Semantics in broad-coverage natural language processing

Ann Copestake

Computer Laboratory
University of Cambridge

October 2006

Broad coverage semantics.

Example

Is there any water in the refrigerator?
lb1:int_m(e2,lb5), lb5:prpstn_m(e2,h6), lb7:be_v_there(e2,x8), lb9:any_q(x8,h10,h11), lb12:water_n_1(x8), lb12:in_p(e13,x8,x14), lb15:the_q(x14,h17,h16), lb18:refrigerator_n_1(x14), h6 $=_q$ lb7, h10 $=_q$ lb12, h17 $=_q$ lb18

Aims:

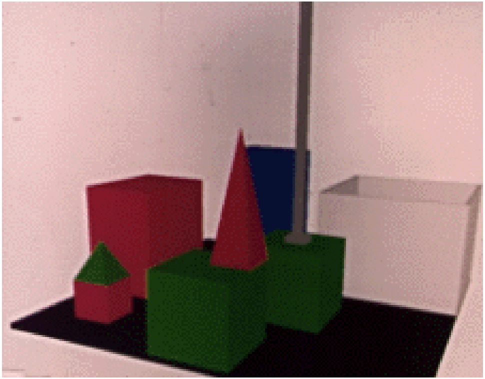
- Build systems to analyse any text to produce a meaning representation (and to generate text from meaning representations).
- Exploit these systems in applications.
- Find out interesting things about language.

Outline.

1. Semantics in computational linguistics.
   - Already done?
   - Or impossible?
   - Objectives.
2. Technology for semantic representation.
   - Parsing technology.
   - Compositional semantic representation.
   - Underspecification.
3. Applications.
   - eScience applications.
   - Some other applications.

Outline.

SHRDLU (Winograd, 1971).



Person: PICK UP A BIG RED BLOCK.
Computer: OK. (does it)
Person: WHAT DOES THE BOX CONTAIN?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

More about SHRDLU.

- A ‘micro-world’: closed domain with small number of objects.
- Impressive demos with under 100 words.
- World state used to resolve linguistic ambiguity etc, but planning of actions and the rules for behaviour of the objects were independent of the language analysis.
- Classic demonstration of ‘strong’ AI.
- Unfortunately, this did not scale up . . .

More about SHRDLU.

- A ‘micro-world’: closed domain with small number of objects.
- Impressive demos with under 100 words.
- World state used to resolve linguistic ambiguity etc, but planning of actions and the rules for behaviour of the objects were independent of the language analysis.
- Classic demonstration of ‘strong’ AI.
- Unfortunately, this did not scale up . . .

Outline.

1. Semantics in computational linguistics.
   - Already done?
   - Or impossible?
   - Objectives.
2. Technology for semantic representation.
   - Parsing technology.
   - Compositional semantic representation.
   - Underspecification.
3. Applications.
   - eScience applications.
   - Some other applications.

Winograd and Flores (1986): Understanding Computers and Cognition.

Example

A: Is there any water in the refrigerator?
B: Yes
A: Where? I don't see it.
B: In the cells of the eggplant.

If A's utterance meant:

$$\text{ynq}(\exists x[\iota y[\text{water}'(x) \wedge \text{fridge}'(y) \wedge \text{in}'(e, x, y) \wedge \text{time}(e) = \text{now}]])$$

then B's response is truthful.

Objective reality and hermeneutics.

Every speech act occurs in a context, with a background shared by speaker and hearer. The 'felicity conditions' depend on mutual knowledge and intentions.

Faced with a problem in representing the contents of admissions folders, the right questions are neither realist ("What is a GPA, really?") nor cognitive ("What is in the concept of GPA?") but conversational ("What is the structure of the discourse in which the distinction 'GPA' emerges?").

Nothing exists except through language. (originally due to Gadamer)

Overstatement of some problems.

- Written communication is possible, despite impoverished context, no interactivity.
- Shared conventions of meaning are required for successful language use.
- There is some meaning independent of individual discourses.
- Negotiation of meaning is usually selection between existing possibilities and fine-tuning.

Implicit underestimate of other problems.

Reliably going from real utterances/sentences to compositional meaning representations is hard.

- Coverage of grammars and lexicons (and behaviour when coverage is lacking).
- Working out plausible semantic representations (without being tied to English).
- Ambiguity (search space, efficiency, number of semantic representations).
- Evaluations not demos!

