Maintaining Consistency of Monolingual Verb Entries with Inter-Annotator Agreement

Silvie Cinková, Martin Holub, Lenka Smejkalová,
Anna Lauschmannová and Jonáš Thál

Institute of Formal and Applied Linguistics
Charles University in Prague
NFL Lund, May 23-27, 2011

Word Sense Disambiguation in NLP

- statistical machine learning
  - hand-annotated data
  - no rules, just hints what matters (**features**)
- the computer learns to mimic human judgment
- ... but it will never perform the task better than humans themselves do!
- humans must agree on what is correct first

Ambiguity vs. vagueness

vagueness

- a concept allows for borderline cases
  - what does *tall* mean?
- we have a common understanding of the concept but there will individual differences on which instances will be associated with the concept:
  - *a man of 170, 189, 190 cm?*

ambiguity

- a word denotes several concepts that are cognitively very far from each other
- context disambiguates
- normally just one option meant

Ambiguity

- **SPRING (n)**

1. *season between winter and summer*
2. *water flowing up from under the ground*
3. *a long thin metal coil/spiral*
4. *a jump*

Semantic modulation

- NEWSPAPER

1. large printed sheets of folded paper containing news, articles, etc.
2. organization that produces a newspaper (1)

A number of newspapers reported these incidents.

Eventually, Murdoch bought a newspaper, too.

