Lexicalised Compositionality: relating distributional and model-theoretic semantics

and MRS ...

Ann Copestake

Computer Laboratory
University of Cambridge

September 2014

Outline

Overview of MRS/RMRS/DMRS

Reflecting on MRS

Relating distributional and model-theoretic approaches

Some big questions . . .

- To what extent are computational semanticists committed to a particular philosophical approach to semantics?
- Could computational approaches lead to advances in theoretical ideas about semantics?
- Maybe an 'ordinary language' / 'meaning as use' philosophical approach can only be tested computationally?

Lexicalised compositionality: part of an attempt to think about relationships between approaches to semantics from a (broad coverage) computational semantics perspective.

Outline

Overview of MRS/RMRS/DMRS

Reflecting on MRS

Relating distributional and model-theoretic approaches

One page history of compositional semantics

- Aristotle (c350 BCE) — syllogisms:
  Every P is D, and every S is a P; so every S is D
- Medieval logicians: *dictum de omni* and *dictum de nullo*
  Rex is a brown dog implies Rex is a dog
  Rex is not a dog implies Rex is not a brown dog

  BUT couldn't handle some patient respects some doctor
  and
  every doctor is a senator implies
  some patient respects some senator
- Frege (1879): modern logic. Solves earlier problems but treats natural language structure as misleading.
- Montague (1970, 1974): symbolic logic systematically generated from natural language fragment.

(Pietroski in Stanford Encyclopedia of Philosophy)

One page history of compositional semantics

- Aristotle (c350 BCE) — syllogisms:
  Every P is D, and every S is a P; so every S is D
- Medieval logicians: *dictum de omni* and *dictum de nullo*
  Rex is a brown dog implies Rex is a dog
  Rex is not a dog implies Rex is not a brown dog

  BUT couldn't handle some patient respects some doctor
  and
  every doctor is a senator implies
  some patient respects some senator
- Frege (1879): modern logic. Solves earlier problems but treats natural language structure as misleading.
- Montague (1970, 1974): symbolic logic systematically generated from natural language fragment.

(Pietroski in Stanford Encyclopedia of Philosophy)

One page history of compositional semantics

- Aristotle (c350 BCE) — syllogisms:
  Every P is D, and every S is a P; so every S is D
- Medieval logicians: *dictum de omni* and *dictum de nullo*
  Rex is a brown dog implies Rex is a dog
  Rex is not a dog implies Rex is not a brown dog

  BUT couldn't handle some patient respects some doctor
  and
  every doctor is a senator implies
  some patient respects some senator
- Frege (1879): modern logic. Solves earlier problems but treats natural language structure as misleading.
- Montague (1970, 1974): symbolic logic systematically generated from natural language fragment.

(Pietroski in Stanford Encyclopedia of Philosophy)

One page history of compositional semantics

- Aristotle (c350 BCE) — syllogisms:
  Every P is D, and every S is a P; so every S is D
- Medieval logicians: *dictum de omni* and *dictum de nullo*
  Rex is a brown dog implies Rex is a dog
  Rex is not a dog implies Rex is not a brown dog

  BUT couldn't handle some patient respects some doctor and
  every doctor is a senator implies
  some patient respects some senator
- Frege (1879): modern logic. Solves earlier problems but treats natural language structure as misleading.
- Montague (1970, 1974): symbolic logic systematically generated from natural language fragment.

(Pietroski in Stanford Encyclopedia of Philosophy)

One page history of compositional semantics

- Aristotle (c350 BCE) — syllogisms:
  Every P is D, and every S is a P; so every S is D
- Medieval logicians: *dictum de omni* and *dictum de nullo*
  Rex is a brown dog implies Rex is a dog
  Rex is not a dog implies Rex is not a brown dog

  BUT couldn't handle some patient respects some doctor
  and
  every doctor is a senator implies
  some patient respects some senator
- Frege (1879): modern logic. Solves earlier problems but treats natural language structure as misleading.
- Montague (1970, 1974): symbolic logic systematically generated from natural language fragment.

