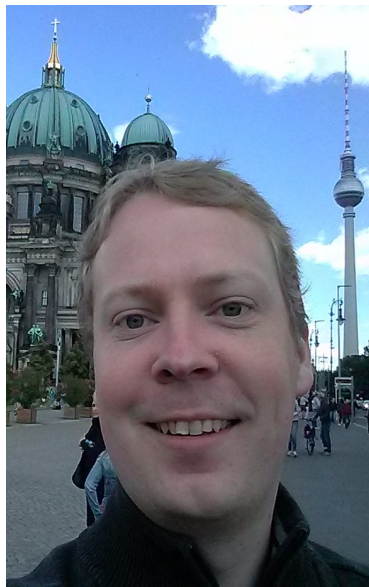# Introduction to Seminar
"Multi-Word Expressions"

William Blacoe

Saarland University, Saarbrücken

About Me

- born in Frankfurt a.M.
- PhD with Mirella Lapata at University of Edinburgh
- PostDoc with Vera Demberg
- inspirations: philosophy, psychology, physics
- passions: travelling, science fiction
- it's best to reach me via
  `wblacoe@coli.uni-saarland.de`

Outline

1 **Multi-Word Expressions**

2 **Bureaucracy**

3 **Presentation**

4 **Essay**

5 **Literature**

Outline

1 Multi-Word Expressions

2 Bureaucracy

3 Presentation

4 Essay

5 Literature

Computational Semantics

- current models of semantics capture word meanings well, e.g. (predicate) logic, (neural) vectors
- composition of word meanings works quite well for certain types of data, e.g. lambda calculus, matrix times vector
- structure of composition is often linear (word by word) or hierarchical (guided by syntax tree)

Artificial Example

"At first I was making eggs and bacon. When I turned around to say 'good morning' to Esther, all of a sudden I burned myself on the stove and she performed ad hoc first aid on me. I almost kicked the bucket."

Problem

- common assumption: phrase meaning is fully compositional
  e.g. "red hat" is intersection of red things and hats
- **WRONG**, most of natural language is not fully compositional
- this class of phrases is called **multi-word expressions**
- MWEs are basically all super-word units that are not fully compositional

"The UN's carrot and stick method did not work on Iran"

MWE vs. metaphor (1)

"The UN's carrot and stick method did not work on Iran"

idiom

- because no carrots or sticks are involved
- definition: "a group of words whose meaning considered as a unit is different from the meanings of each word considered separately."
- meaning: motivation via enticement and punishment

MWE vs. metaphor (2)

"The UN's carrot and stick method did not work on Iran"

metaphor

- because no farmers or doneys are involved
- definition: "an expression that describes a person or object by referring to something that is considered to possess similar characteristics."
- references: Iran=donkey, UN=farmer, UN sanctions=stick, trade incentives=carrot

Tasks in CoLi Literature

- MWE identification: distinguish literal from non-literal usages
- parsing (e.g. Nivre: creating a syntactic and MWE tree)
- identifying literal paraphrases for MWEs, i.e. generating an MWE dictionary
- all kinds of detailed linguistic analyses of MWEs

Downstream Applications

- compressing/summarising text
- machine translation
- information retrieval

Linguistic Definition of MWEs

"A MWE is a combination of words that must be treated as a unit at some level of linguistic processing."

(Nicoletta Calzolari, 2002)

levels:

- lexical, e.g. "ad hoc"
- syntactic, e.g. "at first", "all of a sudden"
- semantic, e.g. "first aid"
- pragmatic/situational, e.g. "good morning"
- statistical, e.g. "salt and pepper"

(Kordoni and Egg, 2013)

Artificial Example (again)

"At first I was making eggs and bacon. When I turned around to say 'good morning' to Esther, all of a sudden I burned myself on the stove and she performed ad hoc first aid on me. I almost kicked the bucket."

Degrees of Compositionality

- literal phrases (i.e. the meaning of the phrase is a straightforward combination of the literal meanings of the words), e.g. "wear a hat", "give a present"
- somewhat abstract, e.g. "take my meaning", "put a price"
- light verb constructions, e.g. "take a bow", "put emphasis"
- idioms, e.g. "kick the bucket", "give it a whirl"

(Fazly and Stevenson, 2008)

Usages of Phrases

- some phrases have a literal and non-literal usage, e.g. "spill the beans"
- in the idiomatic sense the parts need to be combined exactly in this way
- in the literal sense the parts need to be generaliseable,
    - e.g. "drop the beans" because "drop" is similar to "spill"
    - e.g. "spill the popcorn" because "popcorn" is similar to "beans"
- context is necessary to determine the intended meaning

Example MWEs (1)

Example of semi-fixed MWE: "kick the bucket"

- exact words "the" and "bucket" are necessary
- "kick" and "bucket" cannot be replaced by synonyms, e.g. "hit the pail"
- no words can intervene, e.g. "kick the yellow bucket"
- syntax cannot change, e.g. "the bucket was kicked"
- morphology of "kick" **can** vary, e.g. "He is kicking the bucket", "He kicked the bucket"

Example MWEs (2)

Example of more flexible MWE: "elephant in the room"

- exact words "elephant" and "room" are necessary
- "elephant" and "room" cannot be replaced by synonyms, e.g. "rhino in the kitchen"
- some words can intervene, e.g. "big elephant in the living room"
- morphology of "elephant" **can** vary, e.g. "the/an/one/two elephant(s) in the room"

Outline

1 **Multi-Word Expressions**

2 **Bureaucracy**

3 **Presentation**

4 **Essay**

5 **Literature**

Credit Points

- B.Sc.: 7 CP for presentation + essay
- M.Sc.: 4 CP for presentation
- M.Sc.: 7 CP for presentation + essay

Seminar Homepage

the Seminar Homepage
`http://www.coli.uni-saarland.de/courses/mwe16`
will contain

- these slides
- suggested topics
- schedule of who presents what when
- your slides?

