Denotations and distributions

Aurélie Herbelot & Ann Copestake

University of Cambridge
Computer Laboratory

Language Sciences in the 21st Century

Outline

The general intuition

Harris (1954)

Words that appear in the same context are semantically similar.

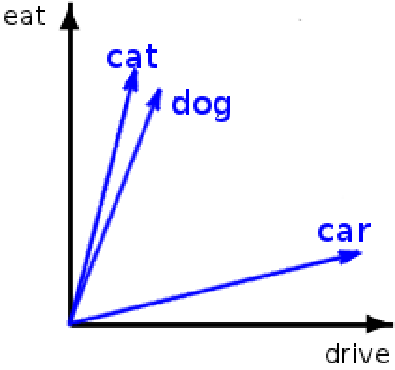
Firth (1957)

‘You shall know a word by the company it keeps.’

- **Distributions** are vectors in a multidimensional semantic space, that is, objects with a magnitude (length) and a direction.
- The **semantic space** has dimensions which correspond to possible contexts, as taken from a given corpus.

A distributional space

- A mini-distributional space, with two possible contexts, *eat* and *drive*.



- In practice, many more dimensions are used:
  cat [...dog 0.8, eat 0.7, joke 0.01, mansion 0.2, zebra 0.1...]

A continuous story

Ludwig Wittgenstein: Words are defined by their usage.

Margaret Masterman: Cambridge Language Research Unit (CLRU: 1955–1986).

Karen Spärck-Jones: Early experiments on distributional semantics: 1963, 1967.

Outline

Distributional semantics and the world

- Kempson (1977) and Cann (1993): a semantics should...
  - provide a hypothesis for how the building blocks of language are related to the world;
  - give an account of the meaning of words, phrases and sentences, and explain the relations between them.

Is distributional semantics Wittgensteinian?

- (Late) Wittgenstein: Meaning is use. Meaning only results from language games, not from a representationalist view of language (correspondence with the world). Not a metaphysical theory.
- Modern distributional semantics equates usage with textual (and sometimes visual) corpora. Reduced version of the Wittgensteinian thesis, as we (mostly) do away with pragmatics at implementation stage.
- Some corpora can be analysed as representing a particular pragmatic context. E.g. Wikipedia is a certain way to provide information, with specific normative rules.

Concrete

- Distribution for *concrete* (noun), as obtained from Wikipedia.

0.542296::and_c+steel_n
0.540451::steel_n+and_c
0.512329::slab_n+of_p()
0.466818::brick_n+and_c
0.463849::steel_n+or_c
0.453806::meter_n+of_p()
0.442502::and_c+glass_n
0.436364::stone_n+and_c
0.428527::and_c+brick_n
0.380303::be_v+material_n
0.374869::glass_n+and_c
0.374346::material_n+such+as_p()
0.374041::and_c+granite_n
0.367402::ton_n+of_p()
0.353181::or_c+stone_n

0.351596::yard_n+of_p()
0.342199::consistency_n+of_p()
0.340048::and_c+concrete_n
0.338328::or_c+metal_n
0.333411::centimeter_n+of_p()
0.331533::concrete_n+and_c
0.323514::exposed_a
0.317804::and_c+clay_n
0.31632::wood_n+and_c
0.31594::strength_n+of_p()
0.314691::foot_n+of_p()
0.312795::inch_n+of_p()
0.306334::Stone_n+and_c
0.304715::material_n+be_v
0.284954::and_c+ton_n

Concrete

- Distribution for *concrete* (noun), as obtained from Wikipedia.

0.542296::and_c+steel_n
0.540451::steel_n+and_c
0.512329::slab_n+of_p()
0.466818::brick_n+and_c
0.463849::steel_n+or_c
0.351596::yard_n+of_p()
0.342199::consistency_n+of_p()
0.340048::and_c+concrete_n
0.338328::or_c+metal_n
0.333411::centimeter_n+of_p()

System

Corpus: WikiWoods. Contexts: semantic dependencies. Semantic space: 100,000 most frequent contexts. Weighting: normalised PMI.

0.380303::be_v+material_n
0.374869::glass_n+and_c
0.374346::material_n+such+as_p()
0.374041::and_c+granite_n
0.367402::ton_n+of_p()
0.353181::or_c+stone_n
0.314691::foot_n+of_p()
0.312795::inch_n+of_p()
0.306334::Stone_n+and_c
0.304715::material_n+be_v
0.284954::and_c+ton_n

Concrete

- Distribution for *concrete* (noun), as obtained from Wikipedia.