Dialogue with autonomous agents is not the only application of computational linguistics (or even a major one, nowadays).

Outline.

Objectives for computational semantics.

- Construct compositional semantics from arbitrary text (i.e., use the information that comes from syntax and morphology to give a logical representation)
- and generate utterances from semantic representations.
- Show utility in applications where:
    - the context is relatively conventional/stable;
    - full understanding is not required;
    - BUT without using toy domains.
- Provide semantics for predicates via:
    - links to ontologies (e.g., *water* is H_2O);
    - task-specific encodings (e.g., *aim*, *goal* etc in scientific text used as cues for extraction);
    - (longer-term) acquisition from corpora (bootstrapping from compositional semantics)

Outline.

1. Semantics in computational linguistics.
   - Already done?
   - Or impossible?
   - Objectives.
2. Technology for semantic representation.
   - Parsing technology.
   - Compositional semantic representation.
   - Underspecification.
3. Applications.
   - eScience applications.
   - Some other applications.

Parsing since SHRDLU.

- Grammars can be directly based on linguistic theories and declarative: easier to modify and maintain, usable with different parsers.
- Coverage for English has increased, grammars available for many other languages as well.
- Work on learning grammars automatically (but mostly from hand-annotated text).
- Statistical techniques for parse selection (both hand-built and manually created grammars).
- Avoid representation of real world knowledge: either very limited domains or very limited inference.

Different ‘depths’ of analysis

- part-of-speech tagging (e.g., Elworthy POS tagger)
- chunking
- grammars without lexicons (e.g., RASP parser, Briscoe and Carroll)
- detailed grammars that can be used for generation as well as parsing (e.g., resources from DELPH-IN Open Source collaboration, Flickinger, Oepen, Copestake, Carroll, Bender et al)

  `http://www.delph-in.net/erg/`

Outline.

1. Semantics in computational linguistics.
   - Already done?
   - Or impossible?
   - Objectives.
2. Technology for semantic representation.
   - Parsing technology.
   - Compositional semantic representation.
   - Underspecification.
3. Applications.
   - eScience applications.
   - Some other applications.

General ideas.

- Compositional semantics is driven by syntax (traditionally FOPC via lambda calculus).
- Alternatives to FOPC include Minimal Recursion Semantics (MRS) and Robust MRS.

Is there any water in the refrigerator?

lb1:int_m(e2,lb5), lb5:prpstn_m(e2,h6), lb7:be_v_there(e2,x8),
lb9:any_q(x8,h10,h11), lb12:water_n_1(x8),
lb12:in_p(e13,x8,x14), lb15:the_q(x14,h17,h16),
lb18:refrigerator_n_1(x14),
h6 $=_q$ lb7, h10 $=_q$ lb12, h17 $=_q$ lb18

Flattening: representation of conjunction.

- conjunction is used to represent modification by (most) adjectives and adverbs, prepositional phrases etc
  $\wedge(\wedge(\wedge(\text{huge}'(x), \text{ugly}'(x)), \text{grey}'(x)), \text{house}'(x))$
- suppose 'huge house' corresponds to 'mansion'
  $\wedge(\text{huge}'(x), \text{house}'(x)) \mapsto \text{mansion}'(x)$
  matching involves unpacking binary conjunction tree
- but why not use n-ary conjunction?
  $\wedge(\text{huge}'(x), \text{ugly}'(x), \text{grey}'(x), \text{house}'(x))$
- or let a list indicate conjunction and use a canonical ordering?
  $(\text{grey}'(x), \text{house}'(x), \text{huge}'(x), \text{ugly}'(x))$

Flattening: representation of conjunction.

- conjunction is used to represent modification by (most) adjectives and adverbs, prepositional phrases etc
  $\wedge(\wedge(\wedge(\text{huge}'(x), \text{ugly}'(x)), \text{grey}'(x)), \text{house}'(x))$
- suppose ‘huge house’ corresponds to ‘mansion’
  $\wedge(\text{huge}'(x), \text{house}'(x)) \mapsto \text{mansion}'(x)$
  matching involves unpacking binary conjunction tree
- but why not use n-ary conjunction?
  $\wedge(\text{huge}'(x), \text{ugly}'(x), \text{grey}'(x), \text{house}'(x))$
- or let a list indicate conjunction and use a canonical ordering?
  $(\text{grey}'(x), \text{house}'(x), \text{huge}'(x), \text{ugly}'(x))$

Flattening: representation of conjunction.