Vagueness

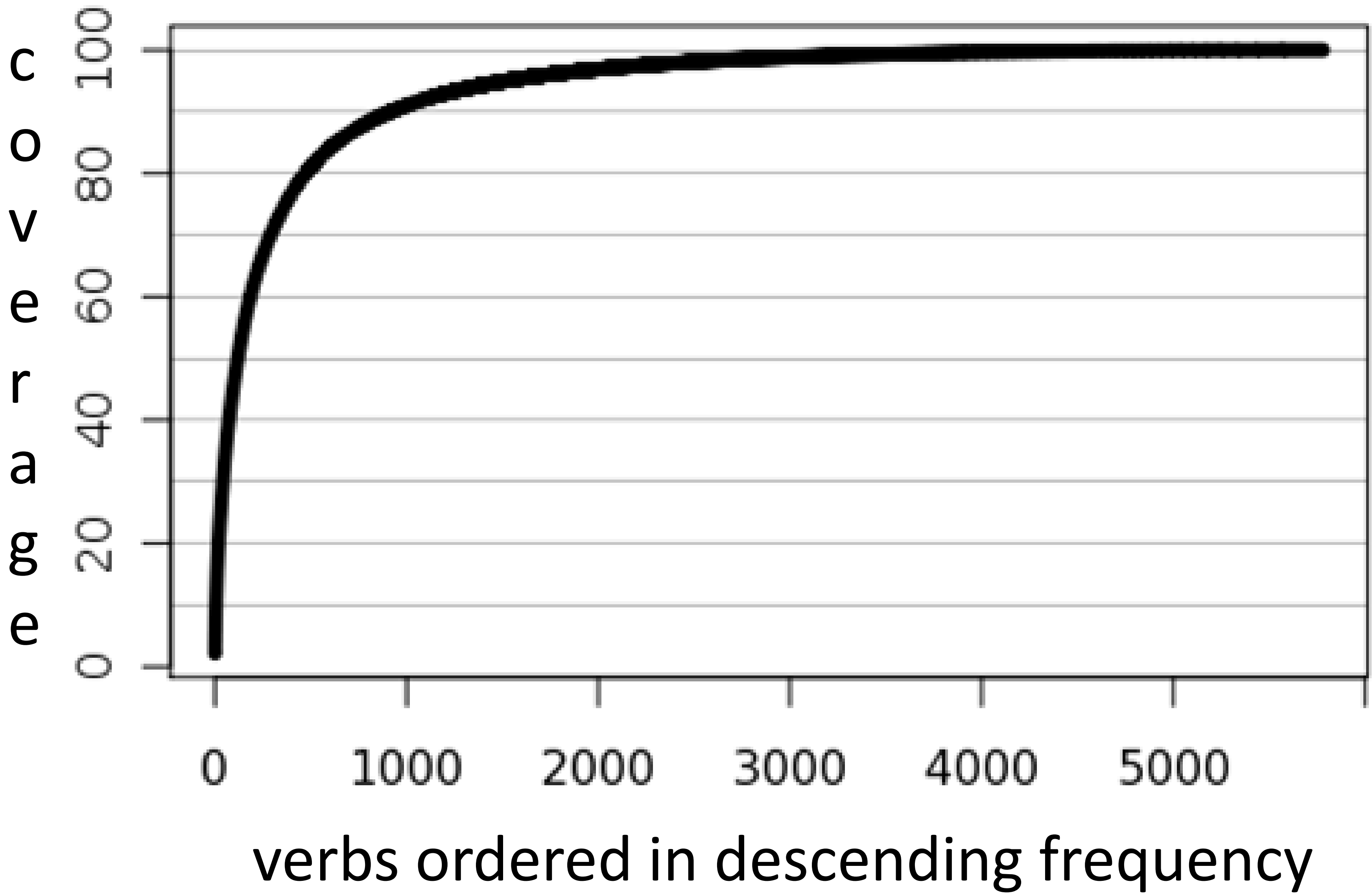
- **GLASS**

1. **hard clear substance**
2. **small container for drinking**
3. (summarizing term for) **attractive artifacts made of glass:** *a collection of Italian glass*
4. mirror ✞
5. barometer✞

Verbs in language

- Verbs denote
  - relations between entities
  - events that entities undergo
- many less words for events/relations than for entities!
- almost any can be described by the top most frequent 1000 verbs