(Pietroski in Stanford Encyclopedia of Philosophy)

Reducing the gap between Frege and natural language

- Event semantics (Davidson, 1967). Also Hobbs (1985) "Ontological promiscuity".
  $\text{chase}'(e, x, y) \wedge \text{quick}'(e)$
- Discourse Representation Theory (Kamp, 1981).
- Generalized quantifiers (Barwise and Cooper, 1981).
  $\text{some}(\lambda x[dog(x)], \lambda y[bark(y)]) \equiv$ some(x,dog(x),bark(x))
- Quantifier scope underspecification (Alshawi and Crouch, 1992).
- Simplified composition:
  Full quantifier scope underspecification means NPs of type e, transitive verbs of type $\langle e, \langle e, t \rangle \rangle$
  - Lambda calculus (possibly with labels).
  - Algebraic approaches (Zeevat, 1989).

Quantifier scope ambiguity

Every cat chased some dog

$$\forall x[\text{cat}'(x) \implies \exists y[\text{dog}'(y) \wedge \text{chase}'(x, y)]]$$
$$\exists y[\text{dog}'(y) \wedge \forall x[\text{cat}'(x) \implies \text{chase}'(x, y)]]$$

- Both scopes valid (grammar fragment shown only gives one — Montague used “quantifying-in rules”).
- Cannot decide between scope on the basis of syntax.
- Requires scope disambiguation (or enumeration) to produce valid logical representation.

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

Minimal Recursion Semantic (MRS)

Some big dog chased every cat

l1:some(x,h1,h2), h1 qeq l2, l2:big(x), l2:dog(x),
l4:chase(e,x,y), l5:every(y,h3,h4), h3 qeq l6, l6:cat(y)

Elementary predications (EPs) and scope constraints (qeqs)

some(x, big(x) $\wedge$ dog(x), every(y, cat(y), chase(e,x)))
h1=l2, h3=l6, h2=l5, h4=l4

every(y, cat(y), some(x, big(x) $\wedge$ dog(x), chase(e,x)))
h1=l2, h3=l6, h2=l4, h4=l1

MRS highlights (involving lots of people)

- Originally (VerbMobil): compositional semantics for typed feature structure grammars: flat structures good for composition and some linguistic phenomena (e.g., idioms). Principle: underspecified structures capturing syntax and morphology, further specification possible, convertible to logical formalisms.
- Algebra: constrained composition, guarantee of scopability (UTool for scoping).
- Other frameworks, other formalisms, other natural languages (DELPH-IN, Grammar Matrix).

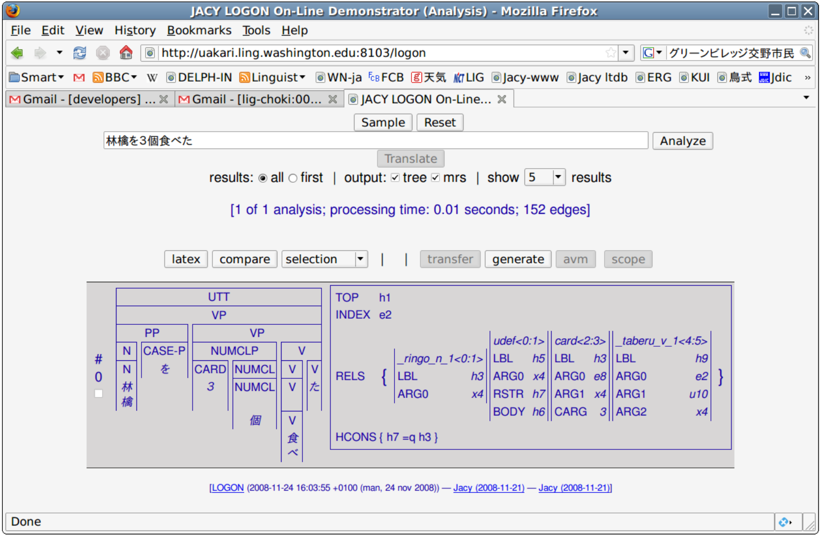
Further MRS highlights (even more people)

- Robust MRS (Deep Thought): more underspecification, composition as specialisation, semantics from shallower systems. *MRS as annotation.
- MT by semantic transfer (LOGON).
- Treebanks with MRSs (Redwoods, WeSearch, Wikiwoods).
- DMRS: dependency structures inter-convertible with MRS (example later): packing, incremental disambiguation (perhaps), graph-based SMT (in progress).
- MRS predicates connected to WordNets.
- *MRS-derived ontologies, distributions.
- Deeper semantics via (broadly) distributional techniques.

