

Review : Automatic Acquisition of Hyponyms from Large Text Corpora

Automatic Acquisition of Hyponyms from Large Text corpora 리뷰

Geonil Yun,

rjsdlf45@gmail.com

2021. 08. 31.

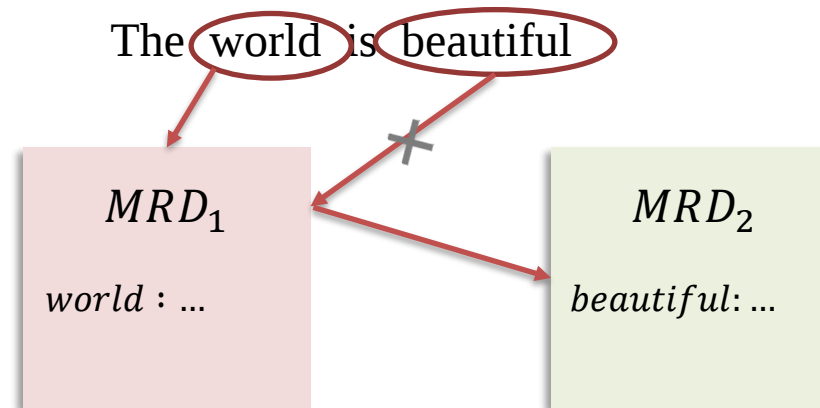


Introduction

- Extracting lexical information

- Machine Readable Dictionary(MRD)

- ✓ fixed entries : find terms and expressions that are not defined MRDs (other textual resource)



- view corpus also as a source of information about language it is written in

Introduction

- Find lexical-relation with pattern

NP = Noun Phrase

- Hyponym relation

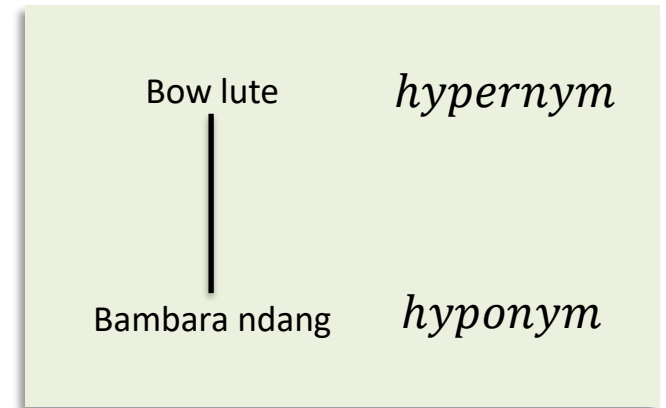
“ The **bow lute**, such as the **Bambara ndang**, is plucked and has an individual curved neck for each string. ”

NP_0 such as $\{NP_1, NP_2, \dots, (and|or)\} NP_n$

imply

for all $NP_i, 1 \leq i \leq n, \text{hyponym}(NP_i, NP_0)$

$\text{hyponym}(\text{"Bambara ndang"}, \text{"bow lute"})$



hyponym

hypernym

Frame : L_0 is a (kind of) L_1

Introduction

- **Pattern-based, Pattern recognition**

- LDOCE : use hierarchy of patterns, part-of-speech(POS), wildcard character pattern
- Taxonomy

- Noun taxonomy in MRD
 - start with "Any" – *NP*

MRD (Webster 7th)

nectar : **any** delicious **drink**

discord : **any** harsh or unpleasant **sound**

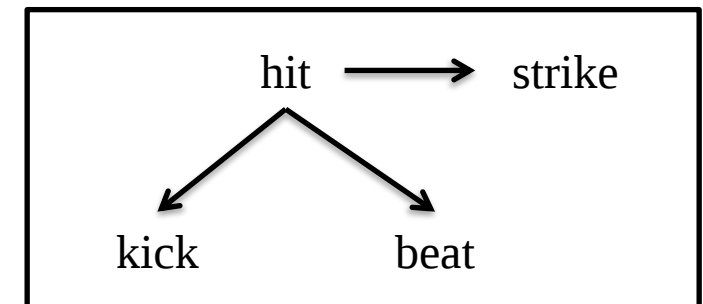
- Key verb(and Key noun, Function noun) taxonomy in MRD
 - verbs are described as : *to* **VERB**