Time

- every Wednesday 12-14 cum tempore
- more than one presentation per session?
- presentation schedule on my website

Outline

1 **Multi-Word Expressions**

2 **Bureaucracy**

3 **Presentation**

4 **Essay**

5 **Literature**

Recommended Readings

- Lecture Slides by Tim Baldwin

  http:

  //people.eng.unimelb.edu.au/tbaldwin/pubs/altss2004.pdf

- Chapter on Multi-Word Expression by Tim Baldwin from Handbook of Natural Language Processing (2nd edition)

  http:

  //people.eng.unimelb.edu.au/tbaldwin/pubs/handbook2009.pdf

- Lectures Slides by Aline Villavicencio and Valia Kordoni

  https://lct-master.org/files/mwe1-6.pdf

Study

- recent research papers
- search using e.g. google scholar
- search by author, topic or title

Purpose of Presenting

- show you understand the paper(s)
- make the content understandable for others
- become familiar with current research and scientific methods

Preparation

- study your paper
- submit slides to me a few days in advance for feedback
- prepare discussion questions
- practise giving your talk (verbally!) at home

Style

- give an easy-to-follow logic/story of the work
- successively build up the contribution
- make transitions between parts gradual
- use your own words to explain things
- don't do "English with English subtitles"

Granularity

- find good level of abstractness/summary of your paper(s)
- balance formal and prosaic parts (formulas decrease attention)
- choose which parts to explain in detail and which to gloss over
- clarify abbreviations

Organisation

- your choice: take questions during or after presentation
- use other media if you like, e.g. white board
- involve the audience
- your choice: sit or stand

Slides

- use LaTeX beamer or any other software
- use bullet points rather than sentences
- use images if possible
- show understandable examples
- not too much information on one slide
- visualise data/results if possible

Structure (1)

- **introduce** by building on basic understanding of linguistics and NLP
- but spending too much time on repeating basics deters attention
- **motivate** matter by showing current problem or shortcoming
- convey the **big idea**
- say what this contribution is **useful** for
- if necessary go into theoretical **preliminaries**

Structure (2)

- set paper in context of **existing work**
- introduce present **model** or theory
- show **strengths** and **advantages** of present approach over existing models
- explain setup of **evaluation**: resources, software, experiments
- **show results** and (some) data and interpret them
- propose **outlook** and future improvements of present approach
- provide properly cited **references** at the end

Substance

- if your paper doesn't seem to contain enough material to fill presentation time, seek more material from related literature
- presentation time: 25 to 35 minutes
- discussion time: the balance of your 45 minutes (?)
- all students fill out feedback form

Discussion Afterwards

- the presenter leads the discussion
- presenter and audience prepare questions for each other before the lesson
- I will participate in the discussion
- participation is a part of your final grade
- you can share ideas for expanding the presented work

Outline

1 **Multi-Word Expressions**

2 **Bureaucracy**

3 **Presentation**

4 **Essay**

5 **Literature**

Procedure

- you must register in time that you want to submit an essay
- submit the essay to me towards the end of the semester
- use LaTeX or any other software to write the essay
- cite properly
- length of essay depends on type of essay (ask me)

Content and Structure

- not necessarily original
- but it **is** possible to do your own experiment(s)
- also possible: write a survey
- also possible: write a research proposal
- either way: put your work in context of current related literature

Topic

choose a topic for your essay, either

- based on your presentation, but significantly different
- or from another suggested paper that was not presented
- or your own suggestion that I and you agree on

Outline

1 **Multi-Word Expressions**

2 **Bureaucracy**

3 **Presentation**

4 **Essay**

5 **Literature**

Recommended Readings (again)

- Lecture Slides by Tim Baldwin

  http:
  //people.eng.unimelb.edu.au/tbaldwin/pubs/altss2004.pdf

- Chapter on Multi-Word Expression by Tim Baldwin from Handbook of Natural Language Processing (2nd edition)

  http:
  //people.eng.unimelb.edu.au/tbaldwin/pubs/handbook2009.pdf

- Lectures Slides by Aline Villavicencio and Valia Kordoni

  https://lct-master.org/files/mwe1-6.pdf

Categories (1)

your selection for paper/essay can be based on

- tasks
  - MWE identification
  - (hybrid) parsing
  - word sense disambiguation
- integration of MWEs into NLP applications
  - machine translation
  - paraphrase detection
  - summarisation

Categories (2)

your selection for paper/essay can be based on

- resources
    - representation of MWEs in dictionaries and ontologies
    - constructed manually and automatically
- resource annotation
    - in treebanks
    - in other datasets

Categories (3)

your selection for paper/essay can be based on

- evaluation of MWE treatment techniques
- lexical, syntactic or semantic aspects of MWEs
- multilingualism
    - models of first and second language acquisition of MWEs
    - cross-linguistic studies of MWEs

Suggested Papers

See course website:

www.coli.uni-saarland.de/courses/mwe16

References

Afsaneh Fazly and Suzanne Stevenson. A distributional account of the semantics of multiword expressions. *special issue of the Italian Journal of Linguistics*, 20(1): 157–179, 2008.

Valia Kordoni and Markus Egg. Multiword expressions and collocations in nlp: A state of the art overview. 2013.

Ralph Grishman Nancy Ide Alessandro Lenci Catherine MacLeod Antonio Zampolli Nicoletta Calzolari, Charles Fillmore. Towards best practice for multiword expressions in computational lexicons. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, pages 1934–1940, Las Palmas, Canary Islands, 2002.