# Compound interpretation as a challenge for computational semantics


Diarmuid Ó Séaghdha



ComAComA, Dublin
24 August 2014



UNIVERSITY OF CAMBRIDGE

# Introduction

- Noun-noun compounding is very common in many languages
- We can make new words out of old
- Expanding vocabulary $\rightarrow$ lots of OOV problems!
- Compounding compresses information about semantic relations
- Decompressing this information ("interpretation") is a non-trivial task
- In this talk I focus on relational understanding

# Compound interpretation as semantic relation prediction

The hut is located in the mountains
The hut is constructed out of timber
The camp produces timber

# Compound interpretation as semantic relation prediction

| | |
|---|---|
| The hut is located in the mountains | LOCATION |
| The hut is constructed out of timber | MATERIAL |
| The camp produces timber | LOCATION/PRODUCER |

# Compound interpretation as semantic relation prediction

| | |
|---|---|
| The hut is located in the mountains | LOCATION |
| The hut is constructed out of timber | MATERIAL |
| The camp produces timber | LOCATION/PRODUCER |

We slept in a mountain hut
We slept in a timber hut
We slept in a timber camp

# Compound interpretation as semantic relation prediction

| | |
|---|---|
| The hut is located in the mountains | LOCATION |
| The hut is constructed out of timber | MATERIAL |
| The camp produces timber | LOCATION/PRODUCER |

We slept in a mountain hut
We slept in a timber hut
We slept in a timber camp

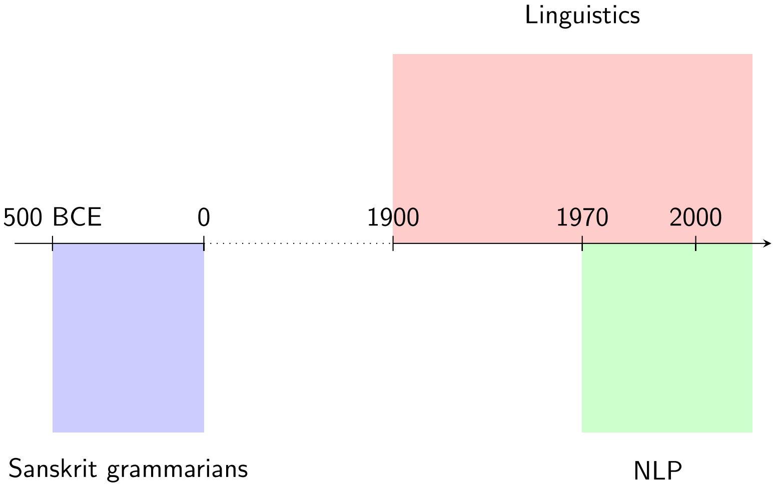
??

# Why compounds?

- Special but very frequent case of information extraction
- In order to interpret compounds, a system must be able to deal with:
  - Lexical semantics
  - Relational semantics
  - Implicit information
  - World knowledge
  - Handling sparsity
- Compound interpretation is an excellent testbed for computational semantics.

# Thoughts and open questions

# A brief history of compound semantics

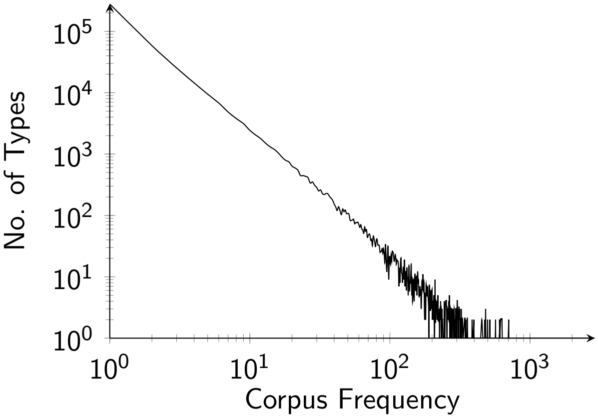


# Open questions

- ... almost all questions are still open!
- Some questions that I am interested in:
  - What are useful representations for compound semantics?
  - What are learnable representations for compound semantics?
  - Should we use representations that are not specific to compounds?
  - What are the applications of compound interpretation?
    - Paraphrasing/lexical expansion (for MT, search,...)
    - Machine reading/natural language understanding
- Many representation options, some more popular than others
- All have pros and cons