- conjunction is used to represent modification by (most) adjectives and adverbs, prepositional phrases etc
$\wedge(\wedge(\wedge(\text{huge}'(x), \text{ugly}'(x)), \text{grey}'(x)), \text{house}'(x))$
- suppose 'huge house' corresponds to 'mansion'
$\wedge(\text{huge}'(x), \text{house}'(x)) \mapsto \text{mansion}'(x)$
matching involves unpacking binary conjunction tree
- but why not use n-ary conjunction?
$\wedge(\text{huge}'(x), \text{ugly}'(x), \text{grey}'(x), \text{house}'(x))$
- or let a list indicate conjunction and use a canonical ordering?
$(\text{grey}'(x), \text{house}'(x), \text{huge}'(x), \text{ugly}'(x))$

Outline.

1. Semantics in computational linguistics.
   - Already done?
   - Or impossible?
   - Objectives.
2. Technology for semantic representation.
   - Parsing technology.
   - Compositional semantic representation.
   - Underspecification.
3. Applications.
   - eScience applications.
   - Some other applications.

Underspecification and Sudoku solving

			7					8
		9					2	
	5			3			9	
8					2			
		6				7		
			4					1
	3			9			6	
	2					4		
7					1			

Solving.

			7					8
		9					2	
	5			3			9	
8					2			
		6				7		
			4					1
	3			9			6	
	2					4		
7					1			

Possibility 1.

			7					8
		9					2	7
	5			3			9	
8					2			
		6				7		
			4					1
	3			9			6	
	2					4		
7					1			

Possibility 2.

			7					8
		9					2	
	5			3			9	7
8					2			
		6				7		
			4					1
	3			9			6	
	2					4		
7					1			

Underspecification.

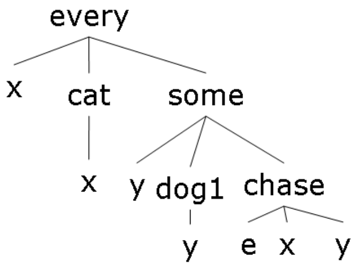
			7					8
		9					2	7
	5			3			9	7
8					2			
		6				7		
			4					1
	3			9			6	
	2					4		
7					1			

Inference on underspecified form.

			7					8
		9					2	7
	5			3			9	7
8					2			
		6				7		
			4					1
	3			9			6	
	2					4	7	
7					1			

Logical representations as trees.

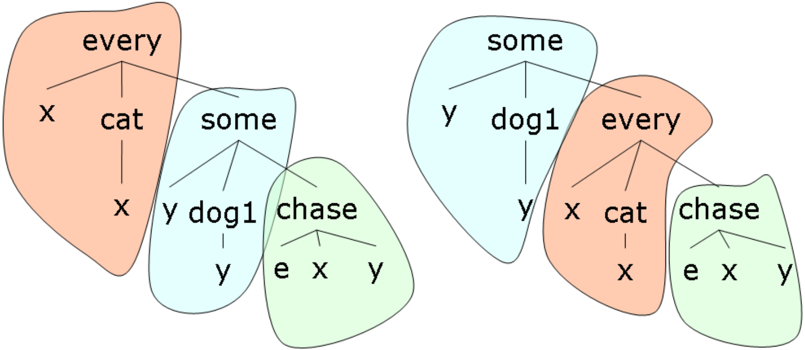
every cat chased some dog



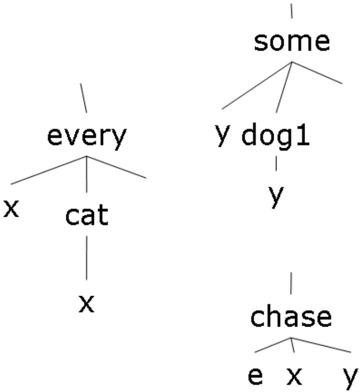
every(x,cat(x),some(y,dog1(y),chase(e,x,y)))

some(y,dog1(y),every(x,cat(x),chase(e,x,y)))

