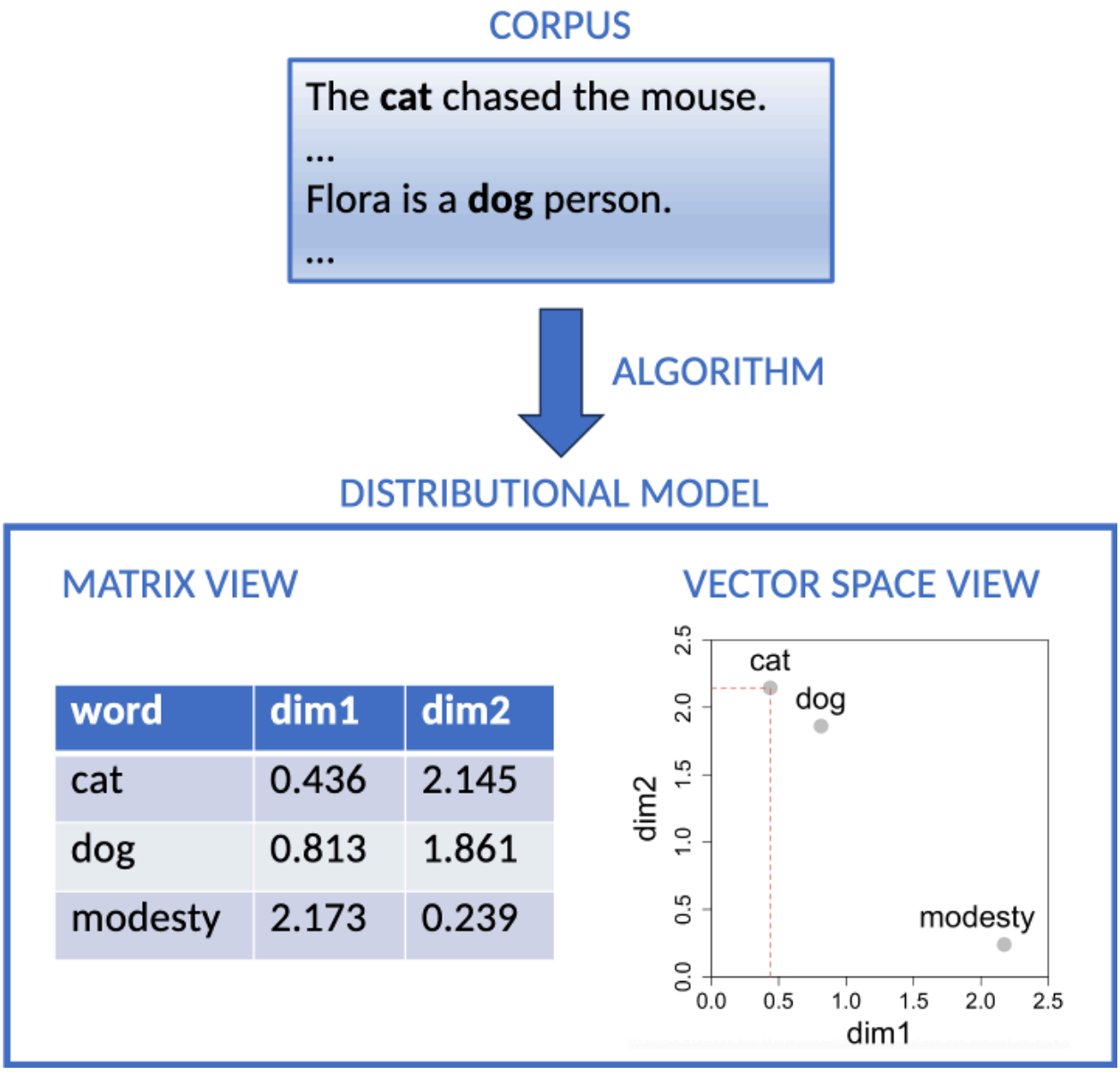

Figure 1

Distributional semantics: from corpus to semantic space via learning algorithm. In this toy model, word vectors have only two dimensions, but in real spaces, they range from dozens to thousands of dimensions.

Distributional models

Distributional models are in the form of vector spaces (Figure 1, lower panels) in which meanings are vectors, or points in the space. These models are usually represented in matrix form (Figure 1, lower panel left) where each row of the matrix is a word (or other linguistic unit) and the columns define the position of the words in the space. Distributional models can be built with different methods; what they all have in common is that words that occur in similar linguistic contexts are assigned similar vectors.

Two prominent algorithms to build distributional methods are latent semantic analysis (LSA; Landauer & Dumais, 1997) and word2vec (Mikolov et al., 2013).

