Distributional semantics for linguists: 3b

Ann Copestake and Aurélie Herbelot

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

August 2012

Outline

Distributional and compositional semantics

Ideal distributions

Actual distributions

Outline.

Starting points

- Compositional semantics is relatively well understood: e.g., generalised quantifiers.
- Reasonably efficient broad-coverage computational grammars with compositional semantics are available for a number of languages.
  - DELPH-IN: grammars using MRS for English, Japanese, German, Norwegian, Spanish, Portuguese: small grammars for all major language families (Emily Bender, Grammar Matrix)
- But conventional notions of denotation (*cat* is cat′ etc) are not satisfactory.
- Can distributional semantics give an alternative, without completely rethinking composition?

Logical representation in MRS

Some big angry dog barks loudly

$\exists x4[\text{big}'(x4) \wedge \text{angry}'(x4) \wedge \text{dog}'(x4) \wedge \text{bark}'(e2, x4) \wedge \text{loud}'(e2)]$

l1:a1:_some_q, BV(a1,x4), RSTR(a1,h5), BODY(a1,h6),
l2:a2:_big_a(e8), ARG1(a2,x4),
l2:a3:_angry_a(e9), ARG1(a3,x4),
l2:a4:_dog_n(x4), l4:a5:_bark_v(e2), ARG1(a5,x4),
l4:a6:_loud_a(e10), ARG1(a6,e2), h5 $=_q$ l2

Logical representation in MRS

Some big angry dog barks loudly

$\exists x4[\text{big}'(x4) \wedge \text{angry}'(x4) \wedge \text{dog}'(x4) \wedge \text{bark}'(e2, x4) \wedge \text{loud}'(e2)]$

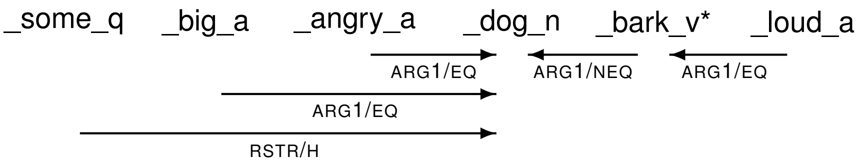
l1:a1:_some_q, BV(a1,x4), RSTR(a1,h5), BODY(a1,h6),
l2:a2:_big_a(e8), ARG1(a2,x4),
l2:a3:_angry_a(e9), ARG1(a3,x4),
l2:a4:_dog_n(x4), l4:a5:_bark_v(e2), ARG1(a5,x4),
l4:a6:_loud_a(e10), ARG1(a6,e2), h5 $=_q$ l2

Logical representation in MRS

Some big angry dog barks loudly

$\exists x4[\text{big}'(x4) \wedge \text{angry}'(x4) \wedge \text{dog}'(x4) \wedge \text{bark}'(e2, x4) \wedge \text{loud}'(e2)]$

l1:a1:_some_q, BV(a1,x4), RSTR(a1,h5), BODY(a1,h6),
l2:a2:_big_a(e8), ARG1(a2,x4),
l2:a3:_angry_a(e9), ARG1(a3,x4),
l2:a4:_dog_n(x4), l4:a5:_bark_v(e2), ARG1(a5,x4),
l4:a6:_loud_a(e10), ARG1(a6,e2), h5 $=_q$ l2



Quantifier-free MRS (this talk)

Some big angry dog barks loudly

Full RMRS:

l1:a1:_some_q, BV(a1,x4), RSTR(a1,h5), BODY(a1,h6),
l2:a2:_big_a(e8), ARG1(a2,x4),
l2:a3:_angry_a(e9), ARG1(a3,x4),
l2:a4:_dog_n(x4), l4:a5:_bark_v(e2), ARG1(a5,x4),
l4:a6:_loud_a(e10), ARG1(a6,e2), h5 $=_q$ l2

Simplified MRS:

some_q(x4), big_a(x4),
angry_a(x4),
dog_n(x4), bark_v(e2,x4),
loud_a(e2)

A longer example

Very few of the Chinese construction companies consulted were even remotely interested in entering into such an arrangement with a local partner.