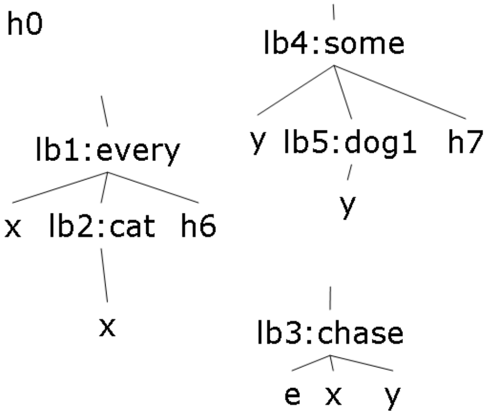
Structure sharing between trees.



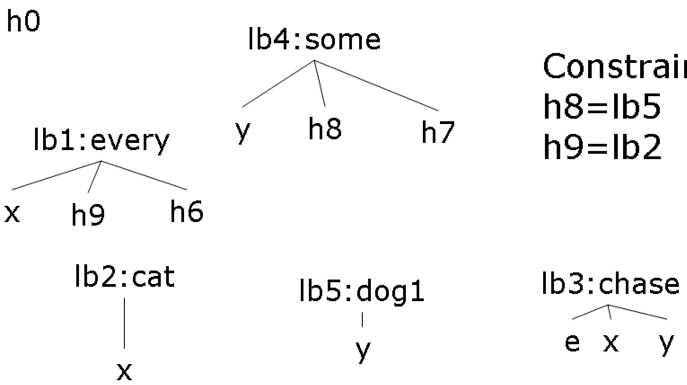
Tree fragments.



Tree fragments with labels.



Maximum splitting.



Constraints:
h8=lb5
h9=lb2

Underspecification and flattening in MRS.

Every cat chased some dog

lb0:prpstn_m(e,h1), lb1:every_q(x,h9,h6),lb2:cat_n(x),
lb4:some_q(y,h8,h7), lb5:dog_n_1(y),lb3:chase_v(e,x,y),
h1 $=_q$ lb3, h8 $=_q$ lb5, h9 $=_q$ lb2

Is there any water in the refrigerator?

lb1:int_m(e2,lb5), lb5:prpstn_m(e2,h6), lb7:be_v_there(e2,x8),
lb9:any_q(x8,h10,h11), lb12:water_n_1(x8),
lb12:in_p(e13,x8,x14), lb15:the_q(x14,h17,h16),
lb18:refrigerator_n_1(x14),
h6 $=_q$ lb7, h10 $=_q$ lb12, h17 $=_q$ lb18

Arguments without lexicons.

- Robust syntactic processing can proceed without a detailed lexicon. e.g., *chase* is a verb, but transitivity?
- But arity of predicates correlates with transitivity.

Splitting off arguments

lb0:prpstn_m(e,h1), lb1:every_q(x), lb1:RSTR(h9), lb1:BODY(h6), lb2:cat_n(x), lb4:some_q(y), lb4:RSTR(h8), lb4:BODY(h7), lb5:dog_n_1(y), lb3:chase_v(e), lb3:ARG1(x), lb3:ARG2(y), h1 $=_q$ lb3, h8 $=_q$ lb5, h9 $=_q$ lb2

- Verb POS tag gives lb3:LEXEME_v(e) — ARG1 and ARG2 added if licensed by syntax.

Arguments without lexicons.

- Robust syntactic processing can proceed without a detailed lexicon. e.g., *chase* is a verb, but transitivity?
- But arity of predicates correlates with transitivity.

Splitting off arguments

lb0:prpstn_m(e,h1), lb1:every_q(x), lb1:RSTR(h9), lb1:BODY(h6), lb2:cat_n(x), lb4:some_q(y), lb4:RSTR(h8), lb4:BODY(h7), lb5:dog_n_1(y), lb3:chase_v(e), lb3:ARG1(x), lb3:ARG2(y), h1 $=_q$ lb3, h8 $=_q$ lb5, h9 $=_q$ lb2

- Verb POS tag gives lb3:LEXEME_v(e) — ARG1 and ARG2 added if licensed by syntax.

Integrating processing.

Shallow processing representations are underspecified compared to deep processing.