Distributional semantics and cognition

Distributional semantics has been proposed as a theory for how humans learn, store, and process word meaning (Landauer & Dumais, 1997) [see Computational Models of Language Learning; Language Acquisition; Word Learning; Word Meaning]. The spatial relationships between word vectors have been shown to accurately reproduce a wide range of human behaviors such as similarity judgments ("cord" is similar to "string" but not to "smile") and lexical priming (people respond faster to "doctor" if they have just read or heard "hospital"). Figure 2 provides an example of how relations in distributional space correspond to semantic relations.

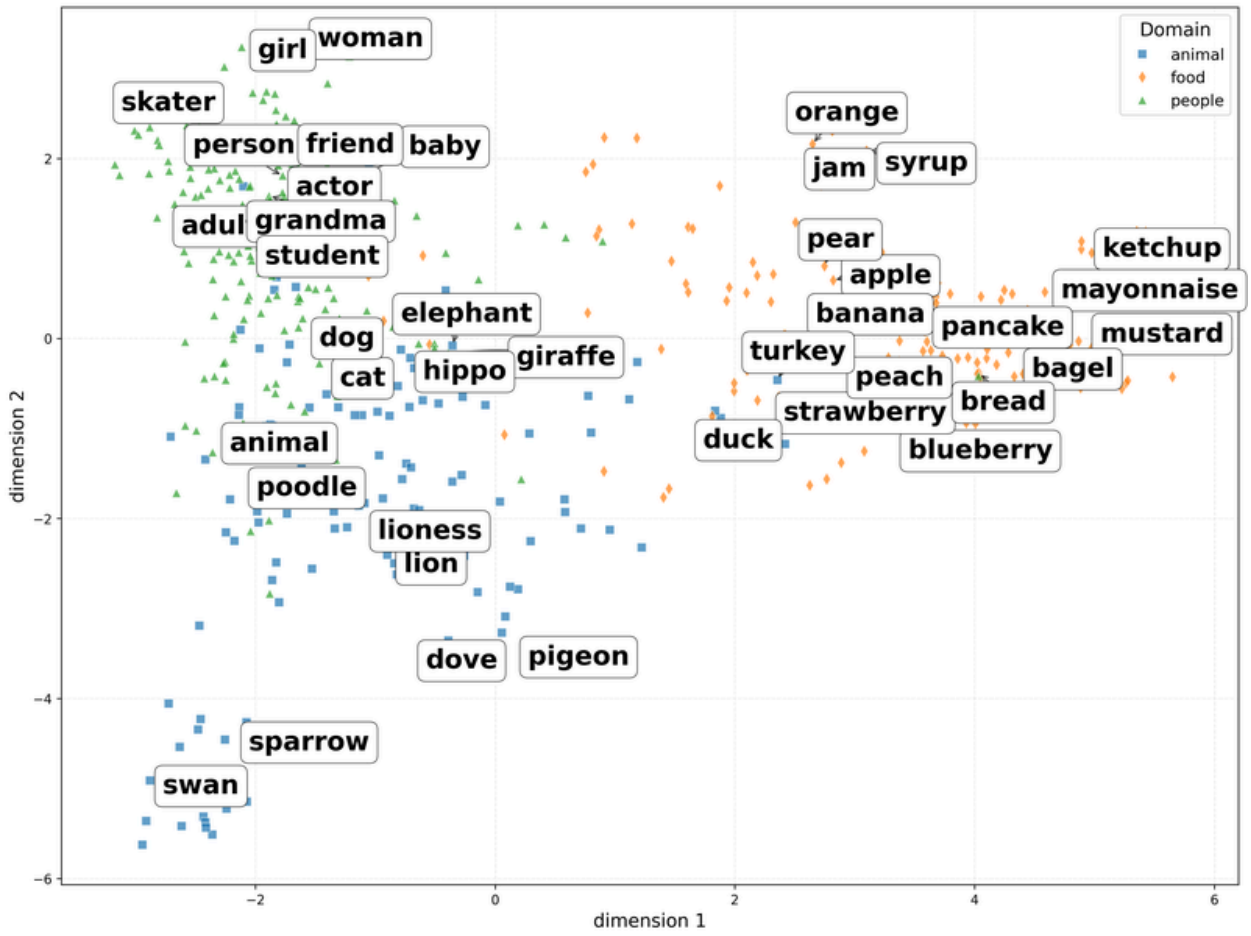

Figure 2

In distributional space, words are meaningfully arranged, both globally (people, animals, and food items are in different parts of the space) and locally (e.g., "mayonnaise," "ketchup," and "mustard" are closer to each other than they are to "jam" or "syrup"). The figure shows a sample of 569 nouns belonging to different domains represented as points in two-dimensional space. Obtained by principal component analysis from 300-dimensional word embeddings from spaCy (**Honnibal et al., 2020**).