DELPH-IN: Deep Linguistic Processing using HPSG

- Informal collaboration on tools and grammars: see `http://www.delph-in.net/`
- Large grammars for English, German and Japanese; medium/growing for Spanish, Norwegian, Portuguese, Korean, French. Many small grammars.
- Common semantic framework: *MRS.
- Parsing and generation (realization), integrated shallower processing.
- Grammar Matrix: framework/starter kit for the development of grammars for diverse languages.

JACY example



JACY example

ringo	wo	3	ko	tabeta
apple	acc	3	classifier	eat past

[pro] ate three apples: default interpretation “I ate three apples”

l_3:_ringo_n_1(x_4),
l_5:udef(x_4, h_7, h_6),
l_3:card(e_8, x_4, 3),
l_9:_taberu_v_1($e_2\{\text{TENSE } past, \text{PROG -}, \text{PERF -}, \text{SF } prop\}$, u_{10}, x_4)
$h_7 =_q l_3$

Leading underscores: predicates correspond to lexeme.
No underscores: ‘grammar’ predicates (shared).
(Token/character positions not shown.)

A real example

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

A real example

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

modified quantifier

A real example

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

partitive

A real example

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

compound nominal

A real example

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

reduced relative

A real example

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

modified modifier

A real example

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

predeterminer

A real example

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}), $h_{48} =_q l_{49}$, $h_{42} =_q l_{43}$, $h_{32} =_q l_{37}$, $h_{18} =_q l_{20}$, $h_{13} =_q l_{14}$, $h_7 =_q l_3$

Compositional semantics in DELPH-IN: information that can be associated with syntax and morphology

- Fully identified (for English): predicate-argument structure (nouns, adjectives, verbs), modifier scope (e.g., *probably*), many multiword expressions, many constructions (e.g., relative clauses, appositives, tag questions, pseudo-partitives), . . .
- Partially identified/underspecified: quantifier scope, compound nouns, tense, aspect, massness, some sense extensions, some idioms . . .
- Mapping to WordNets (Bond et al).
- Many ongoing discussions: e.g., semi-compositional structures (cf Johan Bos examples yesterday).

Semantics and grammar engineering

- Ongoing extensive experimentation with details of the analysis (compare annotation).
- Highly empirical: working with real data for ongoing projects (simple examples for regression testing).
- Interactions can be complex: require implementation to investigate.
- Limitations of semantic literature:
  - base assumptions: ambiguity, lexical resources
  - sometimes ad-hoc or omitted syntax
  - few/no analyses: e.g., modification of quantifiers (*almost every*)
- Claim/hope: if we capture syntax/morphology, MRS can be a basis for deeper semantics.

RMRS

Split off most of EP's arguments: relate to predicate via anchor

MRS:
l1:some(x,h1,h2), h1 qeq l2,
l2:dog(x),
l3:chase(e,x,y),
l4:every(y,h3,h4), h3 qeq l65,
l5:cat(y)

RMRS:
l1:a1:some, BV(a1,x), RSTR(a1,h1), BODY(a1,h2), h1 qeq l2,
l2:a2:dog(x),
l3:a3:chase(e), ARG1(a3,x), ARG2(a3,y),
l4:a4:every, BV(a4,y), RSTR(a4,h3), BODY(a4,h4), h3 qeq l5,
l5:a5:cat(y)

Allows omission or underspecification of arguments.

RMRS

Split off most of EP's arguments: relate to predicate via anchor

MRS:
l1:some(x,h1,h2), h1 qeq l2,
l2:dog(x),
l3:chase(e,x,y),
l4:every(y,h3,h4), h3 qeq l65,
l5:cat(y)

RMRS:
l1:a1:some, BV(a1,x), RSTR(a1,h1), BODY(a1,h2), h1 qeq l2,
l2:a2:dog(x),
l3:a3:chase(e), ARG1(a3,x), ARG2(a3,y),
l4:a4:every, BV(a4,y), RSTR(a4,h3), BODY(a4,h4), h3 qeq l5,
l5:a5:cat(y)

Allows omission or underspecification of arguments.

