Synonymy in an approach to combined distributional and compositional semantics

Ann Copestake and Aurélie Herbelot

Computer Laboratory
University of Cambridge

October 2010
Combining compositional and distributional semantics

- Combining compositional and distributional techniques, based on existing approaches to compositional semantics.
- Replace (or augment) the standard notion of lexical denotation with a distributional notion. e.g., instead of cat’, use cat°: the set of all linguistic contexts in which the lexeme cat occurs.
- Contexts are expressed as logical forms.
- Primary objective: better models of lexical semantics with compositional semantics.
- Psychological plausibility: learnability.
Ideal distribution with grounded utterances

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

Possible utterances (restricted lexemes, no logical redundancy in utterance):

- 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
LC context sets

Logical forms:

- a sphere jiggles: \( a(x1), \text{sphere}(x1), \text{jiggle}(e1, x1) \)
- a black sphere jiggles:
  \( a(x2), \text{black}(x2), \text{sphere}(x2), \text{jiggle}(e2, x2) \)

Context set for \textit{sphere} (paired with \( S_1 \)):

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

Context set: pair of \text{distributional argument tuple} and \text{distributional LF}.
LF assumptions and slacker semantics

Slacker assumptions:

1. don’t force distinctions which are unmotivated by syntax
2. keep representations ‘surfacy’
3. (R)MRS, but simplified LFs here

Main points:

▶ Word sense distinctions only if syntactic effects: don’t even distinguish traditional bank senses.
▶ Underspecification of quantifier scope etc
▶ Eventualities, (neo-)Davidsonian.
▶ Equate entities (i.e., x1 etc) only according to sentence syntax.
Ideal distribution for $S_1$

sphere $\circ = \{ < [x1][a(x1), \text{jiggle} \circ (e1, x1)], S_1 >,$
\hspace{1cm}< [x2][a(x2), \text{black} \circ (x2), \text{jiggle} \circ (e2, x2)], S_1 > \}$

cube $\circ = \{ < [x3][a(x3), \text{rotate} \circ (e3, x3)], S_1 >,$
\hspace{1cm}< [x4][a(x4), \text{white} \circ (x4), \text{rotate} \circ (e4, x4)], S_1 > \}$

object $\circ = \{ < [x5][a(x5), \text{jiggle} \circ (e5, x5)], S_1 >,$
\hspace{1cm}< [x6][a(x6), \text{black} \circ (x6), \text{jiggle} \circ (e6, x6)], S_1 >,$
\hspace{1cm}< [x7][a(x7), \text{rotate} \circ (e7, x7)], S_1 >,$
\hspace{1cm}< [x8][a(x8), \text{white} \circ (x8), \text{rotate} \circ (e8, x8)], S_1 > \}$

jiggle $\circ = \{ < [e1, x1][a(x1), \text{sphere} \circ (x1)], S_1 >,$
\hspace{1cm}< [e2, x2][a(x2), \text{black} \circ (x2), \text{sphere} \circ (x2)], S_1 >,$
\hspace{1cm}< [e5, x5][a(x5), \text{object} \circ (x5)], S_1 >,$
\hspace{1cm}< [e6, x6][a(x6), \text{black} \circ (x6), \text{object} \circ (x6)], S_1 > \}$
Ideal distribution for $S_1$, continued

\[
\text{rotate}^\circ = \{ < [e3, x3][a(x3), \text{cube}^\circ(x3)], S_1 >, \\
< [e4, x4][a(x4), \text{white}^\circ(x4), \text{cube}^\circ(x4)], S_1 >, \\
< [e7, x7][a(x7), \text{object}^\circ(x7)], S_1 >, \\
< [e8, x8][a(x8), \text{white}^\circ(x8), \text{object}^\circ(x8)], S_1 > \}
\]

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

\[
\text{white}^\circ = \{ < [x4][a(x4), \text{cube}^\circ(x4), \text{rotate}^\circ(e4, x4)], S_1 >, \\
< [x8][a(x8), \text{object}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 > \}
\]
Relationship to standard notion of extension

For a predicate \( P \), the distributional arguments of \( P^\circ \) in \( lc_0 \) correspond to \( P' \), assuming real world equalities.

\[
sphere^\circ = \{ < [x1][a(x1), \text{jiggle}^\circ(e1, x1)], S_1 >, \\
< [x2][a(x2), \text{black}^\circ(x2), \text{jiggle}^\circ(e2, x2)], S_1 > \}
\]
distributional arguments \( x1, x2 =_{rw} a \) (where \( =_{rw} \) stands for real world equality):

\[
object^\circ = \{ < [x5][a(x5), \text{jiggle}^\circ(e5, x5)], S_1 >, \\
< [x6][a(x6), \text{black}^\circ(x6), \text{jiggle}^\circ(e6, x6)], S_1 >, \\
< [x7][a(x7), \text{rotate}^\circ(e7, x7)], S_1 >, \\
< [x8][a(x8), \text{white}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 > \}
\]
distributional arguments \( x5, x6 =_{rw} a, x7, x8 =_{rw} b \)
Ideal distribution properties