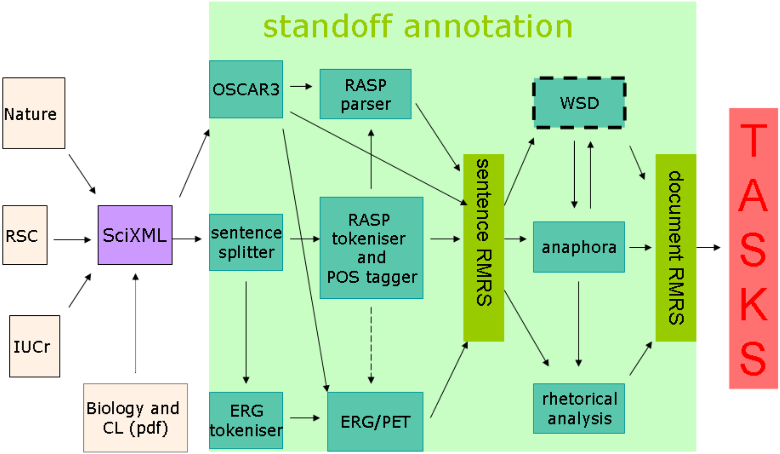
lb1:every_q(x),	lb1:every_q(x),
lb1:RSTR(h9),	
lb1:BODY(h6),	
lb2:cat_n(x),	lb2:cat_n(x),
lb4:some_q(y),	lb4:some_q(y),
lb1:RSTR(h8),	
lb1:BODY(h7),	
lb5:dog_n_1(y),	lb5:dog_n(y),
lb3:chase_v(e),	lb3:chase_v(e)
lb3:ARG1(x),	
lb3:ARG2(y)	

Integrating processing.

Shallow processing representations are underspecified compared to deep processing.

lb1:every_q(x),	lb1:every_q(x),
lb1:RSTR(h9),	
lb1:BODY(h6),	
lb2:cat_n(x),	lb2:cat_n(x),
lb4:some_q(y),	lb4:some_q(y),
lb1:RSTR(h8),	
lb1:BODY(h7),	
lb5:dog_n_1(y),	lb5:dog_n(y),
lb3:chase_v(e),	lb3:chase_v(e)
lb3:ARG1(x),	
lb3:ARG2(y)	

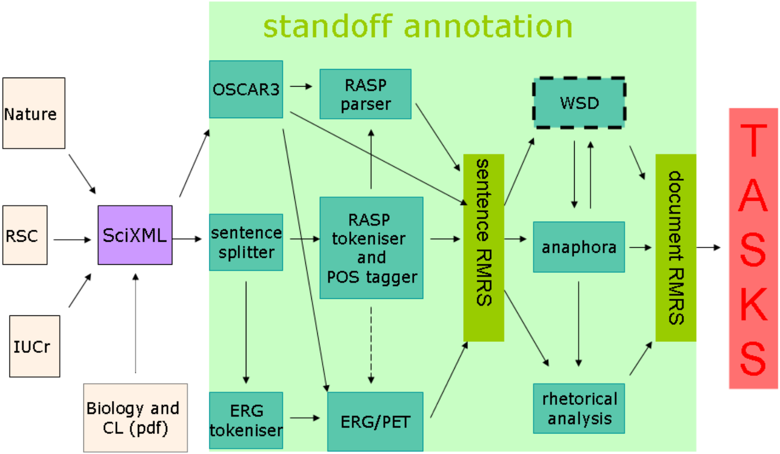
Combined processing architecture (SciBorg).



Outline.

1. Semantics in computational linguistics.
   - Already done?
   - Or impossible?
   - Objectives.
2. Technology for semantic representation.
   - Parsing technology.
   - Compositional semantic representation.
   - Underspecification.
3. Applications.
   - eScience applications.
   - Some other applications.

Extracting the science from scientific publications.



Searches on Chemistry Papers.

Papers about synthesis of Tröger's base from anilines:

Paper 1: The synthesis of 2,8-dimethyl-6H,12H-5,11 methanodibenzo[b,f][1,5]diazocine (Troger's base) from p-toluidine and of two Troger's base analogs from other anilines

Paper 2: . . . Tröger's base (TB) . . . The TBs are usually prepared from para-substituted anilines

Variation in expression.