RMRS

Split off most of EP's arguments: relate to predicate via anchor

MRS:
l1:some(x,h1,h2), h1 qeq l2,
l2:dog(x),
l3:chase(e,x,y),
l4:every(y,h3,h4), h3 qeq l65,
l5:cat(y)

RMRS:
l1:a1:some, BV(a1,x), RSTR(a1,h1), BODY(a1,h2), h1 qeq l2,
l2:a2:dog(x),
l3:a3:chase(e), ARG1(a3,x), ARG2(a3,y),
l4:a4:every, BV(a4,y), RSTR(a4,h3), BODY(a4,h4), h3 qeq l5,
l5:a5:cat(y)

Allows omission or underspecification of arguments.

RMRS

Split off most of EP's arguments: relate to predicate via anchor

MRS:
l1:some(x,h1,h2), h1 qeq l2,
l2:dog(x),
l3:chase(e,x,y),
l4:every(y,h3,h4), h3 qeq l65,
l5:cat(y)

RMRS:
l1:a1:some, BV(a1,x), RSTR(a1,h1), BODY(a1,h2), h1 qeq l2,
l2:a2:dog(x),
l3:a3:chase(e), ARG1(a3,x), ARG2(a3,y),
l4:a4:every, BV(a4,y), RSTR(a4,h3), BODY(a4,h4), h3 qeq l5,
l5:a5:cat(y)

Allows omission or underspecification of arguments.

RMRS

Split off most of EP's arguments: relate to predicate via anchor

MRS:
l1:some(x,h1,h2), h1 qeq l2,
l2:dog(x),
l3:chase(e,x,y),
l4:every(y,h3,h4), h3 qeq l65,
l5:cat(y)

RMRS:
l1:a1:some, BV(a1,x), RSTR(a1,h1), BODY(a1,h2), h1 qeq l2,
l2:a2:dog(x),
l3:a3:chase(e), ARG1(a3,x), ARG2(a3,y),
l4:a4:every, BV(a4,y), RSTR(a4,h3), BODY(a4,h4), h3 qeq l5,
l5:a5:cat(y)

Allows omission or underspecification of arguments.

RMRS

Split off most of EP's arguments: relate to predicate via anchor

MRS:
l1:some(x,h1,h2), h1 qeq l2,
l2:dog(x),
l3:chase(e,x,y),
l4:every(y,h3,h4), h3 qeq l65,
l5:cat(y)

RMRS:
l1:a1:some, BV(a1,x), RSTR(a1,h1), BODY(a1,h2), h1 qeq l2,
l2:a2:dog(x),
l3:a3:chase(e), ARG1(a3,x), ARG2(a3,y),
l4:a4:every, BV(a4,y), RSTR(a4,h3), BODY(a4,h4), h3 qeq l5,
l5:a5:cat(y)

Allows omission or underspecification of arguments.

Semantics via incremental annotation (RMRS)

Most cats noisily chased a large dog

most_DAT cat_NN2 noisily_RR chase_VVD a_AT1 large_JJ dog_NN1

l1:a1:most_q
l2:a2:cat_n(x2)
l3:a3:noisy(e3)
l4:a4:chase(e4)
l5:a5:a(x5)
l6:a6:large(e6)
l7:a7:dog(x7)

Semantics via incremental annotation (RMRS)

Most cats noisily chased a large dog

most_DAT cat_NN2 noisily_RR chase_VVD a_AT1 large_JJ dog_NN1

l1:a1:most_q	a1:BV(x2)	
l2:a2:cat_n(x2)		
l3:a3:noisy(e3)		
l4:a4:chase(e4)		
l5:a5:a(x5)	x5=x7	
l6:a6:large(e6)	a6:ARG1(x7)	l6=l7
l7:a7:dog(x7)		

Semantics via incremental annotation (RMRS)