Verbs in BNC 50

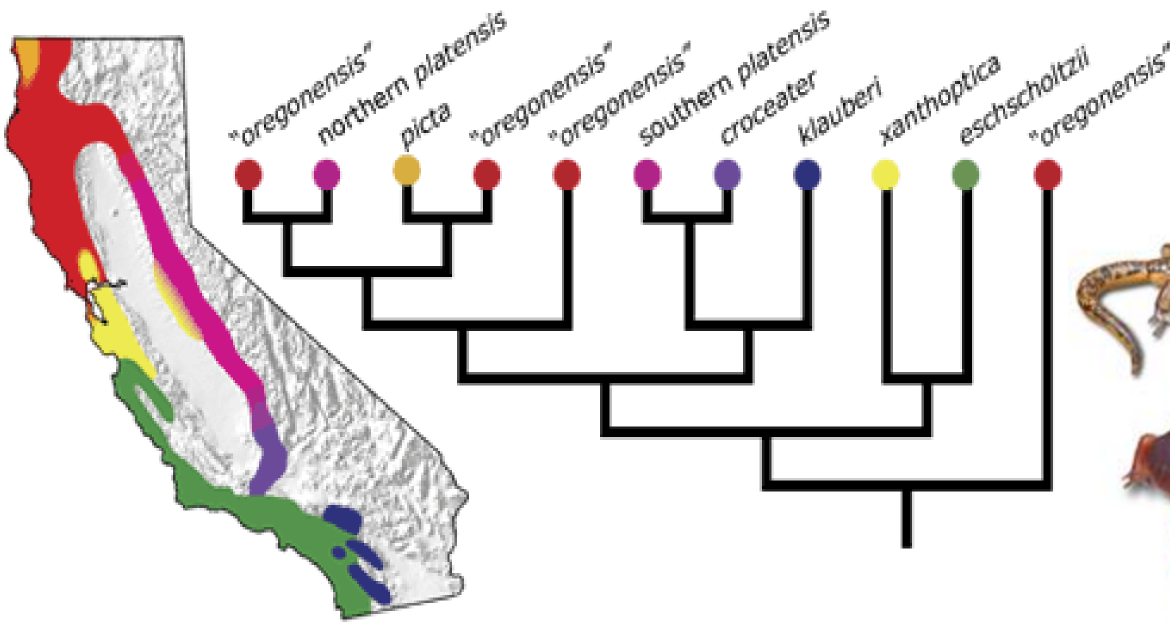


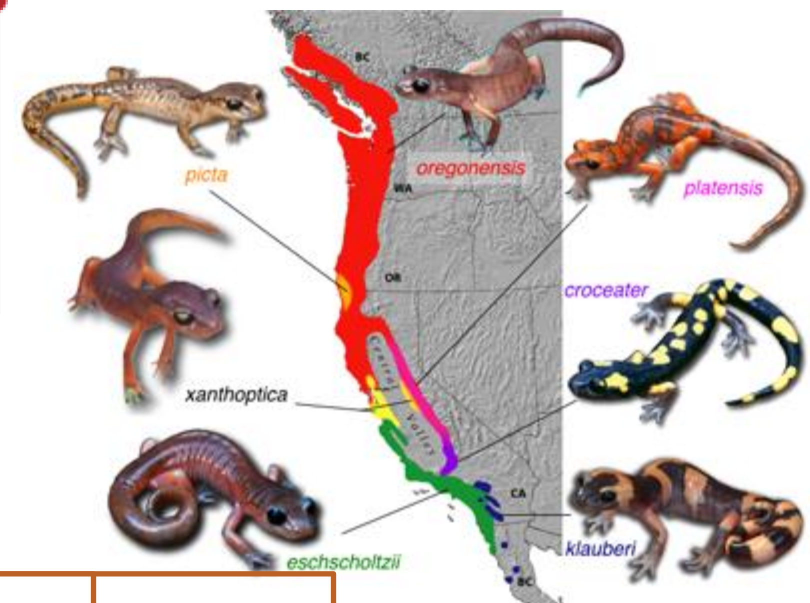
One concept – different events

- with different participants
  - A boy was **throwing/hurling/tossing/blowing** stones.
  - Fast driving on gravel roads **throws** up **rocks** which can scar the car./My wheels **spit** **gravel** and I long for a bigger windshield.
  - Tracey Morton was **thrown** into the air and carried down the road by the motorbike./ I saw this one bus **thrown** across the road/She was **thrown** from her horse.
  - ... fragmental material **thrown** into the air by explosive volcanic activity . The volcano was **throwing/spewing** stones and lava.
  - Dawn **threw/cast** sunlight across the ruins of the old city.

Vagueness: Ensatina salamander

van Deemter, 2010





E1 ❤	E2	❤	E3	❤	E4	❤	E5	❤ E6

Grandchildren can interbreed where parents & grandparents could not.

Inherently vague concept: *throw*

1. Human uses **hands** to **propel** a physical **object** in a direction for a short distance

   *Tourists are encouraged to **throw** coins into the fountains for good fortune.*

2. Human violently **pushes** or shoves or kicks another **human** so that the other human loses control over his movements and **falls**

   He threw her to the ground/against the cupboard…

3. Human **discards** or gets rid of an artifact or **stuff**

   He threw the paper straight into the bin/threw it away, threw it out.

4. Human (**murderer**) disposes of or **discards** or **hides** the **body** of his **victim** to some place

   *Their corpses were **thrown** down a well.*

5. Human **feeds** an **animal/animals** with a physical object or a substance

   *It was like **throwing** a piece of meat to sharks.*

Ambiguity vs. vagueness

ambiguity = no interbreeding
vagueness = interbreeding

propel stuff ❤ discard stuff

propel stuff ❤ feed animals with stuff

discard stuff X discard person

propel stuff X propel person

feed animals with stuff ❤ discard stuff

None of these can breed with others from the set
(throw light, throw into disarray, throw a glance, horse throws...)