MRD (LDOCE)

strike : **to hit**

kick: **to hit** with the foot

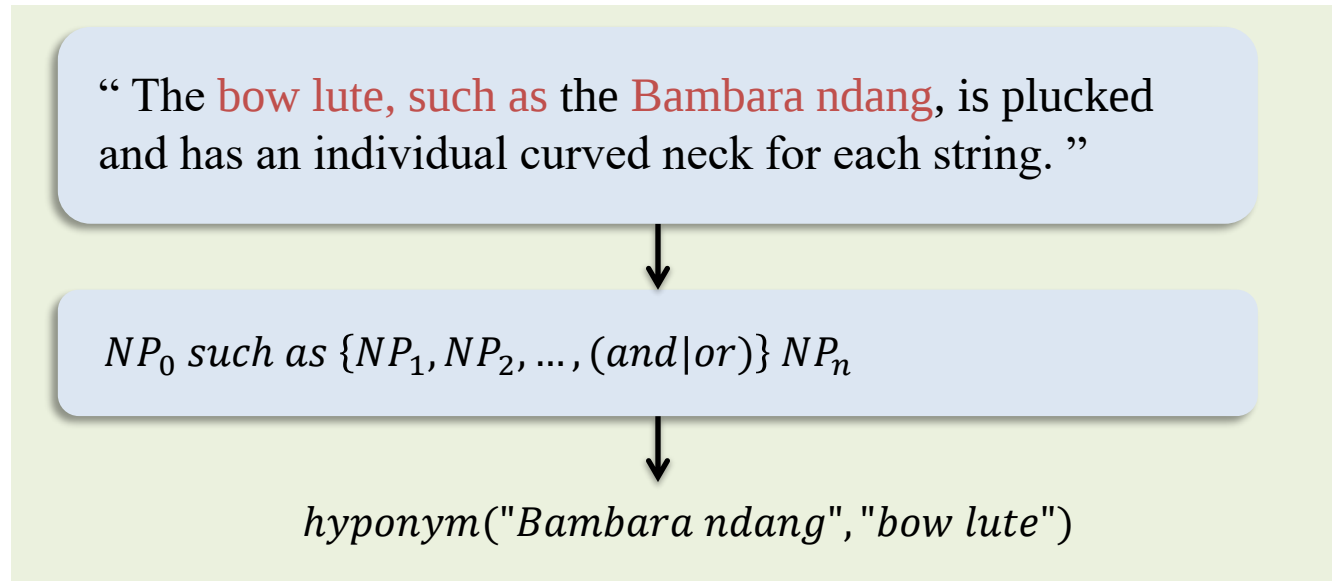
beat : **to hit** many times



Introduction

- **Pattern recognition vs Parsing**

- Pattern recognition is more accurate and efficient than parsing
 - Apply pattern-based relation recognition



- Building lexical hierarchy **automatically**

pattern-based > complex parser, knowledge base

Introduction

- Identify hyponymy relation

- Finding constructions that indicate the relation of interest (frequent, reliable)
- Identify a set of lexico-syntactic patterns

- I. They occur frequently and in **many text genres**.
=> many instances of the relation
- II. They (almost) always indicate the **relation of interest**.
=> information will not be erroneous
- III. They can be recognized with **little or no pre-encoded knowledge**.
=> not require the tools that help build