A longer example

Very few of the Chinese construction companies consulted were even remotely interested in entering into such an arrangement with a local partner.

l_3:part_of(x_4{PERS *3*, NUM *pl*}, x_5{PERS *3*, NUM *pl*}),
l_6:udef_q(x_4, h_7, h_8),
l_3:_very_x_deg(e_9,e_{10}{SF *prop*}),
l_3:_few_a(e_{10}, x_4),
l_{11}:_the_q(x_5, h_{13}, h_{12}),
l_{14}:compound(e_{16}{SF *prop*, TENSE *untensed*, MOOD *indicative*, PROG -, PERF -}, x_5, x_{15}),
l_{17}:udef_q(x_{15}, h_{18}, h_{19}),
l_{20}:_chinese_a_1(e_{21}{SF *prop*, TENSE *untensed*, MOOD *indicative*}, x_{15}),
l_{20}:_construction_n(x_{15}),
l_{14}:_company_n(x_5),
l_3:_consult_v_1(e_{24}{SF *prop*, TENSE *untensed*, MOOD *indicative*, PROG -, PERF -}, p_{25}, x_4),
l_{27}:_even_a_1(e_{28},e_2{SF *prop*, TENSE *past*, MOOD *indicative*, PROG -, PERF -}),
l_{27}:_remotely_x_deg(e_{29}{SF *prop*, TENSE *untensed*, MOOD *indicative*, PROG -, PERF -}, e_2),
l_{27}:_interested_a_in(e_2, x_4, x_{30}{PERS *3*, NUM *sg*, GEND *n*}),
l_{31}:udef_q(x_{30}, h_{32}, h_{33}),
l_{34}:_enter_v_1(e_{35}{SF *prop*, TENSE *untensed*, MOOD *indicative*, PROG +, PERF -}, p_{36}),
l_{37}:nominalization(x_{30}, h_{34}),
l_{34}:_into_p(e_{38},e_{35}, x_{39}{PERS *3*, NUM *sg*, IND +}),
l_{40}:_such+a_q(x_{39}, h_{42}, h_{41}),
l_{43}:_arrangement_n_1(x_{39}),
l_{37}:_with_p($e_{44}x_{30}$, x_{45}{PERS *3*, NUM *sg*, IND +}),
l_{46}:_a_q(x_{45}, h_{48}, h_{47}),
l_{49}:_local_a_1(e_{50}{SF *prop*, TENSE *untensed*, MOOD *indicative*}, x_{45}),
l_{49}:_partner_n_1(x_{45}), $h_{48} =_q l_{49}$, $h_{42} =_q l_{43}$, $h_{32} =_q l_{37}$, $h_{18} =_q l_{20}$, $h_{13} =_q l_{14}$, $h_7 =_q l_3$

LF assumptions and slacker semantics

Slacker assumptions:

1. don't force distinctions which are unmotivated by syntax
2. keep representations 'surfacy'

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: linguistic entities.
- Separate step of equating to real world entities.

Lexicalised compositionality (LC)

- 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 combined with compositional semantics.

Distributions and semantics

- Conventional distributions fail to capture semantic ideas:
  - Full vs near synonymy, homonymy, antonymy.
  - Quantification.
  - Senses (perhaps).

  What's missing is any notion of an individual entity.
- So, 'deeper' distributional semantics (cf Clark and Pulman 2007)
- We start with an idealized notion of a distribution ...

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

Outline.

Distributional and compositional semantics

Ideal distributions

Actual distributions

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

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.