# The lexical analysis

- **Idea:** Treat compounds as if they were words.
- Frequent/idiomatic compounds (e.g., in WordNet)
- **Pro:** Flexible
- **Con:** Productivity



# The "pro-verb" analysis

- **Idea:** Underspecified single relation for all compounds
- Adequate when parsing to logical form or e.g. Minimal Recursion Semantics:

| | |
|---|---|
| *car tyre* | `compound_nn_rel(car,tyre)` |
| *history book* | `compound_nn_rel(history,book)` |

- **Pro:** Easy to integrate with parsing/structured prediction
- **Con:** Not very expressive!

# The inventory analysis

- **Idea:** Select a relation label from a (small) set of candidates

| | |
|---|---|
| *car tyre* | PART-WHOLE |
| *mountain hut* | LOCATION |
| *cheese knife* | PURPOSE |
| *headache pill* | PURPOSE |

- Earliest, most common approach [Su, 1969; Russell, 1972; Nastase and Szpakowicz, 2003; Girju et al., 2005; Tratz and Hovy, 2010]
- Some relation extraction datasets span compounds and other constructions [Hendrickx et al., 2010]
- **Pro:** Learnable as multiclass classification; annotation is feasible
- **Con:** Conflates subtleties (*sleeping pill* vs *headache pill*); requires annotated training data

# The vector analysis

- **Idea:** Represent a compound by composing vectors for each constituent to produce a new vector
- Lots of work on vector composition; some work on noun-noun composition [Mitchell and Lapata, 2010; Reddy et al., 2011; Ó Séaghdha and Korhonen, 2014]
- **Pro:** Learnable from unlabelled data
- **Con:** Difficult to interpret

# The paraphrase analysis

- **Idea:** Represent the implicit relation(s) with a distribution over explicit paraphrases.
- Allowable paraphrases can use prepositions [Lauer, 1995], verbs [Nakov, 2008; Butnariu et al., 2010], free paraphrases [Hendrickx et al., 2013]

```
virus that causes flu         38
virus that spreads flu        13
virus that creates flu         6
virus that gives flu           5
...
virus that is made up of flu  1
virus that is observed in flu 1
```

- Suitable for similarity, data expansion
- **Pro:** Learnable from unannotated text
- **Con:** Paraphrases can be ambiguous/synonymous

# The frame analysis

- We could recover implicit relational structure in terms of FrameNet-like frames:

| | |
|---|---|
| *cheese knife* | `Cutting(f)` $\wedge$ `Instrument(f,knife)` $\wedge$ `Item(f,cheese)` |
| *kitchen knife* | `Cutting(f)` $\wedge$ `Instrument(f,knife)` $\wedge$ `Place(f,kitchen)` |
| *student demonstration* | `Protest(f)` $\wedge$ `Protestor(f,student)` |
| *headache pill* | `Cure(f)` $\wedge$ `Affliction(f,headache)` $\wedge$ `Medication(f,pill)` |

- Connection to cognitive/frame semantics [Ryder, 1994; Coulson, 2001]
- SRL usually assumes explicit verbal predicates or nominalisations
- **Pro:** More stuctured than paraphrases, more fine-grained than traditional relations
- **Con:** Annotation

# Conclusion

The first part of this talk has no conclusion!