- linguistic variation and syntactic relationship: synthesis of X, synthesize X, prepare X ...
- coreference
- chemistry names
- ontological relationships

Could expand out query terms, but how to search for papers describing Tröger's base syntheses which don't involve anilines?

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs 'attacking Vogons are annoying' and 'spelling contests are boring')
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs 'the attack of the Vogons')
- aspirin's synthesis (noun+pos noun) (vs 'the Vogons' attack')
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs 'attacking Vogons are annoying' and 'spelling contests are boring')
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun)
  (vs 'the attack of the Vogons')
- aspirin's synthesis (noun+pos noun)
  (vs 'the Vogons' attack')
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs 'attacking Vogons are annoying' and 'spelling contests are boring')
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs 'the attack of the Vogons')
- aspirin's synthesis (noun+pos noun) (vs 'the Vogons' attack')
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs ‘attacking Vogons are annoying’ and ‘spelling contests are boring’)
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs ‘the attack of the Vogons’)
- aspirin’s synthesis (noun+pos noun) (vs ‘the Vogons’ attack’)
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs ‘attacking Vogons are annoying’ and ‘spelling contests are boring’)
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs ‘the attack of the Vogons’)
- aspirin’s synthesis (noun+pos noun) (vs ‘the Vogons’ attack’)
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs ‘attacking Vogons are annoying’ and ‘spelling contests are boring’)
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs ‘the attack of the Vogons’)
- aspirin’s synthesis (noun+pos noun) (vs ‘the Vogons’ attack’)
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs 'attacking Vogons are annoying' and 'spelling contests are boring')
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs 'the attack of the Vogons')
- aspirin's synthesis (noun+pos noun) (vs 'the Vogons' attack')
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs ‘attacking Vogons are annoying’ and ‘spelling contests are boring’)
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun)
(vs ‘the attack of the Vogons’)
- aspirin’s synthesis (noun+pos noun)
(vs ‘the Vogons’ attack’)
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs ‘attacking Vogons are annoying’ and ‘spelling contests are boring’)
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs ‘the attack of the Vogons’)
- aspirin’s synthesis (noun+pos noun) (vs ‘the Vogons’ attack’)
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs ‘attacking Vogons are annoying’ and ‘spelling contests are boring’)
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun)
(vs ‘the attack of the Vogons’)
- aspirin’s synthesis (noun+pos noun)
(vs ‘the Vogons’ attack’)
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

Syntactic variability.

- Hoffman synthesized/synthesised aspirin (verb+ed NP)
- aspirin was synthesised by Hoffman (NP be verb+ed)
- synthesising aspirin is easy (verb+ing NP) (vs 'attacking Vogons are annoying' and 'spelling contests are boring')
- the synthesised aspirin (verb+ed/adj noun)
- the synthesis of aspirin (noun of noun) (vs 'the attack of the Vogons')
- aspirin's synthesis (noun+pos noun) (vs 'the Vogons' attack')
- aspirin synthesis (noun noun)

Common semantic pattern (ideally):
lb1:synthesise(e), lb1:ARG2(y), lb3:aspirin(y)

AZ (Simone Teufel) in SciBorg

Synthesis of pyrazole and pyrimidine Tröger's base analogues

Rodrigo Abonia, Andrea Albornoz, Hector Larrahondo, Jairo Quiroga, Braulio Insuasty, Henry Insuasty, Angelina Hormaza, Adolfo Sánchez, Manuel Nogueras

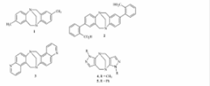
Perkin 1

Tröger's-base analogues bearing fused pyrazolic or pyrimidinic rings were prepared in acceptable to good yields through the reaction of 3-alkyl-5-amino-1-arylpyrazoles and 6-aminopyrimidin-4(3*H*)-ones with formaldehyde under mild conditions (*i.e.*, in ethanol at 50 °C in the presence of catalytic amounts of acetic acid).
Two key intermediates were isolated from the reaction mixtures, which helped us to suggest a sequence of steps for the formation of the Tröger's bases obtained. The structures of the products were assigned by **1** H and **13** C NMR, mass spectra and elemental analysis and confirmed by X-ray diffraction for one of the obtained compounds.