Ideal distribution for S_1

$$\begin{aligned}
\text{sphere}^\circ = \quad & \{ \quad < [x1][a(x1), \text{jiggle}^\circ(e1, x1)], S_1 >, \\
& \quad\;\; < [x2][a(x2), \text{black}^\circ(x2), \text{jiggle}^\circ(e2, x2)], S_1 > \} \\
\text{cube}^\circ = \quad & \{ \quad < [x3][a(x3), \text{rotate}^\circ(e3, x3)], S_1 >, \\
& \quad\;\; < [x4][a(x4), \text{white}^\circ(x4), \text{rotate}^\circ(e4, x4)], S_1 > \} \\
\text{object}^\circ = \quad & \{ \quad < [x5][a(x5), \text{jiggle}^\circ(e5, x5)], S_1 >, \\
& \quad\;\; < [x6][a(x6), \text{black}^\circ(x6), \text{jiggle}^\circ(e6, x6)], S_1 >, \\
& \quad\;\; < [x7][a(x7), \text{rotate}^\circ(e7, x7)], S_1 >, \\
& \quad\;\; < [x8][a(x8), \text{white}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 > \} \\
\text{jiggle}^\circ = \quad & \{ \quad < [e1, x1][a(x1), \text{sphere}^\circ(x1)], S_1 >, \\
& \quad\;\; < [e2, x2][a(x2), \text{black}^\circ(x2), \text{sphere}^\circ(x2)], S_1 >, \\
& \quad\;\; < [e5, x5][a(x5), \text{object}^\circ(x5)], S_1 >, \\
& \quad\;\; < [e6, x6][a(x6), \text{black}^\circ(x6), \text{object}^\circ(x6)], S_1 > \}
\end{aligned}$$

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.

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

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

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

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

Context sets as vectors

	jiggle $^\circ$(e,x)	rotate $^\circ$(e,x)	sphere $^\circ$(x)	cube $^\circ$(x)	object $^\circ$(x)
sphere $^\circ$	1	0	0	0	0
cube $^\circ$	0	1	0	0	0
object $^\circ$	1	1	0	0	0
black $^\circ$	1	0	1	0	1
white $^\circ$	0	1	0	1	1

- One way of generalising over the context sets.
- Variant semantic representations allow more possibilities.

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.

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 primarily expansions of actual distributions: e.g., *sphere* implies *object*.
- Frequency is relevant to actual distributions but not to ideal distributions.

Lexicalised compositionality: status and plans

- Investigation of various semantic phenomena from the ideal distribution perspective.
- Pilot experiments (Aurélie, Friday)
- Experiments with child language data?
- Build distributions based on predicates applied to particular entities: requires anaphora resolution etc.

Outline.

Distributional and compositional semantics

Ideal distributions

Actual distributions

Actual distributions and corpora

- LC actual distributions are an individual's experience, but this is highly problematic with existing corpora.
- Google-scale models MAY approximate real world knowledge, but not representative of individual's word use.
  - We don't even know how many words 'typical' individuals hear in a day . . .
  - For low-to-medium frequency words, individuals' experiences must differ.
    e.g., 100 million word BNC very roughly equivalent to 5 years exposure but quite unlike any individual's experience.
  - In BNC, *rancid* occurs 77 times: frequent for some people and almost unknown for others?
  - A different type of corpus is essential to model individual differences, negotiation of meaning.

Individuated, situation-annotated corpora

- Collect data based on known individuals' experience.
- Ideally, all language heard and read, spoken and written over a period of time.
- Some (not all) contexts involve perceptual grounding: some indication of this would be useful.
- Technologically feasible, legally complex!
- Approximations: e.g., web data with known authorship?
- Not just for LC!

Individuated, situation-annotated corpora

- Collect data based on known individuals' experience.
- Ideally, all language heard and read, spoken and written over a period of time.
- Some (not all) contexts involve perceptual grounding: some indication of this would be useful.
- Technologically feasible, legally complex!
- Approximations: e.g., web data with known authorship?
- Not just for LC!

Summary

- LC: one of a number of attempts to look at combining distributional and compositional semantics.
- Current aim: provide a theoretical account which has the necessary properties.
- Full-scale experiments would require new corpora, but pilot experiments now.