Most cats noisily chased a large dog

most_DAT cat_NN2 noisily_RR chase_VVD a_AT1 large_JJ dog_NN1

l1:a1:most_q	a1:BV(x2)
l2:a2:cat_n(x2)	
l3:a3:noisy(e3)	l3=l4 e3=e4
l4:a4:chase(e4)	a4:ARG1(x2) a4:ARG2(x5)
l5:a5:a(x5)	x5=x7
l6:a6:large(e6)	a6:ARG1(x7) l6=l7
l7:a7:dog(x7)	

Semantics via incremental annotation (RMRS)

Most cats noisily chased a large dog

most_DAT cat_NN2 noisily_RR chase_VVD a_AT1 large_JJ dog_NN1

l1:a1:most_q	a1:BV(x2) a1:RSTR(h1) h1 qeq l2
l2:a2:cat_n(x2)	
l3:a3:noisy(e3)	l3=l4 e3=e4
l4:a4:chase(e4)	a4:ARG1(x2) a4:ARG2(x5)
l5:a5:a(x5)	x5=x7 a5:RSTR(h5) h5 qeq l6
l6:a6:large(e6)	a6:ARG1(x7) l6=l7
l7:a7:dog(x7)	

Semantics via incremental annotation (RMRS)

Most cats noisily chased a large dog

most_DAT cat_NN2 noisily_RR chase_VVD a_AT1 large_JJ dog_NN1

l1:a1:most_q	a1:BV(x2) a1:RSTR(h1) h1 qeq l2	a1:BODY(l5)
l2:a2:cat_n(x2)		
l3:a3:noisy(e3)	l3=l4 e3=e4	
l4:a4:chase(e4)	a4:ARG1(x2) a4:ARG2(x5)	
l5:a5:a(x5)	x5=x7 a5:RSTR(h5) h5 qeq l6	a1:BODY(l3)
l6:a6:large(e6)	a6:ARG1(x7) l6=l7	
l7:a7:dog(x7)		

Semantics via incremental annotation (RMRS)

Most cats noisily chased a large dog

most_DAT cat_NN2 noisily_RR chase_VVD a_AT1 large_JJ dog_NN1

l1:a1:most_q	a1:BV(x2) a1:RSTR(h1) h1 qeq l2	a1:BODY(l3)
l2:a2:cat_n(x2)		
l3:a3:noisy(e3)	l3=l4 e3=e4	
l4:a4:chase(e4)	a4:ARG1(x2) a4:ARG2(x5)	
l5:a5:a(x5)	x5=x7 a5:RSTR(h5) h5 qeq l6	a1:BODY(l1)
l6:a6:large(e6)	a6:ARG1(x7) l6=l7	
l7:a7:dog(x7)		

DMRS

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

DMRS

Some big angry dog barks loudly

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

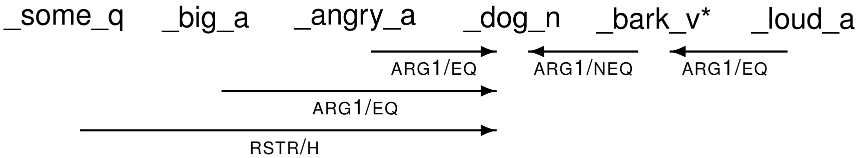
l1:a1:_some_q, BV(a1,x4), RSTR(a1,h5), BODY(a1,h6),
l2:a2:_big_a(e8), ARG1(a2,x4),
l2:a3:_angry_a(e9), ARG1(a3,x4),
l2:a4:_dog_n(x4), l4:a5:_bark_v(e2), ARG1(a5,x4),
l4:a6:_loud_a(e10), ARG1(a6,e2), h5 $=_q$ l2

DMRS

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



Dependency MRS (DMRS)

- predicates with simple inventory of links, no variables
- all information is retained so inter-convertible with MRS (without external information source)
- structure is minimal (no redundancy)
- applicable to different grammars, robust to changes in grammars

BUT: No direct logical interpretation.

Outline

Overview of MRS/RMRS/DMRS

Reflecting on MRS

Relating distributional and model-theoretic approaches

Big questions again

- To what extent are we (MRS developers) committed to a particular philosophical approach to semantics?
- Could computational approaches lead to advances in theoretical ideas about semantics?
- Maybe an ‘ordinary language’ / ‘meaning as use’ philosophical approach can only be tested computationally?