Introduction

Although the first Tröger's base **1** was obtained more than a century ago from the reaction of p-toluidine and formaldehyde,[1] recently the study of these compounds has gained importance due to their potential applications. They posses a relatively rigid chiral structure which makes them suitable for the development of possible synthetic enzyme and artificial receptor systems, [2] chelating and biomimetic systems, [3] and transition metal complexes for regio- and stereoselective catalytic reactions.[4]

For these reasons, numerous Tröger's-base derivatives have been prepared bearing different types of substituents and structures (*i.e.*, **2–5** Scheme 1), with the purpose of



Scheme 1 The original Tröger's base 1 and some interesting derivatives and analogues.

However, some of the above methodologies possess tedious work-up procedures or include relatively strong reaction conditions, such as treatment of the starting materials for several hours with an ethanolic solution of conc. hydrochloric acid or TFA solution, with poor to moderate yields, as is the case for analogues **4** and **5.**

Considering these potential applications, we now report a simple synthetic method for the preparation of 5,12-dialkyl-3,10-diaryl-1,3,4,8,10,11-hexaazatetracyclo[6.6.1.0 2,6 .0 9,13]pentadeca-2(6),4,9(13),11-tetraenes **8a–e** and 4,12-dimethoxy-1,3,5,9,11,13-hexaazatetracyclo[7.7.1.0 2,7.0 10,15heptadeca-2(7),3,10(15),11-tetraene-6,14-diones **10a,b** based on the reaction of 3-alkyl-5-amino-1-arylpyrazoles **6** and 6-aminopyrimidin-4(3*H*)-ones **9** with formaldehyde in ethanol and catalytic amounts of acetic acid. Compounds **8** and **10** are new Tröger's-base analogues bearing heterocyclic rings instead of the usual phenyl rings in their aromatic parts.

Results and discussion

In an attempt to prepare the benzotriazolyl derivative **7a**, which could be used as an intermediate in the synthesis of new hydroquinoline analogues of interest,[6] a mixture of 5-amino-3-methyl-1-phenylpyrazole **6a**, formaldehyde and benzotriazole in 10 mL of ethanol, with catalytic amounts of acetic acid, was heated at 50 °C for 5 minutes. A solid precipitated from the solution while it was still hot. However, no consumption of benzotriazole was observed by TLC.

The reaction conditions were modified and the same product was obtained when the reaction was carried out without using benzotriazole, as shown in Chart 1. On the basis of NMR and mass spectra and X-ray crystallographic analysis we established that the structure of this compound is 5,12-dimethyl-3,10-diphenyl-1,3,4,8,10,11-hexaazatetracyclo[6.6.1.0 2,6.0 9,13] pentadeca-2(6),4,9(13),11-tetraene **8a**, a new pentagonal Tröger's-

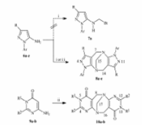


Chart 1 Reaction of 5-aminopyrazoles 6 and 6-aminopyrimidin-4(3H)-ones 9 with formaldehyde.

1588 *J.Chem. Soc, Perkin Trans. 1, 2002, 1588-1591*
DOI: 10.1039/b200862a

Legenda:
- Background
- Other
- Own
- Based
- Contrast
- Textual
- Aim

Identifying cues

> *The primary aims of the present study are (i) the synthesis of an amino acid derivative that can be incorporated into proteins via standard solid-phase synthesis methods, and (ii) a test of the ability of the derivative to function as a photoswitch in a biological environment.*

Specify cues in RMRS:

`lb1:objective(x), ARG1(lb1,y), lb2:research(y)`

`objective` generalises the predicates for *aim*, *goal* etc and `research` generalises *study, work* etc. (i.e., ontology for rhetorical structure).

SciBorg objective: extended information extraction

Searching for papers describing Tröger's base syntheses which don't involve anilines.

```
retrieve all papers X:
PAPER-AIM(X,lb1), lb1:synthesis, lb1:SYN-RESULT(<TB
lb1:SYN-SOURCE(y), NOT(aniline(y))
```