A really vague sentence

- *Osbern has his father killed by a lowly mob and* ***thrown*** *to birds and wild animals.*

1. propel stuff
2. push/shove a person
3. discard stuff
4. murderer discard/hide corpse
5. feed animals with stuff

1-5 apply to different degree.
All they have in common outside this context is perhaps the “away” direction.

Other approaches

- "a translation task rather than a classification task" (Liberman, 2009)
- "I don't believe in word senses" (Kilgarriff, 1997)
  - there is no static inventory of meanings; they depend on the final application
- **Non-linguists** agree well on whether or not a word in two contexts is used in the same sense (Rumshisky, Verhagen and Moszkowicz, 2009)
- Textual Entailment: does text *t* entail text *t1*?
  - reasonable interannotator agreement (Bentivogli et al., 2010)

Corpus Pattern Analysis (P. Hanks)

1 **[no object]**
[[Human]^[Animate]^[Vehicle]] zoom AdvDirection
[[Human | Animate| Vehicle]] moves very quickly [[Direction]]

2 **[no object]**
[Artifact | {camera} = Camera] zoom [{in on} Physical Object]
[[Artifact = Camera]] focuses on [[Physical Object]]

3 **[Asset = Price] zoom AdvDirection**
[[Asset = Price]] increases very quickly and dramatically

an archaeologist who	zoomed	1	around the world thrashing Naz
ued that the cameras	zoomed	2	in more often on his Gucci loaf
r pre-tax profits have	zoomed	3	ahead by 92% to almost £27m, a
year to September 30	zoomed	3	from £27.5m to £36.5m.
iant Cable & Wireless	zooming	3	ahead 20p to 667p after 2.4m sh

Corpus Patterns

- Different from dictionary readings
  - meaningful clusters that
    - share one paraphrase
    - the participants in each syntactic position form a homogeneous group
    - have the same morphosyntactic behavior
  - “norms” and “exploitations”

  ride a horse – ride a cat

  being thrown into a pool of desire

What are we doing?

- 40 verb entries made by Patrick Hanks (PDEV)
- 250+ BNC concordances sorted by PH checked/revised by SC
- 3x annotated 50 other concordances
- interannotator agreement measurement
- revisions of entries to ensure interannotator agreement above 0.6 (F/C kappa, Kr. alpha)

Our objectives

- check for each verb whether people can agree on the patterns
- remove superfluous disagreements such as

  “concordance is either a figurative use of X or a syntactic deviation of Y”
- revise bad patterns, revise data (gold standard)
- model superclusters of patterns that come into consideration when context is underspecified
  - in the data rather than in the entries
- teach the computer to make similar (super)clusters
- find out whether these clusters help applications (machine translation, paraphrasing, entailment)

References

- van Deemter, Kees. 2010. Not exactly. In praise of vagueness. Oxford University Press.
- Brown, Susan Windisch et al. 2010. Number or Nuance: Which Factors Restrict Reliable Word Sense Annotation? Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valetta, Malta.
- Rumshisky, Anna et al. 2009. The Holy Grail of Sense Definition: Creating a Sense-Disambiguated Corpus from Scratch. Fifth International Workshop on Generative Approaches to the Lexicon (GL 2009). Pisa, Italy.
- Kilgarriff, Adam. “I Don’t Believe in Word Senses”. 1997. Computers and the Humanities, 13 (2).
- Bentivogli, Luisa et al. 2010. Building Textual Entailment Specialized Data Sets: A Methodology for Isolating Linguistic Phenomena Relevant to Inference. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valetta, Malta.
- Hanks, Patrick Wyndham. in press. Lexical Analysis. Norms and Exploitations. MIT Press.
- Erk, Katrin. 2010.*What Is Word Meaning, Really? (And How Can Distributional Models Help Us Describe It?)* Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics, ACL 2010, pages 17–26, Uppsala, Sweden
- McCarthy, Diana; Gaylord, Nicholas. 2009. *Investigations on Word Senses and Word Usages. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*