0.542296::and_c+steel_n
0.540451::steel_n+and_c
0.512329::slab_n+of_p()
0.466818::brick_n+and_c
0.463849::steel_n+or_c
0.453806::meter_n+of_p()
0.442502::and_c+glass_n
0.436364::stone_n+and_c
0.428527::and_c+brick_n
0.380303::be_v+material_n
0.374869::glass_n+and_c
0.374346::material_n+such+as_p()
0.374041::and_c+granite_n
0.367402::ton_n+of_p()
0.353181::or_c+stone_n
0.351596::yard_n+of_p()
0.342199::consistency_n+of_p()
0.340048::and_c+concrete_n
0.338328::or_c+metal_n
0.333411::centimeter_n+of_p()
0.331533::concrete_n+and_c
0.323514::exposed_a
0.317804::and_c+clay_n
0.31632::wood_n+and_c
0.31594::strength_n+of_p()
0.314691::foot_n+of_p()
0.312795::inch_n+of_p()
0.306334::Stone_n+and_c
0.304715::material_n+be_v
0.284954::and_c+ton_n

An account of the meaning of words

- Distributional semantics does give an elegant solution to the problem of representing content words.
- But what about function words? *not, can, the, most*

Distributions and extension

- Distributions do not model extension (of course!)
- There is no satisfactory account of quantification in distributional semantics: *most, more than 37*. Compare with Tarskian approach where words refer to things in the world.
- Some lexical phenomena like antonymy cannot be represented without access to the denoted individuals. But corpora are no exhaustive description of the world.

Beyond raw usage

- Statistics extracted from corpora directly reflect language use but are perhaps insufficient to fully model meaning.
- Some sentences have more importance than others: *aardvarks are mammals*.
- Meaning might come from usage, but how?

Meaning as use *and* denotation

- The pragmatic stance of Wittgenstein makes it compatible with using set theory where needed. So we should perhaps have separate theories for separate parts of language.
- Computational semantics is working on a unifying theory. Why? Because it is not clear which blocks of language should be dealt with distributionally / set-theoretically (McNally, 2013).
- Meaning (including set theory) comes from usage. Can we have a theory that lets us go from distributionalism to set theory?
- Perhaps, by adding some inference...

Outline

Ideal distributions: the intuition

- If we want to relate usage-based theories with extensions we require both:
  - Possible utterances: i.e., sentences.
  - Correspondence with real world entities and relations.
- Ideal distribution: **all** the sentences that could be ‘appropriately’ uttered about some microworld.

Ideal distribution with grounded utterances

Microworld S_1: A jiggling black sphere (a) and a rotating white cube (b)

Possible utterances (restrict lexemes to *a*, *sphere*, *cube*, *object*, *rotate*, *jiggle*, *black*, *white*):

a sphere jiggles
a black sphere jiggles
a cube rotates
a white cube rotates
an object jiggles
a black object jiggles
an object rotates
a white object rotates

Also: *a black black sphere jiggles* . . .

LC context sets

Logical forms in simplified MRS:
a sphere jiggles: $a(x1), \text{sphere}^\circ(x1), \text{jiggle}^\circ(e1, x1)$
a black sphere jiggles: $a(x2), \text{black}^\circ(x2), \text{sphere}^\circ(x2), \text{jiggle}^\circ(e2, x2)$

x1, x2 . . . correspond to **linguistic entities**: obtained by parsing.

Context set for *sphere* (paired with S_1):

$$\text{sphere}^\circ = \{ \; < [x1][a(x1), \text{jiggle}^\circ(e1, x1)], S_1 >, \; < [x2][a(x2), \text{black}^\circ(x2), \text{jiggle}^\circ(e2, x2)], S_1 > \}$$

Context set: pair of **distributional argument tuple** and **distributional LF**.