# Experiments with a multi-granularity relation inventory

# Relation Inventory

COARSE

| | |
|---|---|
| BE | *guide dog* |
| HAVE | *car tyre* |
| IN | *air disaster* |
| ACTOR | *committee discussion* |
| INST | *air filter* |
| ABOUT | *history book* |

# Relation Inventory

COARSE DIRECTED

BE

HAVE — HAVE$_1$ *car tyre*

HAVE — HAVE$_2$ *hotel owner*

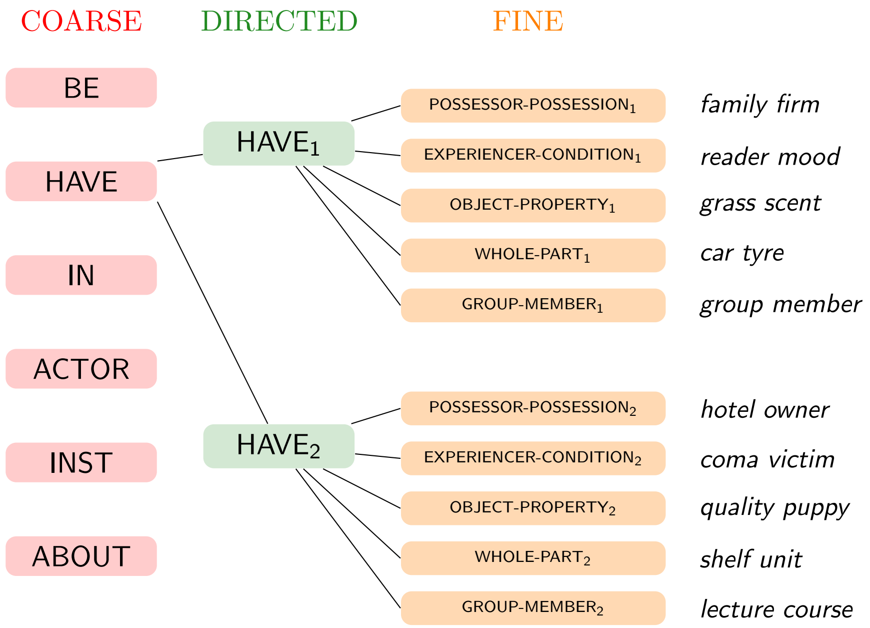
IN

ACTOR

INST

ABOUT

# Relation Inventory



# 1443-Compounds Dataset

- 2,000 candidate two-noun compounds sampled from the British National Corpus
- Filtered for extraction errors and idioms
- 1,443 unique compounds labelled with semantic relations at each level of granularity

| Granularity | Labels | Agreement ($\kappa$) | Random Baseline |
|---|---|---|---|
| Coarse | 6 | 0.62 | 16.3% |
| Directed | 10 | 0.61 | 10.0% |
| Fine | 27 | 0.56 | 3.7% |

- Try it out yourself: `http://www.cl.cam.ac.uk/~do242/Resources/1443_Compounds.tar.gz`

# Information sources for relation classification

**Lexical information:** Information about the individual constituent words of a compound.

**Relational information:** Information about how the entities denoted by a compounds constituents typically interact in the world.

**Contextual information:** Information derived from the context in which a compound occurs.

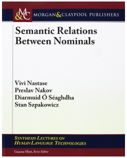
# Information sources for relation classification

**Lexical information:** Information about the individual constituent words of a compound.

**Relational information:** Information about how the entities denoted by a compounds constituents typically interact in the world.

**Contextual information:** Information derived from the context in which a compound occurs.

[Nastase et al., 2013]



# Information sources for *kidney disease*

| | |
|---|---|
| LEXICAL: | |
| *modifier (coord)* | *liver*:460 *heart*:225 *lung*:186 *brain*:148 *spleen*:100 |
| *head (coord)* | *cancer*:964 *disorder*:707 *syndrome*:483 *condition*:440 *injury*:427 |
| RELATIONAL: | *Stagnant water breeds fatal diseases of liver and kidney such as hepatitis* |
| | *Chronic disease causes kidney function to worsen over time until dialysis is needed* |
| | *This disease attacks the kidneys, liver, and cardiovascular system* |
| CONTEXT: | *These include the elderly, people with chronic respiratory disease, chronic heart disease, kidney disease and diabetes, and health service staff* |