MRS and relationships between approaches

- Original MRS (and other approaches): relating logic to syntactic information.
- RMRS: tagging, chunking as underspecified semantics.
- DMRS: semantic and syntactic dependency representations.

Theory neutrality?

- Compositionality essential to all semantic approaches, including non-representational approaches (inferentialism, ordinary language philosophy etc).
- Some of (most of?) what MRS captures about language is therefore relevant for any approach.

Is MRS actually a semantic representation? Do we want it to be?

- MRS as ‘deep’ syntax? attempt to abstract away from language-specific idiosyncrasy?
- MRS competitive in SemEval tasks etc but hardly used in ‘proper’ inference (as opposed to pattern matching etc).
- DELPH-IN goal is complete coverage, for all text types: but generics? liar sentences? imperatives, greetings?
- Combining the different logics that have been proposed for different phenomena leads to a mess . . .
- So why insist on model theory? Current answer: model-theory is important in working out lexical/constructional structures.

Alternative philosophical underpinnings?

One candidate is inferentialism, specifically Brandom, ‘Making It Explicit’.

- Consistency with using human judgments of ‘good’ inferences rather than grounding a theory on ‘truth’.
- Consistency with taking ‘linguistic’ entities as primary.
- Logic still plays an important role.
- Full theory would necessarily integrate pragmatics (but corpora like Wikipedia involve limited speech acts).

Linguistic entities

- Only practical to 'ground' entities in limited domains:

  Very few of the Chinese construction companies consulted ...
- Lexical chains can't work without retaining lexical information:

  Der Bus ist das Zuhause der Band.
  Es ist sehr gemütlich.

  OR

  Er fährt nicht sehr schnell.
- Are there non-linguistic notions of entities?

Outline

Overview of MRS/RMRS/DMRS

Reflecting on MRS

Relating distributional and model-theoretic approaches

Lexicalised compositionality (LC)

- Joint work with Aurélie Herbelot.
- Lexicalised compositionality: part of an attempt to think about underpinnings and relationships
- Maybe also contributing to theory of distributional semantics.
- Maybe informative for practically integrating lexical semantics and MRS?

Distributional semantics

Distributional semantics: family of techniques for representing word meaning based on (linguistic) contexts of use.

it was authentic scrumpy, rather sharp and very strong

we could taste a famous local product — scrumpy

spending hours in the pub drinking scrumpy

Distributional semantics

Distributional semantics: family of techniques for representing word meaning based on (linguistic) contexts of use.

it was authentic scrumpy, rather sharp and very strong

we could taste a famous local product — scrumpy

spending hours in the pub drinking scrumpy

Distributional semantics

Distributional semantics: family of techniques for representing word meaning based on (linguistic) contexts of use.

it was authentic scrumpy, rather sharp and very strong

we could taste a famous local product — scrumpy

spending hours in the pub drinking scrumpy

Distributional semantics: the intuitions

- Humans typically learn word meanings (concepts) from context: sometimes perceptually grounded, sometimes not.
- Possibly processed to some different representation, but perhaps mental representation directly reflects context?
- Distributional semantics uses linguistic context to represent meaning (partially).
- Meaning seen as a space, with dimensions corresponding to elements in the context (features).
- Computational techniques generally use vectors (semantic space models, vector space models).

LC main ideas

- Combining compositional and distributional techniques, using 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.
- Distributions from contexts expressed in MRS.
- Ideal distributions: everything someone **could** say (contrasted with 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):

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

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

Ideal distribution for S_1

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

Ideal distribution for S_1, continued

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

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

$$
\begin{aligned}
\text{white}^\circ = \quad \{ \quad & < [x4][a(x4), \text{cube}^\circ(x4), \text{rotate}^\circ(e4, x4)], S_1 >, \\
& < [x8][a(x8), \text{object}^\circ(x8), \text{rotate}^\circ(e8, x8)], S_1 > \}
\end{aligned}
$$

Relationship to standard notion of extension

For a predicate P, the distributional arguments of P$^\circ$ 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.

Back to the beginning?

Ludwig Wittgenstein:
Philosophical Investigations
(i.e., later work).

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

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