Ideal distribution for S_1

$$\text{sphere}^\circ = \{ < [x1][a(x1), \text{jiggle}^\circ(e1, x1)], S_1 >, < [x2][a(x2), \text{black}^\circ(x2), \text{jiggle}^\circ(e2, x2)], S_1 > \}$$

$$\text{cube}^\circ = \{ < [x3][a(x3), \text{rotate}^\circ(e3, x3)], S_1 >, < [x4][a(x4), \text{white}^\circ(x4), \text{rotate}^\circ(e4, x4)], S_1 > \}$$

$$\text{object}^\circ = \{ < [x5][a(x5), \text{jiggle}^\circ(e5, x5)], S_1 >, < [x6][a(x6), \text{black}^\circ(x6), \text{jiggle}^\circ(e6, x6)], S_1 >, < [x7][a(x7), \text{rotate}^\circ(e7, x7)], S_1 >, < [x8][a(x8), \text{white}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 > \}$$

$$\text{jiggle}^\circ = \{ < [e1, x1][a(x1), \text{sphere}^\circ(x1)], S_1 >, < [e2, x2][a(x2), \text{black}^\circ(x2), \text{sphere}^\circ(x2)], S_1 >, < [e5, x5][a(x5), \text{object}^\circ(x5)], S_1 >, < [e6, x6][a(x6), \text{black}^\circ(x6), \text{object}^\circ(x6)], S_1 > \}$$

Ideal distribution for S_1, continued

$$
\begin{aligned}
\text{rotate}^\circ = \quad \{ \quad & < [e3, x3][a(x3), \text{cube}^\circ(x3)], S_1 >, \\
& < [e4, x4][a(x4), \text{white}^\circ(x4), \text{cube}^\circ(x4)], S_1 >, \\
& < [e7, x7][a(x7), \text{object}^\circ(x7)], S_1 >, \\
& < [e8, x8][a(x8), \text{white}^\circ(x8), \text{object}^\circ(x8)], S_1 >\} \\
\text{black}^\circ = \quad \{ \quad & < [x2][a(x2), \text{sphere}^\circ(x2), \text{jiggle}^\circ(e2, x2)], S_1 >, \\
& < [x5][a(x5), \text{object}^\circ(x5), \text{jiggle}^\circ(e5, x5)], S_1 >\} \\
\text{white}^\circ = \quad \{ \quad & < [x4][a(x4), \text{cube}^\circ(x4), \text{rotate}^\circ(e4, x4)], S_1 >, \\
& < [x8][a(x8), \text{object}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 >\}
\end{aligned}
$$

Relationship to standard notion of extension

For a predicate P, the distributional arguments of P° in lc_0 correspond to P', assuming real world equalities.

$$\text{sphere}^\circ = \{ \; < [x1][a(x1), \text{jiggle}^\circ(e1, x1)], S_1 >, \; < [x2][a(x2), \text{black}^\circ(x2), \text{jiggle}^\circ(e2, x2)], S_1 > \}$$

distributional arguments $x1, x2 =_{rw} a$ (where $=_{rw}$ stands for real world equality)

$$\text{sphere}' = \{a\}$$

Relationship to standard notion of extension

$$\text{object}^\circ = \{ \; < [x5][a(x5), \text{jiggle}^\circ(e5, x5)], S_1 >,$$
$$< [x6][a(x6), \text{black}^\circ(x6), \text{jiggle}^\circ(e6, x6)], S_1 >,$$
$$< [x7][a(x7), \text{rotate}^\circ(e7, x7)], S_1 >,$$
$$< [x8][a(x8), \text{white}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 > \}$$

distributional arguments $x5, x6 =_{rw} a$, $x7, x8 =_{rw} b$

Ideal and actual distributions

- Ideal distributions: all the things a speaker could say about the situation.
- Can (perhaps) be thought of in terms of a speaker's competence.
- Speaker dependent: *cup* or *mug*?
- Actual distributions correspond to things a speaker says and hears.
- Ideal distributions are expansions of actual distributions: e.g., *sphere* implies *object*.
- Frequency is relevant to actual distributions but not to ideal distributions.

A inferentialist view?

- Obtaining ideal distributions from actual distributions (language use) relies on making correct inferences.
- Brandom (1994): meaning is use *in inferences*.
    - When stating *Tweety is a bird*, I commit to being able to provide the justification *Tweety lays eggs and has wings*.
    - The language game of ‘providing reasons’ is similar to the process of inferring ideal from actual distributions.
- The task of computational semantics is to model *how* such inferences are made.

Outline

Conclusion

- Distributional semantics aims at modelling ‘language as use’.
- Without further qualification, it is unclear how observed language use should account for phenomena dealt with by set-theoretic semantics.
- An inferentialist approach may explain how models are obtained from raw linguistic occurrences.

Distributional semantics and language sciences

- Distributional techniques are practical computational methods for (some aspects of?) lexical semantic representation.
  - Experiments with learner data.
  - Syntax-semantics interface in linguistics.
- Learning from contexts has psychological plausibility, but:
  - Grounding.
  - Corpus size.
  - More information from context: deeper distributional semantics.
- Philosophical aspects.