Distributional semantics meets MRS?

Ann Copestake and Aurélie Herbelot

Computer Laboratory, University of Cambridge
and
Department Linguistik, Universität Potsdam

July 2012

Distributional semantics and DELPH-IN

Distributional semantics: family of techniques for representing word meaning based on contexts of use. Simplest approaches use vector representation based on characteristics words extracted from window. Parsed data sometimes better.

it was authentic scrumpy, rather sharp and very strong
we could taste a famous local product — scrumpy
spending hours in the pub drinking scrumpy

1. Theoretical issues: lexicalised compositionality (Copestake and Herbelot)
2. Distributions from DELPH-IN output
3. Distributional techniques improving DELPH-IN performance?
4. Providing deeper semantics?

Distributional semantics and DELPH-IN

Distributional semantics: family of techniques for representing word meaning based on contexts of use. Simplest approaches use vector representation based on characteristics words extracted from window. Parsed data sometimes better.

it was authentic scrumpy, rather sharp and very strong
we could taste a famous local product — scrumpy
spending hours in the pub drinking scrumpy

1. Theoretical issues: lexicalised compositionality (Copestake and Herbelot)
2. Distributions from DELPH-IN output
3. Distributional techniques improving DELPH-IN performance?
4. Providing deeper semantics?

Distributional semantics and DELPH-IN

Distributional semantics: family of techniques for representing word meaning based on contexts of use. Simplest approaches use vector representation based on characteristics words extracted from window. Parsed data sometimes better.

it was authentic scrumpy, rather sharp and very strong
we could taste a famous local product — scrumpy
spending hours in the pub drinking scrumpy

1. Theoretical issues: lexicalised compositionality (Copestake and Herbelot)
2. Distributions from DELPH-IN output
3. Distributional techniques improving DELPH-IN performance?
4. Providing deeper semantics?

Distributional semantics and DELPH-IN

Distributional semantics: family of techniques for representing word meaning based on contexts of use. Simplest approaches use vector representation based on characteristics words extracted from window. Parsed data sometimes better.

it was authentic scrumpy, rather sharp and very strong
we could taste a famous local product — scrumpy
spending hours in the pub drinking scrumpy

1. Theoretical issues: lexicalised compositionality (Copestake and Herbelot)
2. Distributions from DELPH-IN output
3. Distributional techniques improving DELPH-IN performance?
4. Providing deeper semantics?

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$^\circ$: 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: Hebbian learnability.

`http://www.cl.cam.ac.uk/~aac10/papers/lc1-0web.pdf`

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

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

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

Context set: pair of distributional argument tuple and 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

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

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

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

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

Ideal distribution for S_1, continued

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

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.

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

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

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

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 and ‘individuated’, situation-annotated corpora

- 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.

Lexicalised compositionality: status and plans

- Investigation of various semantic phenomena from the ideal distribution perspective.
- Possible pilot experiments with corpus acquisition and/or language learner corpora.
- Build distributions based on predicates applied to particular entities: feasible, but implies anaphora resolution, hence ERG parsing unsuitable without robustness.

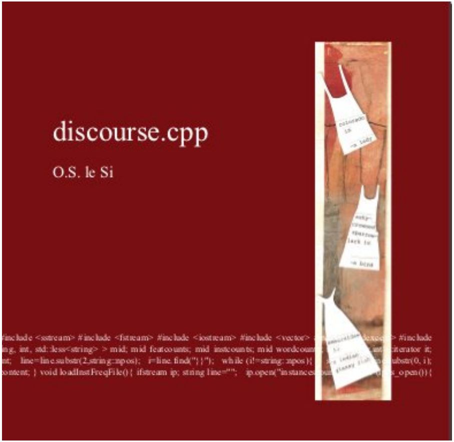
Adjective and binomial ordering

- *gigantic striped box* not *striped gigantic box*
- *brandy and soda* not *soda and brandy*
- ordering principles partially semantic
- lots of discussion about gendered examples: e.g., *boy and girl*
- our hypothesis: humans maintain order of known examples, order unseen by semantic similarity with seen

Adjective and binomial ordering: Kumar (2012)

- Same type of model for adjectives and binomials: unseen cases ordered by k-nearest neighbour comparison to seen examples using distributional similarity.
- Unparsed WikiWoods data: significantly better than using positional probabilities.
- Parsed WikiWoods converted to DMRS, limited relations: similar results to positional probabilities (but much less data).
- Expect further improvement using phonological features in addition.

Discourse.cpp by O.S. le Si, edited by Aurélie Herbelot



http://www.peerpress.de/

Characteristic contexts for **strength**

reflect_v ARG1 membership_n ARG2 *
decrease_v ARG1 pressure_n ARG2 *
assess_v ARG1 player_n ARG2 *
attack_v ARG1 * ARG2 Prussia
begin_v ARG1 * ARG2 bleed_v
describe_v ARG1 part_n ARG2 *
describe_v ARG1 point_n ARG2 *
draw_v ARG1 * ARG2 reaction_n
help_v ARG1 * ARG2-4 overcome_v
inhibit_v ARG1 * ARG2 growth_n
moreover_r ARG1 interaction_n ARG2 *
provide_v ARG1 hull_n ARG2 *
provide_v ARG1 soil_n ARG2 *
reach_v ARG1 bond_n ARG2 *

from Discourse.cpp

Similarities for **strength**

strength 1
companionship 0.0410899
discretion 0.0325424
needle 0.0282791
standing 0.0249236
battlefield 0.0242123
depth 0.0164379
representation 0.0160898

battalion 0.0157682
myth 0.0149577
factor 0.0143694
knowledge 0.0137592
detail 0.0117955
soldier 0.0115114
advance 0.0108719
tone 0.0107681

strength the poem: Content selected out of the 16 nouns most similar to strength. Two nouns changed into gerunds. Prepositions and conjunctions added afterwards.

from Discourse.cpp

Strength

Needle standing battlefield
Depth of representation
Battalion myths and
Soldier advancing tone

OR

Companionship

Whisky

some fubar song with a wawa... the phoneme p... the
backwash starts... liquor, turpentine... the broth: marijuana
beverage with expressionism... with honey... it curds and clogs
and mashes its own debris with a snobbery of naturalist...
- - - banker with his - - - leather... - - - banker...

I chill.
I age.
I darken.
I blend.

Like an old punk, sulphide in her veins.