# Information sources for *holiday village*

| | |
|---|---|
| Lexical: | |
| *modifier (coord)* | *weekend*:507 *sunday*:198 *holiday*:180 *day*:159 *event*:115 |
| *head (coord)* | *municipality*:9417 *parish*:4786 *town*:4526 *hamlet*:1634 *city*:1263 |
| Relational: | *He is spending the holiday at his grandmother's house in the village of Busang in the Vosges region*<br>*The Prime Minister and his family will spend their holidays in Vernet, a village of 2,000 inhabitants located about 20 kilometers south of Toulouse*<br>*Other holiday activities include a guided tour of Panama City, a visit to an Indian village and a helicopter tour* |
| Context: | *For FFr100m ($17.5m), American Express has bought a 2% stake in Club Méditerranée, a French group that ranks third among European tour operators, and runs holiday villages in exotic places* |

# Contextual information doesn't help

- Contextual information does not have discriminative power for compound interpretation [Ó Séaghdha and Copestake, 2007]

  *We slept in a mountain hut*
  *We slept in a timber hut*
  *We slept in a timber camp*

  *I cut it with the cheese knife*
  *I cut it with the kitchen knife*
  *I cut it with the steel knife*

- Sparsity also an issue
- Not considered further here

# Experimental setup

- 5-fold cross-validation on 1443-Compounds
- All experiments use a Support Vector Machine classifier (LIBSVM)
- SVM cost parameter ($c$) set per fold by cross-validation on the training data
- Kernel derived from Jensen-Shannon divergence [Ó Séaghdha and Copestake, 2008; 2013]:

$$k_{JSD(linear)}(\mathbf{p}, \mathbf{q}) = -\sum_i p_i \log_2 \left( \frac{p_i}{p_i + q_i} \right) + q_i \log_2 \left( \frac{q_i}{p_i + q_i} \right)$$

# Lexical features

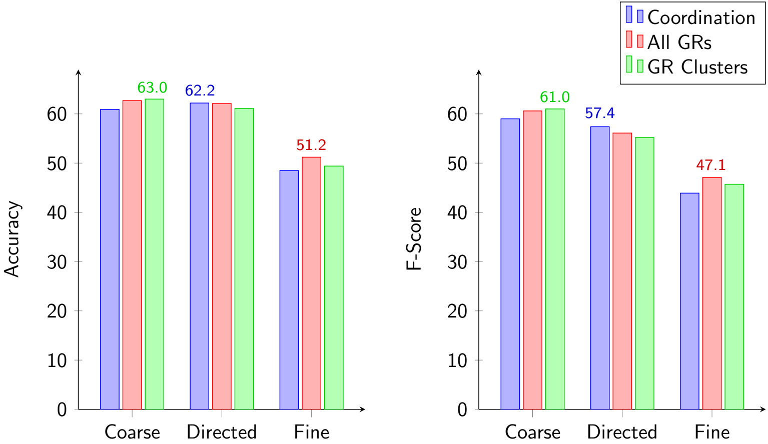
- Distributional features extracted from parsed BNC and Wikipedia corpora.
- One vector for each constituent:

  **Coordination** Distribution over nouns co-occurring in a coordination relation

  **All GRs** Distribution over all lexicalised grammatical relations involving a noun, verb, adjective or adverb

  **GR Clusters** 1000-dimensional representation learned with Latent Dirichlet Allocation from **All GRs** data [Ó Séaghdha and Korhonen, 2011; 2014]

