A Taxonomic Classification of WordNet Polysemy Types

Abed Alhakim Freihat
(Qatar Computing Research Institute – Qatar)

Fausto Giunchiglia
(University of Trento – Italy)

Biswanath Dutta
(Indian Statistical Institute – India)
Overview

- Polysemy and Polysemy Types in WordNet
- Structural Patterns
- Classification Principles
- Metaphoric Structural Patterns
- Specialization Polysemy Structural Patterns
- Homonymy Structural Patterns
- Approach Overview
- Results
Polysemy in WordNet

- Polysemous words: Words that have more than one meaning (in different contexts).
- WordNet 2.1. contains:
  - 147,257 words,
  - 117,597 synsets,
  - and 207,019 word-sense pairs.
- Among them: 27,006 polysemous words,
- 14,530 of them are nouns (≈ 27,000 Synsets)
  (Polysemous nouns at concept level only)
Polysemy in WordNet

- Polysemous words: Words that have more than one meaning (in different contexts).
- WordNet 2.1. contains:
  - 147,257 words,
  - 117,597 synsets,
  - and 207,019 word-sense pairs.
- Among them: 27,006 polysemous words,
- 14530 of them are nouns (≈ 27,000 Synsets).
  (Polysemous nouns at concept level only)
Polysemy Types in WordNet

- **Compound Noun Polysemy:**
  - read/write head, head
  - drumhead, head

- **Metonymy:**
  - Chicken (bird)
  - Chicken (food)
Polysemy Types in WordNet:

- **Specialization Polysemy**:
  - *Australian turtledove*, *turtledove* (australian turtledoves)
  - *Turtledove* (old world turtledoves)

- **Metaphors**:
  - *Fox* (animal)
  - *Fox* (person)

- **Homonymy**:
  - *Bank* (river bank)
  - *Bank* (financial institution)
A polysemy instance is a triple \([\{T\}, s_1, s_2]\), where
- \(s_1\) and \(s_2\) are two polysemous synsets that have the terms \(\{T\}\) in common.
- The term \textit{bazaar} has three polysemy instances:
  \([[\text{bazaar}, \text{bazar}], #1, #2], [[\text{bazaar}], #1, #3], [[\text{bazaar}], #2, #3] \)
  - \#1 bazaar, bazar: a shop where a variety of goods are sold.
  - \#2 bazaar, bazar: a street of small shops.
  - \#3 bazaar, fair: a sale of miscellany; often for charity.
Structural Pattern

• WordNet organizes noun synsets in a hierarchy
  – Entity is the root of the hierarchy.
  – Direct Hypernym/Hypernym is the used relation.
  – Any two synsets have at least one common subsumer.
• The structural pattern of $I=\{T\},s_1,s_2\}$ is a triple
  $<r,p_1,p_2>$, where :
  • $r$ is the least common subsumer of $s_1,s_2$,
  • $p_1/p_2$ is a hypernym of $s_1/s_2$,
  • $r$ is a direct hypernym of $p_1$ and $p_2$.  
Structural Pattern Example

mercantile establishment, ...

a place of business for retailing goods

marketplace, mart

an area in a town where a public mercantile establishment is set up

bazaar, bazar

a street of small shops (especially in Orient)

shop, store

a mercantile establishment for the retail sale of goods or services

bazaar, bazar

a shop where a variety of goods are sold
Common Parent structural pattern

- $P = <r, p_1, p_2>$ of a polysemy instance $I = [\{T\}, s_1, s_2]$ is a common parent structural pattern if $p_1 = s_1$ or $p_2 = s_2$. 
Classification Principles

- Exclusiveness property:
  - Let $\langle r, p_1, p_2 \rangle$,
  - $p_1$ and $p_2$ fulfill the exclusiveness property if they are disjoint.
  - $\langle \text{entity}, \text{abstract entity}, \text{physical entity} \rangle$ fulfill the exclusiveness property.
  - $\langle \text{person}, \text{expert}, \text{scientist} \rangle$ do not fulfill the property.
Classification Principles ..

- Collective Exhaustiveness
  - Let $P = \langle r, p1, p2 \rangle$,
  - $p1$ and $p2$ fulfill the exhaustiveness property if they constitute subclasses of the class $r$.
  - $\langle \text{person, male, female} \rangle$ fulfill the exhaustiveness property.
  - $\langle \text{person, female, worker} \rangle$ do not fulfill the property.
A structural pattern $p = \langle r, p_1, p_2 \rangle$ is metaphoric if $p_1$ and $p_2$ do not fulfill the collectively exhaustiveness property.

Possible violations of the property:

- $p_1$ and $p_2$ are not compatible (Class/role) and can not be subsumed by the pattern root $r$.
  - $\langle \text{person, female, worker} \rangle$
    - gold digger (worker) vs. gold digger (female)
  - $p_1$ subsumes $p_2$ or $p_2$ subsumes $p_1$.
    - $\langle \text{organism, animal, person} \rangle$
      - Fox (animal) vs. fox (person)
A structural pattern $p = <r, p_1, p_2>$ is specialization polysemy structural pattern if $p_1$ and $p_2$
- fulfill the collectively exhaustiveness property, and
- do not fulfill the exclusiveness property.

- <person, expert, scientist>
  - statistician (expert) vs. statistician (scientist)
Homonymy structural pattern

- A structural pattern \( p = \langle r, p_1, p_2 \rangle \) is a Homonymy structural pattern if \( p_1 \) and \( p_2 \)
  - fulfill the exclusiveness property, and
  - fulfill the collectively exhaustiveness property
- \( \langle \text{organism, animal, plant} \rangle \)
  - red fox (animal) vs. red fox (plant)
Approach Overview

- **Structural pattern discovery: (automatic):**
  - The algorithm returns structural patterns associated with their corresponding polysemy instances.
  - **Notes:**
  - Compound noun polysemy precedes this procedures.
  - Why? Compound noun polysemy is a source of noise :)
  - The structural patterns whose pattern root resides in the first and second level in WordNet hierarchy were excluded.
    - Why?
    - These patterns belong mainly to metonymy (CORELEX classes)
• **Structural pattern classification (manual):**
  - Classify each of the discovered structural patterns to metaphor, specialization polysemy, or homonymy patterns.

• **Identifying false positives (manual):**
  - Assign false positives to their corresponding polysemy type.
Classification of Polysemy Types

Results

- **Structural pattern classification**

| Polysemy type    | #patterns | #instances |
|------------------|-----------|------------|
| Spec. Polysemy   | 823       | 9902       |
| Metaphoric       | 134       | 1697       |
| Homonymy         | 71        | 1389       |
| **Total**        | **1028**  | **12988**  |

- **False positive Identification**

| Polysemy Type     | #Instances | #False Positives |
|-------------------|------------|------------------|
| Spec. Polysemy    | 9902       | 1740             |
| Metaphoric        | 1697       | 175              |
| Homonymy          | 1389       | 295              |
| **Total**         | **12988**  | **2210**         |
Evaluation

- To evaluate our approach, 3797 polysemy instances were evaluated by two evaluators.
  - Two master students were taught and trained to classify polysemy.
  - High Agreement 96%
- But the evaluators were
  - not experts, and
  - not native speakers.
Conclusion and future Work

• we have presented how to use two taxonomic principles for classifying the polysemy types in WordNet.

• We have demonstrated the usefulness of our approach on classifying three polysemy types.

• We were able to discover all specialization polysemy structural patterns and subsets of the metaphoric and homonymy structural patterns.

• We aim to continue our work to study the metonymy patterns in the upper level of WordNet hierarchy.
Thank you for your attention :)

Classification of Polysemy Types