

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing

Kiyonori Ohtake[†]

Kazuhide Yamamoto^{†‡}

[†]ATR Spoken Language Translation Research Laboratories, Japan

[‡]Department of Electrical Engineering, Nagaoka University of Technology, Japan

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 1

Objective

To answer the questions:

1. How many paraphrases that are automatically extracted can be directly applied to paraphrasing?
2. What types of paraphrasing will be achieved by the extracted examples?

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 3

Background

- Paraphrasing is translation
⇒ Example-based paraphrasing
- Extract paraphrases from bilingual corpora
- How many paraphrases are desirable, valuable?

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 2

Target of analysis

1. Adjectival predicate → non-adjectival one
2. Passive form → non-passive form

Adjectival predicate and passive formed sentences are

- frequently used, and
- play several functional roles
→ express several meanings

Happy if we can paraphrase them into other expression.

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 4

Results - applicable paraphrases

	adjectival predicate	passive forms
(A)	56.45%	60.62%
(B)	33.69%	37.53%
(C)	9.86%	1.86%

- (A): context independent
(B): depend on context
(C): improper

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 5

Results - types of paraphrases

	adjectival predicate	passive forms
Lexical	30.20%	12.59%
Syntactical	17.18%	8.50%
Lexical and Syntactical	15.54%	6.80%
Semantic	37.08%	72.11%

Desirable for ex.-based paraphrasing

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 6

Conclusion

- How many paraphrases are directly applicable?
Almost **60%** (both adjectival predicates and passive forms)
- What types of paraphrases are extracted?
Desirable examples:
adjectival predicates - 35%,
passive forms - 70%

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing - p. 7