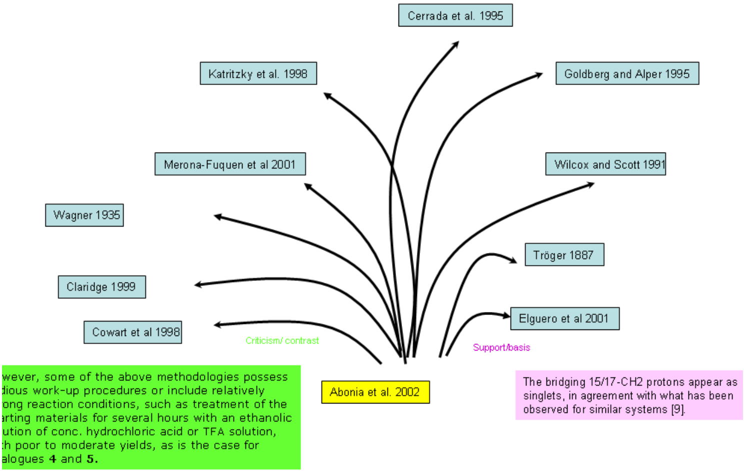
where `<TB>` relates to some precise chemistry structure (represented in CML), `SYN-RESULT` and `SYN-SOURCE` are specific to Chemistry syntheses.

Citation type classification using RMRS: Siddharthan and Teufel, 2006

As we are using the conceptual graph formalism to represent our definitions, we can use the graph matching operations defined in Sowa (1984).

- Matches: lb1:use(e), ARG1(lb1,**authors**), ARG2(lb2,**citation**)
- Clue to classifying this citation as USE (vs CONTRAST etc)

Citation maps



Outline.

Other applications using MRS/RMRS

- Reasoning about meetings (Schlangen et al, 2003)
  A. Can we meet next Monday? B. How about Tuesday?
- Machine Translation using semantic transfer (Verbmobil, LOGON, Japanese-English open source)
- Ontology extraction from dictionaries (NTT)
  doraiba: jidosha wo unten suru hito
  driver: a person who drives a car
- Ontology extraction from Wikipedia (Aurelie Herbelot)
- Email response (YY Software†)
- Question answering (QUETAL project)
- IE and sentiment classification (Deep Thought)

Why B should have said no.

Universal quantification is always over a contextually salient set.
There is no water in the fridge.

$$\forall x[\text{water}'(x) \implies \neg\text{in}'(x, \text{the-fridge})$$

Not all water, but all water in a contextually salient class:

$$\forall x[\text{water}'(x) \wedge \text{SALIENT}_c(\text{water}')(x) \implies \neg\text{in}'(x, \text{the-fridge})$$

A and B might agree perfectly on water' but still have misunderstanding due to different assumptions about contextual salience. But, no direct information about SALIENT_c, so no point putting it in the compositional representation.

Why B should have said no.

Universal quantification is always over a contextually salient set. There is no water in the fridge.

$$\forall x[\text{water}'(x) \implies \neg\text{in}'(x, \text{the-fridge})$$

Not all water, but all water in a contextually salient class:

$$\forall x[\text{water}'(x) \wedge \text{SALIENT}_c(\text{water}')(x) \implies \neg\text{in}'(x, \text{the-fridge})$$

A and B might agree perfectly on water' but still have misunderstanding due to different assumptions about contextual salience. But, no direct information about SALIENT_c, so no point putting it in the compositional representation.

Summary

- Broad coverage compositional semantics is feasible and a useful basis for applications.
- Ongoing improvements in representation technology, coverage, depth of analysis, efficiency and accuracy.
- Prerequisite for ‘real’ natural language understanding (if and when . . .)

Summary

- Broad coverage compositional semantics is feasible and a useful basis for applications.
- Ongoing improvements in representation technology, coverage, depth of analysis, efficiency and accuracy.
- Prerequisite for ‘real’ natural language understanding (if and when . . .)

Credits

Dan Flickinger, Simone Teufel, CJ Rupp, Ben Waldron, Advaith Siddharthan, Peter Corbett, Peter Murray-Rust, Ted Briscoe, John Carroll, Ivan Sag, Carl Pollard, Anette Frank, Alex Lascarides, David Schlangen, Stephan Oepen, Emily Bender, Rob Malouf, Francis Bond, Tim Baldwin, Aline Villavicencio, Melanie Siegel, Lars Hellan, Dorothee Beerman, Ulrich Callmeier, Ulrich Schäfer, Bernd Kiefer, Guido Minen, Victor Poznanski, Susanne Riehemann

with funding from BMBF, CSLI IAP, NSF, EPSRC, NTT, European Commission and Boeing.