

# Measuring the effects of paraphrasing on readability (proposal)

Michiko Nakamura

COE postdoc

Computational Linguistics Laboratory

1

# Written texts

Primary means of transmitting information

- Information sources (books, newspapers, Internet, etc.)
- Interpersonal communication (e-mails, chat, letters, etc.)



2

# Readability assessment

- If our goals in this ubiquitous society are
  - to enhance information exchange and facilitate internet use
  - to make information available for various users (e.g. elderly people, children & second language learners, people with disability, etc.)
- Examining the optimal format and content of a text is an essential task
- Focus: readability assessment of a text

3

# Paraphrasing (2/2)

For readers (internet users)

- Simplify texts
- Comprehension aid / reading assistance

e.g.) Lexical paraphrasing

- Mary burst into tears and ran out of the room
- Mary cried and ran out of the room

e.g.) Structural paraphrasing

- It is John that sold the textbook to Mary
- John sold the textbook to Mary

more frequent

more familiar

Structurally simpler

5

# Paraphrasing (1/2)

- Provides alternative expressions for a given phrase/sentence
  - Paraphrasing is used:
    - For machines (e.g., Q&A, MT)
      - Q: How late is Daiei open?
      - Q: What time does Daiei close?
- 

4

# Paraphrasing as comprehension aid

Assumption

Paraphrased sentences are simpler, and thus easier for the reader

Questions to be asked

- Are paraphrased sentences actually easier to read than original sentences?
- Can we test the effect of paraphrasing empirically?

6

## Previous studies

- ◆ Far more attention to paraphrase generation
- ◆ Very few studies on evaluating the effect of paraphrasing on readability
  - User-specific readability criteria
    - ♦ Hearing impaired people (Inui & Yamamoto 2001; Inui 2001; Inui et al. 2003)
    - ♦ English-speaking aphasics (Canning & Taito 1999; Carroll 1998)
  - Subjective readability judgment data (Inui & Yamamoto 2001)
  - Dealt with paraphrased sentences in isolation (See <http://cl.naist.jp/lab/kura/ParaCorpus/> for more complete list of bibliography)

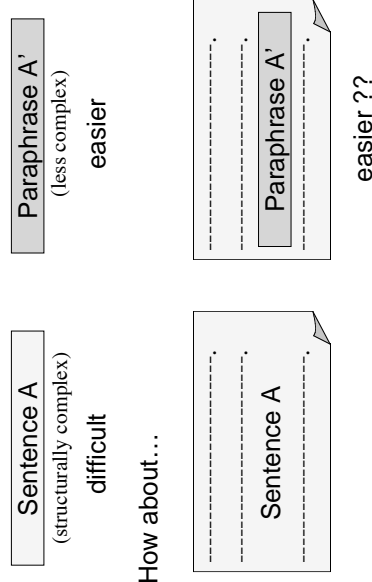
7

## Remaining issues

- ◆ Certain structures are difficult to other users as well (e.g., children, L2 learners, even adults, etc.)
  - 'more general' readability criteria?
  - future application to the Internet
- ◆ Still not clear as to whether:
  - paraphrased sentences are actually easier
  - the effect can be observed during reading
- ◆ Discourse-level effect of paraphrasing

8

## Discourse-level effect



9

## Discourse-level effect (example)

(Inui & Nogami 2001)

- ◆ Småland, which is located to the south-west of Stockholm, is called “The Kingdom of Glass”. The reason is that there are 16 glass manufacturers in this area.

Småland is called “The Kingdom of Glass”.  
It is located to the south-west of Stockholm.  
The reason is that there are 16 glass manufacturers in this area.

10

## Present study (proposal)

The details will be presented in the talk.

11

## Limitations & future work

- ◆ Beginning of a series of studies
- ◆ Needs investigation
  - Wide range of paraphrasing types (structural, semantic, lexical, etc.)
  - Their interaction effects
  - Wide range of users
- ◆ Issues
  - Do we need different readability criteria for different users? (e.g., Inui & Yamamoto 2001)
  - ‘Readability’ may be specific to:
    - ♦ Users, purpose of reading, genre of texts, etc.
- ◆ Implementation in NLP

12

---

## To summarize

- ◆ Assessing the readability of a text is an important issue
- ◆ Focus on 'paraphrasing' that has a potential realizing an optimal content of a written text for various users
- ◆ Propose an experimental study that is designed to test the effect of paraphrasing on the readability of a text
- ◆ The results will give us important clues as to what kinds of factors affect text readability & thus need to be considered in creating an optimal content of internet documents