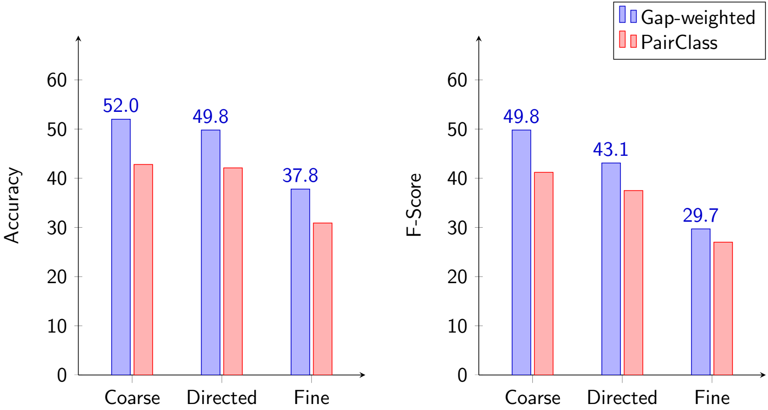
# Results - lexical features



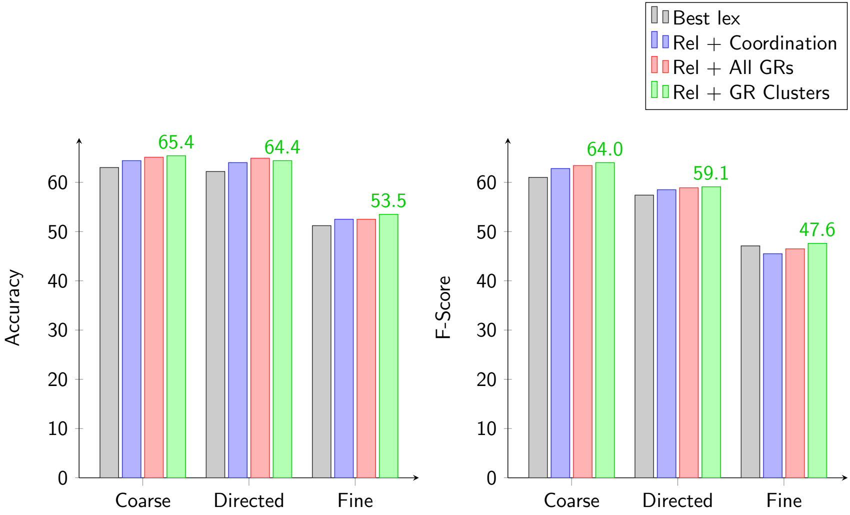
# Relational features

- **Context set** for a compound $N_1 N_2$: the set of all contexts in a corpus where $N_1$ and $N_2$ co-occur
- Context sets for all compounds extracted from Gigaword and BNC corpora
- Embeddings for strings:
  - **Gap-weighted:** all discontinuous *n*-grams [Lodhi et al., 2002]
  - **PairClass:** fixed length (up to 7-word) patterns with wildcards [Turney, 2008]
- Context set representation is the average of its members' embeddings