NP_0 such as $\{NP_1, NP_2, \dots, (and|or)\} NP_n$

Introduction

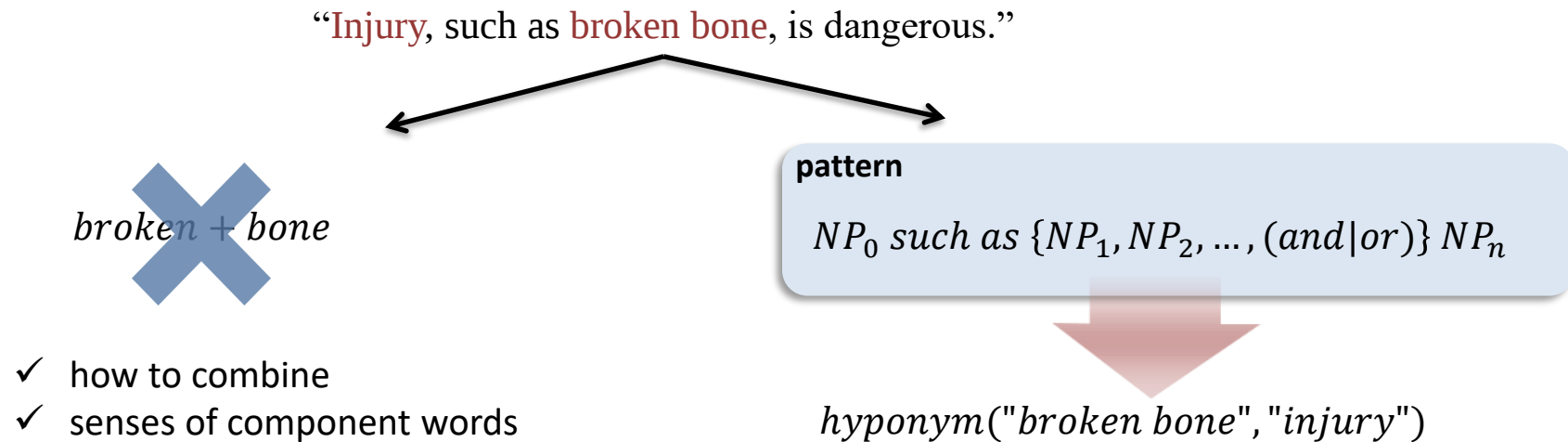
- Usage hyponymy relation

- Lexicon Augmentation

- **Augment** and **verify** existing lexicons, including ones built from MRDs (WordNet, thesaurus etc.)

- Noun Phrase Semantics

- Identification of the general **meaning of an unfamiliar noun phrases**.



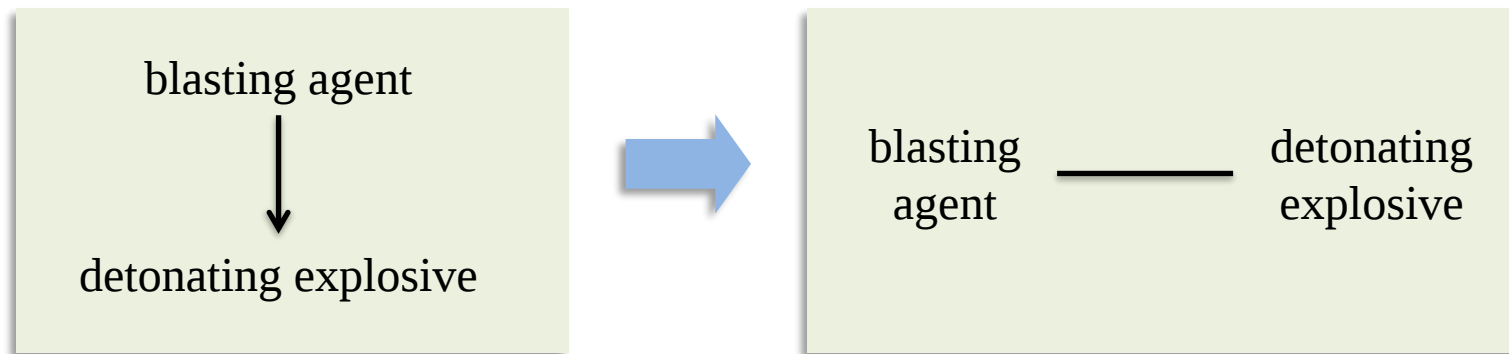
Introduction

- Usage hyponymy relation

- Sematic Relatedness Information

- Relations as **closely** related **semantically** instead of as hyponymic.

