Shimmering Semantics: Some Insights for Ontologies from Lexicography

Ken Baclawski
Ontology Summit 2021
28 April 2021
Caveats and Sources

This presentation is derived from a number of sources:
- Charles Sanders Peirce (from 1882 to 1891)
- My own work (1992 to 1997)
- Articles by Patrick Hanks (Hanks 2008, 2021)
- Wikipedia articles (verified with other sources)

I am not a lexicographer, but I have some background in it
Lexicography

- Practical lexicography
  - concerned with developing dictionaries
- Theoretical lexicography
  - scholarly discipline
  - not a sub-branch of applied linguistics
Objectives of Lexicography

- Analyze the semantic, syntagmatic, and paradigmatic relationships within a lexicon (vocabulary)
- Develop theories of dictionary components and structures linking the data in dictionaries
- Accommodate the needs for information by users in specific types of situations
  - profile intended users via competences
  - identify needs
- Design techniques for efficiently accessing dictionaries
Gary Berg-Cross pointed out that Charles Sanders Peirce contributed over 16,000 entries to The Century Dictionary.

The following is from Peirce’s Logic Notebook:
A dictionary definition will be (or at any rate contains) a definition proper in the case of a scientific or other exact conception; but an ordinary word needs an explanation, not a definition which almost itself needs to be expounded… An “explanation” familiarizes the mind with the use of the word by bringing together in the briefest terms possible the subclasses of occasions in which it is used and giving an interpretation of it in each of them. [Italics added] Peirce Project Newsletter Vol. 3 No. 1 page 2
Understanding meaning in language

- **Norms** are the basic shared conventions
  - including conventional metaphors

- **Exploitations** are freshly created metaphors, unusual phrasing, etc.

- Consequently, there are two different rule systems that interact with one another
Exploitation Rules

- **Ellipsis**: Omitting what has been said (in a dialog) or what is obvious (background knowledge)

- **Anomalous argument**: Common device for comedy but could be serious as well
  - "One morning I shot an elephant in my pajamas. How he got into my pajamas I'll never know."
  - "Always vacuum your moose from the snout up, and brush your pheasant with freshly baked bread, torn not sliced."
Exploitation Rules

- **Metaphor**: Explaining one thing by means of another
  - "All the world's a stage"
- **Simile**: Comparing one thing with another
- Other figurative devices: Wikipedia lists over 200 figure of speech patterns, from **accismus** (a feigned refusal of something earnestly desired), to **zoomorphism** (applying animal characteristics to humans or gods)
• **A synset** is a set of synonymous words or short word phrases representing one word sense.
  - Note: Patrick Hanks calls these "lexical sets", but lexical set currently means something else.
  - The term synset is from WordNet®

• The problem with synsets is that even in basic (norm) usage they "shimmer".
  - Different uses will contract or expand a synset, cutting across semantic categories (classes)
  - Exploitation has even more dramatic effects by creating entirely new synsets
Assigning a word sense to a single fixed semantic category is problematic

- "Every verb picks out its own set of typical direct objects from within a broad semantic class." (Hanks 2008)

Example: meeting vs wedding

- One can attend, arrange, plan and organize a meeting or a wedding
- One can convene, summon, call and chair a meeting but not a wedding
- One can celebrate and perform a wedding but not a meeting
More Examples

- **Example: wash and amputate**
  - One can wash or amputate a leg, arm or foot.
  - One can wash one's face or hair, but one cannot (prototypically) amputate them.

- **Assuming one wishes to specify the domain and range of predicates, how can one deal with this?**
Strategy I

• Semantic categories (classes)
  – Should one create a new semantic category for every sense of every verb when needed?
  – This would result in a large number of overlapping semantic categories.
  – The categories are statistical with some members being more representative than others.
Strategy II

- Semantic coercion
  - forces a term to be a member of a different semantic category in a particular context
  - shifts the basic meaning of a word due to semantic requirements imposed by other words in the given context
- In many cases, one can give convincing arguments for each of the two strategies.
Lexicography and Semantics

- Meanings in language are associated with words in prototypical phraseological **patterns** (not with words in isolation).
- Meanings in text are interpreted by **pattern matching** – mapping bits of text onto the patterns in our heads.
- Members of a language community share prototypical patterns.
- Good pattern matches are "norms".
- Poor pattern matches are either "exploitations" or mistakes.
  - A set of exploitation rules is necessary to distinguish creative usages (and to understand them) from mistakes.
Identifying phraseological patterns

- Traditional manual methods
- Machine learning methods
  - Corpus-driven linguistics (i.e., based purely on the corpus itself)
  - Combination of traditional and statistical
Some Consequences for Ontologies

- **Use of natural language terms in identifiers can result in misunderstanding**
  - Words can have many senses and usage patterns.
    - The intended sense and usage patterns of an ontology identifier may clash with normal senses and usage patterns.
  - Context affects both the sense and normal usage patterns of a word
    - Reusing an ontology term or an entire ontology in another context may change both the intended sense and usage patterns.

- **Using techniques from lexicography could help reduce misunderstanding**
Example of a Definition

- **Pounce (mass noun)**
- 1 A fine resinous powder formerly used to prevent ink from spreading on unglazed paper or to prepare parchment to receive writing.
- 2 Powdered charcoal or other fine powder dusted over a perforated pattern to transfer the design to the object beneath.
  - ‘As actual practice can demonstrate, if a pricked design is pounced from its verso, the pounce marks register more distinctly, than if pounced from the recto.’

Source: OED
Conclusions

- Lexicography and ontology have similar objectives and could benefit from each other.

- A human language is a system with two interlinked systems of rules: normal and exploitative
  - The two systems interact with one another
  - Today’s exploitation may be tomorrow’s norm

- There are interesting opportunities for neuro-symbolic methods to be used for identifying exploitations.
References

- Image of dictionaries is from Wikipedia page on dictionaries License
- Image of scrabble is from Erin McKean License
- The Peirce Project Newsletter https://peirce.sitehost.iu.edu/news/3_1/3.1.htm
- P. Hanks, Exploitation Rules, Research Institute of Information and Language Processing, University of Wolverhampton, nd.
- The “elephant in my pajamas” is due to Groucho Marx in Animal Crackers.