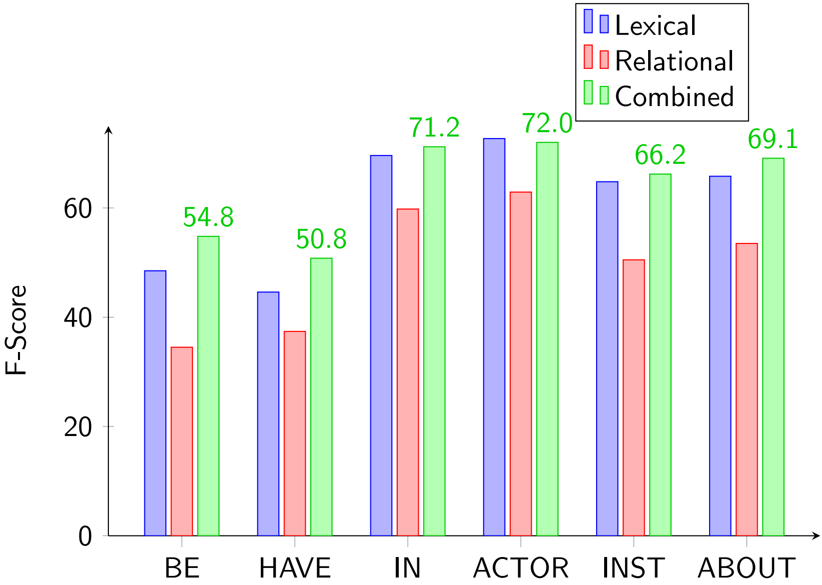
# Results - relational features



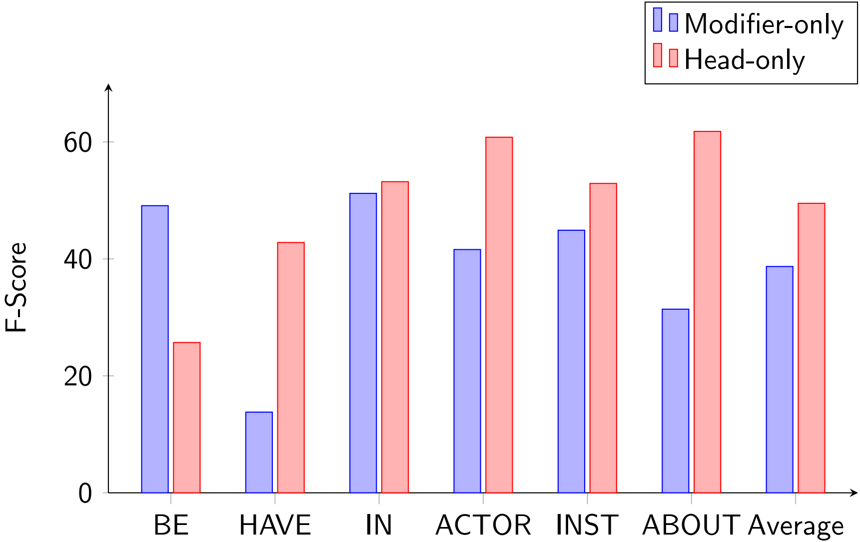
# Results - combined features



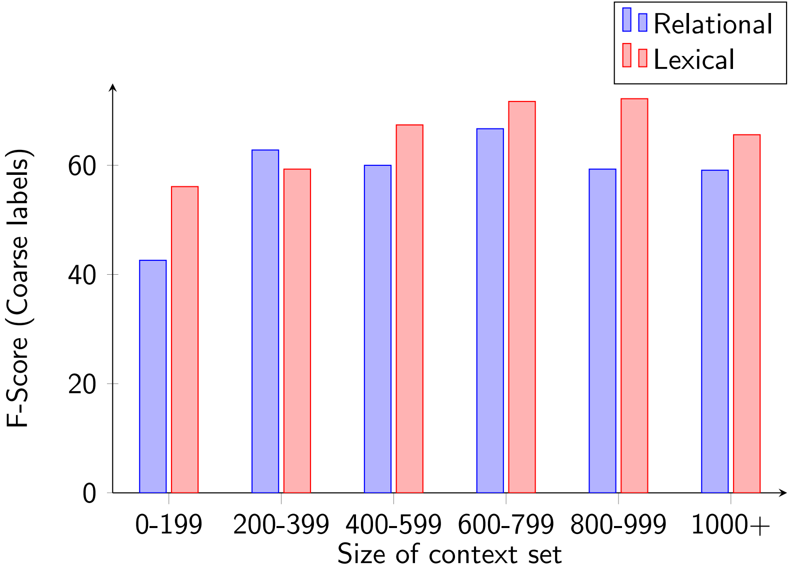
# Performance on individual relations



# Head-only vs modifier-only features



# Effect of context set size



# Conclusions

- Compound interpretation is fun!
- Combining lexical and relation information leads to state-of-the-art performance.
- Previous best performance on 1443-Compounds: 63.6% accuracy on coarse labels [Tratz and Hovy, 2010]
- Our best:

| | Accuracy | F-Score |
|---|---|---|
| Coarse | 65.4 | 64.0 |
| Directed | 64.4 | 59.1 |
| Fine | 53.5 | 47.6 |