hyponym("detonating explosive","blasting agent")



- Not ISA relation. but, imply that the terms' meanings are close.
 - Useful for improved **synonym expansion** and for finding **chains of semantically related phrases**.

Lexico-Syntactic Patterns for Hyponymy

- Hyponym relation from patterns

- Patterns

- (1) *NP such as {NP , }*{(or|and)} NP*
- (2) *such NP as {NP , }*{(or|and)} NP*
- (3) *NP {, NP}*{, } or other NP***
- (4) *NP {, NP}*{, } and other NP*
- (5) *NP {, } including {NP , }*{(or|and)} NP*
- (6) *NP {, } especially {NP , }*{(or|and)} NP*

Sentence

Bruise, wounds, broken bones or other injuries...

Pattern

(3) *NP {, NP}*{, } or other NP*

Hyponym relation

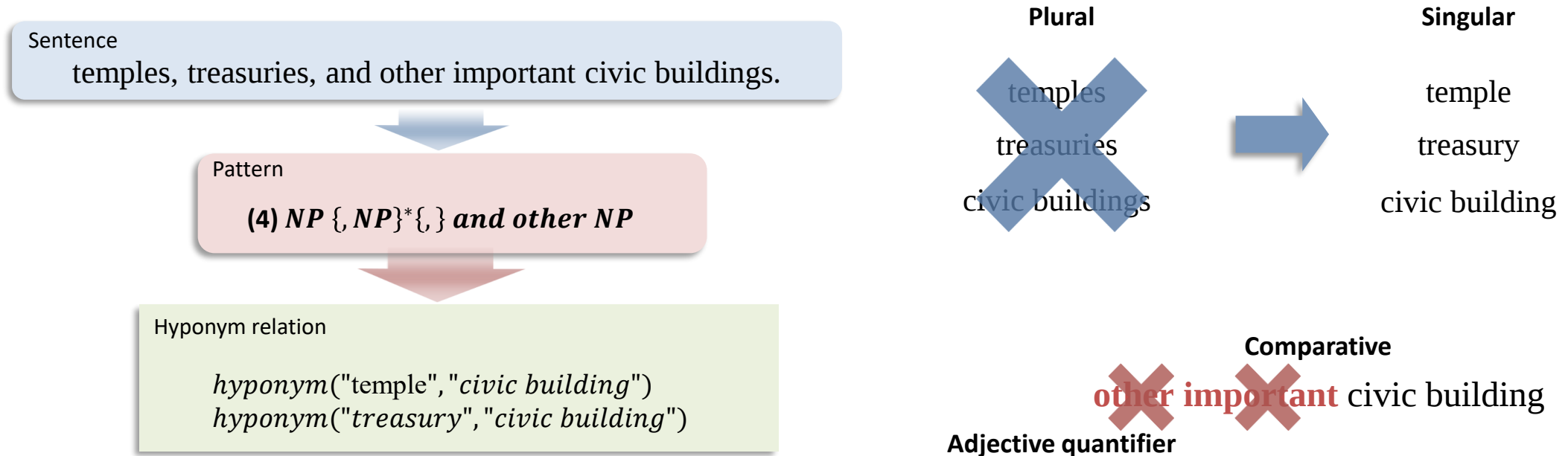
hyponym("Bruise", "injury")
hyponym("wound", "injury")
hyponym("broken bone", "injury")