Questions, controversies, and new developments

Contribution to theories of cognition

The success of distributional models suggests that humans acquire a significant portion of semantic knowledge simply through statistical learning of how words co-occur in their linguistic environment. However, how this fits into an overall theory of meaning and cognition is still unknown (Lenci & Sahlgren, 2023; Piantadosi et al., 2024).

Grounding

A major philosophical critique of distributional semantics has been that its notion of meaning is not *grounded* in the real world (Harnad, 1990) because the models learn only from statistical relationships between words [see Grounded Cognition]. This has two facets: whether the model's knowledge comes from text only (newer, multimodal models instead incorporate, e.g., visual information) and whether the model can solve references ("cat" in "The cat is sleeping" uttered by me on November 14, 2025, at 5:35 p.m. is used to refer to a particular cat, Lola). Newer vision and language models can simulate reference to some extent (e.g., they can produce quite accurate captions for images), but this issue is also still unresolved.

Other limitations of distributional models

Traditional distributional models produce a single vector for each word, but words can take different meanings in different contexts (e.g., "mouse" in "A mouse ran/broke"). This is partially solved by composing word vectors (for instance, the "mouse" vector contains both animal- and device-related semantic information, which can be selectively activated depending on whether the vector is combined with "ran" or "broke"). However, traditional distributional models afford only limited composition of words into phrases and sentences [see Compositionality]. Newer deep learning models have largely overcome both issues (Devlin et al., 2019). Finally, models absorb and often amplify the societal biases present in the data such as gender or racial prejudices (Bolukbasi et al., 2016). How to best deal with bias is an active research topic in artificial intelligence.

Broader connections

Distributional semantics is connected to broader debates about the nature of meaning (Baroni et al., 2014); to other domains in artificial intelligence that have also adopted deep learning, such as computer vision (having a common representational format has made it easier to incorporate other sources of information than text into distributional models, such as visual information from images); and to neuroscience, with some studies showing that distributional models can predict neural activity in human brains (Mitchell et al., 2008). It has traditionally had less influence in theoretical linguistics, but this is starting to change (Boleda, 2020; Erk, 2022).

Further reading

- Jurafsky, D., & Martin, J. H. (2025). Embeddings. In D. Jurafsky & J. H. Martin (Eds.), *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, with language models* (3rd ed.). Stanford University.
- Landauer, T. K., & Dumais, S. T. (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. https://doi.org/10.1037/0033-295X.104.2.211

- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1-31.
- Lenci, A., & Sahlgren, M. (2023). *Distributional semantics*. Cambridge University Press.

References

- Baroni, M., Bernardi, R., & Zamparelli, R. (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technology*, *9*, 241-346. https://aclanthology.org/2014.lilt-9.5/

  ↩
- Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, *6*(1), 213-234. https://doi.org/10.1146/annurev-linguistics-011619-030303

  ↩
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *arXiv*. https://doi.org/10.48550/arXiv.1607.06520

  ↩
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

  ↩
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv*. https://doi.org/10.48550/arXiv.1810.04805

  ↩
- Erk, K. (2022). The probabilistic turn in semantics and pragmatics. *Annual Review of Linguistics*, 8, 101-121. https://doi.org/10.1146/annurev-linguistics-031120-015515

  ↩
- Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. In J. Firth (Ed.), *Studies in linguistic analysis* (pp. 10-32). Basil Blackwell.

  ↩
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335-346. https://doi.org/10.1016/0167-2789(90)90087-6

  ↩

- Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146-162. https://doi.org/10.1080/00437956.1954.11659520

↩

- Landauer, T. K., & Dumais, S. T. (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. https://doi.org/10.1037/0033-295X.104.2.211

↩

- Lenci, A., & Sahlgren, M. (2023). *Distributional semantics*. Cambridge University Press.

↩

- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*. https://doi.org/10.48550/arXiv.1301.3781

↩

- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195. https://doi.org/10.1126/science.1152876

↩

- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

↩

- Piantadosi, S. T., Muller, D. C., Rule, J. S., Kaushik, K., Gorenstein, M., Leib, E. R., & Sanford, E. (2024). Why concepts are (probably) vectors. *Trends in Cognitive Sciences*, *28*(9), 844-856. https://doi.org/10.1016/j.tics.2024.06.011

↩

- Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613-620. https://doi.org/10.1145/361219.361220

↩

- Schütze, H. (1992). Dimensions of meaning. In R. Werner (Ed.), *Supercomputing '92: Proceedings of the 1992 ACM/IEEE conference on supercomputing* (pp. 787-796). IEEE Computer Society Press.

↩

- Sparck Jones, K. (1986). *Synonymy and semantic classification*. Edinburgh University Press. (Original work published 1964)

↩

- Wittgenstein, L. (1958). *Philosophical investigations (*G. E. M. Anscombe, Trans.). Basil Blackwell. (Original work published 1953)

  ↩