- Logical inference is possible.
- Lexical similarity, hyponymy, (denotational) synonymy in terms of context sets.
- Word ‘senses’ as subspaces of context sets.
- Given context sets, learner can associate lexemes with real world entities on plausible assumptions about perceptual similarity.
- Ideal distribution is unrealistic, but a target to approximate (partially) from actual distributions.
Actual distributions

- Actual distributions correspond to an individual’s language experience (problematic with existing corpora).
  - For low-to-medium frequency words, individuals’ experiences will differ.
    e.g., BNC very roughly equivalent to 5 years exposure(?): *rancid* occurs 77 times, *rancorous* 20.
    Essential to model individual differences, negotiation of meaning.
  - Google-sized distributional models MAY help approximate real world knowledge, but not realistic for knowledge of word use.
- Some (not all) contexts involve perceptual grounding.
- Word frequencies are apparent in actual distributions.
Assumptions about synonymy

- Near-synonymy is frequent, absolute synonymy relates to dialect etc.

- Synonymy is more interesting for its absence than its presence:
  - Language learners (and others) tend to assume non-synonymy.
    e.g., “labeling entities with distinct words leads infants to create representations of two distinct individuals” (Carey, 2009:p 277)
  - Blocking: preemption by synonymy (higher frequency forms preferred).

- With respect to a specific context, near-synonyms will often be substitutable.

- Word sense assumptions affect synonymy assumptions.
Synonymy in LC context sets

- Full denotational synonyms have identical ideal context sets, near-synonyms overlapping ideal context sets (identical for some situations).
- Synonyms and near-synonyms both expected to have similar actual distributions (but sparse data, dialect etc).
- No hard line between near-synonyms and non-synonyms.
- Lack of word sense distinctions affects synonymy assumptions.
- Degree of synonymy between two lexemes will vary between individuals.
Near-synonymy and meaning acquisition

- Readers only need around three uses to obtain a working idea of a new word’s meaning.
- Hypothesis: understanding a new word (without definition) can be modelled by two-phase context set comparison:
  - initial approximation: e.g., *rancid* is similar to *off*
  - acquisition of differentiating information **characteristic contexts**: e.g., *rancid* tends to appear with fatty foods (or dairy foods, or . . .)
- Sometimes obtain expert knowledge: e.g., *rancid* refers to oxidation of fat.
- People’s beliefs about low-to-medium frequency words may differ but approximation is usually good enough for communication.
Are *frumpy* and *dowdy* synonyms?

My intuition (pre data check): both negative, both refer to women/women’s clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.

BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  - *frumpy*: *old* (twice), *new*
  - *dowdy*: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

*slightly dowdy elegance — if there could be such a thing*
Are *frumpy* and *dowdy* synonyms?

My intuition (pre data check): both negative, both refer to women/women’s clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.

BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  - frumpy: *old* (twice), *new*
  - dowdy: *plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

*slightly dowdy elegance — if there could be such a thing*
Are *frumpy* and *dowdy* synonyms?

My intuition (pre data check): both negative, both refer to women/women’s clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.

BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.

- Conjoined adjectives
  
  *frumpy*: *old* (twice), *new*
  
  *dowdy*: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

*slightly dowdy elegance — if there could be such a thing*
Are *frumpy* and *dowdy* synonyms?

My intuition (pre data check): both negative, both refer to women/women’s clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless.*

BNC:

- *frumpy*: 17 total. 8 clothing, 9 people.
- *dowdy*: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  - *frumpy*: *old* (twice), *new*
  - *dowdy*: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

slightly dowdy elegance — if there could be such a thing
Are *frumpy* and *dowdy* synonyms?

My intuition (pre data check): both negative, both refer to women/women’s clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.

BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.

- Conjoined adjectives
  - frumpy: old (twice), new
  - dowdy: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

*slightly dowdy elegance — if there could be such a thing*
Full synonymy

- We hypothesize that full synonyms are acquired differently from near-synonyms, generally by (relatively) explicit definition:

  *The aubergine (eggplant) has to be one of my favourite vegetables.*

- Full synonyms allow substitution in the ideal distribution, i.e., they share context sets.

- Contrast with near-synonyms which maintain their own distributions.
Conclusions

- Lexicalised compositionality is very preliminary . . .
- Our proposed approach differs from standard distributional accounts in:
  - Being based on compositional semantics and hence allowing (in principle) for logical inference.
  - Ideal distribution as target for manipulations of actual distributions.
  - Emphasis on the individual’s experience.
- Synonymy:
  - Near-synonymy as (graded) context set similarity, full synonymy as context set identity in ideal distributions.
  - Emphasis on individual distributions: speakers may vary.
  - Explicit definition as well as distribution.
Blocking

- *sank/sunk* but *dreamt/dreamed*
- *curious/curiosity, glorious/glory/*gloriosity*
- *stealer/thief*
  - ? She was a stealer. She was a scene stealer/stealer of fast cars.
- *lamb, rabbit, ?pig (pork), ?cow (beef)*
- *bigger/?more big, odder/more odd, obscurer/more obscure*

Assumption: speakers use the highest frequency form to convey a particular meaning (plus connotation etc) (Briscoe and Copestake, 1999)