NP isn't broken down, analyzed

Lexico-Syntactic Patterns for Hyponymy

- Considerations

- Record form



- Almost **plural** form, but want to record **singular** form
- **Adjectival quantifier** can be eliminated. (“some”, “other”)
- **Comparative** can be eliminated. (“important”, “smaller”)

Lexico-Syntactic Patterns for Hyponymy

- Discovery of New Patterns

- Procedures

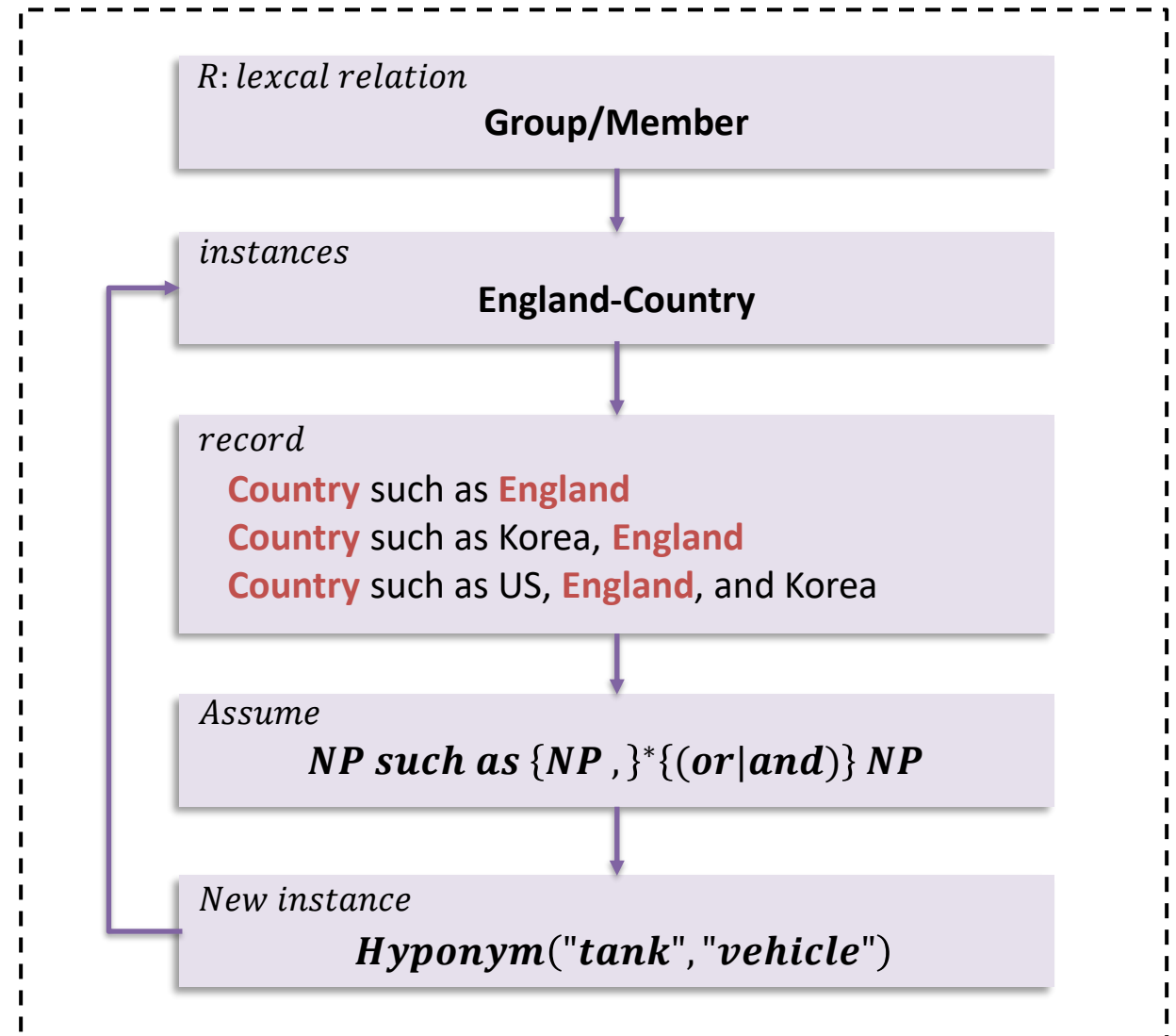
- ① Decide **R** (lexical relation)
- ② Gather list of terms(**instances**)
- ③ Find **environment** that expressions occur near
- ④ Assume **patterns** that indicate the R
- ⑤ **Use new patterns** to gather more instances

- Result

- England-Country to **(1) ~ (5)** patterns.
- Tank-Vehicle to **(6)** pattern.

- Other task

- For **Meronymy**, not successful.
- Some success at identification of **more specific relation**.



Incorporating Results into WordNet

- Validate acquisition method

- Compare results with WordNet

- WordNet is a hand-built thesaurus.
- Possible 3 outcomes after validation

- Outcomes

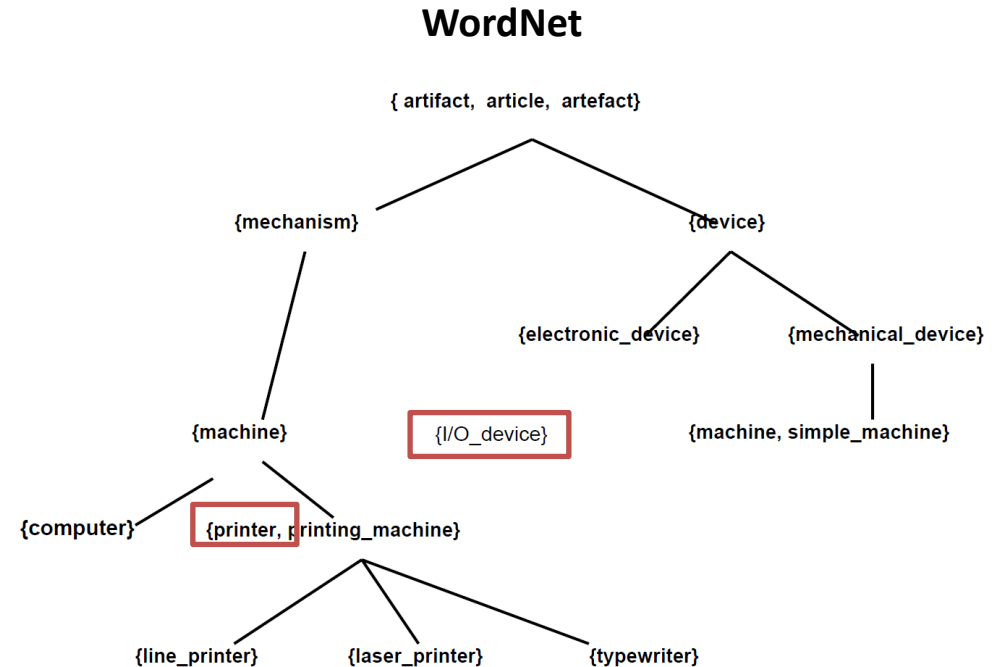
- **Verify** : N_0, N_1 in WordNet. $hyponym(N_0, N_1)$ in **hierarchy**.
- **Critique** : N_0, N_1 in WordNet. $hyponym(N_0, N_1)$ *not* in **hierarchy**.
- **Augment** : one or both of N_0, N_1 are *not* in WordNet.

Critique

Sentence

Other input-output devices, such as printers, color plotters...

$hyponym(\text{"printer"}, \text{"input - output device"})$



Incorporating Results into WordNet

- Results and Evaluation

- Results

Encyclopedia text (8.6M words)

- 7,067 sentences contain “such as”.
- 152 relations fit the patterns.
 - 180 out of 226 unique words existed in the hierarchy.
 - 61 out of 106 feasible relations were found.
- 330 relations fit the patterns. (ease restriction)

New York Times text (20M words)

- 3,178 sentences contain “such as”.
- 46 relations fit the patterns.

- Some Problems

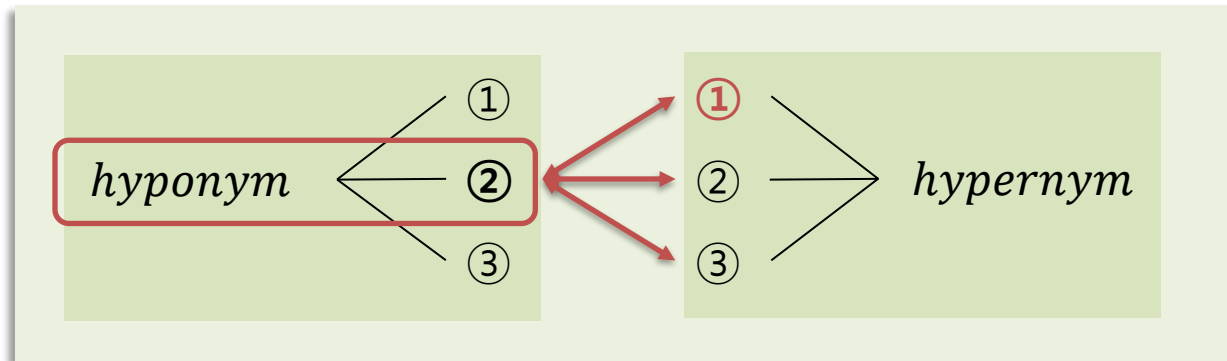
- **Metonymy** - *hyponym("king", "institution")*
- **Underspecification** - *hyponym("steatornis", "species"), hyponym("plot", "device")*
- **Context dependent** - *hyponym("aircraft", "target")*
- **High-level terms** - *hypernym(substance), hypernym(form)*

Incorporating Results into WordNet

- Automatic Updating

- How to automatically insert relation in hierarchy

- Both lexical expressions are present in the noun hierarchy, but **have more than one sense**.
 - Assume hyponym has only one sense, but hypernym has several.
 - Determining which hypernym link to the hyponym. – **use lexical disambiguation algorithm.**



- One or both terms are *not* present in hierarchy.
 - Determine **synsets** that belong terms.
 - Do **sense determination algorithm** above.

Conclusion

- **Automatic Acquisition of Hyponym**
 - Usage of Hyponym relation
 - Lexicon augmentation, NP semantics, Semantic relatedness information
 - Patterns for acquisition of hyponym relation
 - Some patterns with examples
 - Some considerations when recording NP terms and expressions
 - Automatic procedures to discover new patterns that indicate interest relation
 - Low-cost approach than previous (statistically based approach)
 - Require a single instance of interest relation.
 - Evaluate with WordNet
 - Verify, Critique, Augment
 - Use encyclopedia, New York Times text and get successful result
 - Some problems in relation

References

- [1] : Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Coling 1992 volume 2: The 15th international conference on computational linguistics*.
- [2] : Alshawih, H. (1987). Processing dictionary definitions with phrasal pattern hierarchies. *Computational Linguistics*, 13, 195-202.
- [3] : Markowitz, J., Ahlswede, T., & Evens, M. (1986, July). Semantically significant patterns in dictionary definitions. In *24th Annual Meeting of the Association for Computational Linguistics* (pp. 112-119).
- [4] : Nakamura, J. I., & Nagao, M. (1988). Extraction of semantic information from an ordinary English dictionary and its evaluation. In *Coling Budapest 1988 Volume 2: International Conference on Computational Linguistics*.
- [5] : Ahlswede, T., & Evens, M. (1988, June). Parsing vs. text processing in the analysis of dictionary definitions. In *26th Annual Meeting of the Association for Computational Linguistics* (pp. 217-224).

Thank you