# Eye-tracking experiments on reading comprehension of Swedish clinical text

**Gintarė Grigonytė [a], Kristina Nilsson Björkenstam [a], Maria Kvist [bc],**
**Ellen Marklund [a], and Lena Renner [a]**

[a] Department of Linguistics, Stockholm University
[b] Department of Clinical Immunology and Transfusion Medicine, Karolinska University Hospital
[c] Computer and System Sciences, Stockholm University

{gintare, kristina.nilsson, ellen, lena.renner}@ling.su.se,
maria.kvist@karolinska.se

## Abstract

We describe an eye-tracking experiment on clinical text simplification. We report first results on the impact of various simplification procedures.

## 1. Introduction

We describe an eye-tracking experiment on clinical text simplification. We report first results on the impact of various simplification procedures: terminology simplification, abbreviation expansion, and syntactic reconstruction. This research is motivated by increasing demand for text simplification following the Swedish government's strategy for National eHealth, which involves giving citizens on-line access to their clinical records.

Clinical text represents a specialised professional domain. Its language is packed with non-standard syntax, excessive use of abbreviations and great spelling variation. This style has evolved as an efficient means of communication between health care professionals, but presents challenges for laypersons.

Our long term goal is to devise text simplification methods for clinical text. To this end we are conducting an eye-tracking experiment to quantify the cognitive load incurred by different linguistic information

## 2. Design and procedure

We have selected sentences from anonymised clinical daily notes from the Stockholm EPR Corpus[1] (Dalianis et al, 2012). A total of 103 sentences were used as eye-tracking stimuli followed by comprehension questions (yes/no).

We have divided the original sentences into 2 subgroups: easy and difficult. We have created a simplified equivalents for all difficult sentences, by expanding abbreviations, simplifying terminology and reconstructing syntax. For example:

1. Original sentence: *Försörjer sig p.o.*
En. Lit. "Feeding p.o."

*2.* No abbreviation: *Försörjer sig per os.*
En. Lit. "Feeding per os."

3. Modified syntax: *Patienten försörjer sig p.o.*
En. Lit. "The patient is feeding p.o."

Simplified equivalents were also created for simple sentences with omitted subjects and verbs.

Each participant was provided with a randomized sequence of 50 sentences with the constraint that the original and simplified version of the same sentence will not be shown during the same trial.

### 2.1 Subjects

29 layperson participants: researchers (14), students (11), administrators (4) between the ages of 22 and 57; 13 male, 16 female.

### 2.1 Apparatus

Eye movements were recorded by using a Tobii T120 eye-tracking system. Stimulus was presented via E-prime 2.0 software. Tobii Studio 3.2.0 was used for recording and analysis.

## 3. Results

The full paper presentation will contain a detailed analysis of the impact of various simplification procedures: abbreviation expansion, simplified terminology, and syntax reconstruction. We compare these 3 types of simplifications with 2 original categories (easy and difficult). Due to space limitations, we present the summary of comprehension results for each of the categories in figure 1. Overall abbreviation expansion, simplified terminology, and syntax reconstruction makes for better comprehension as compared to the unmodified sentences. However the simplified terminology has a slightly negative effect.

---

# References

Dalianis, H., Hassel, M., Henriksson, A., & Skeppstedt, M. (2012, October). Stockholm EPR corpus: a clinical database used to improve health care. In *Swedish Language Technology Conference* (pp. 17-18), Lund.
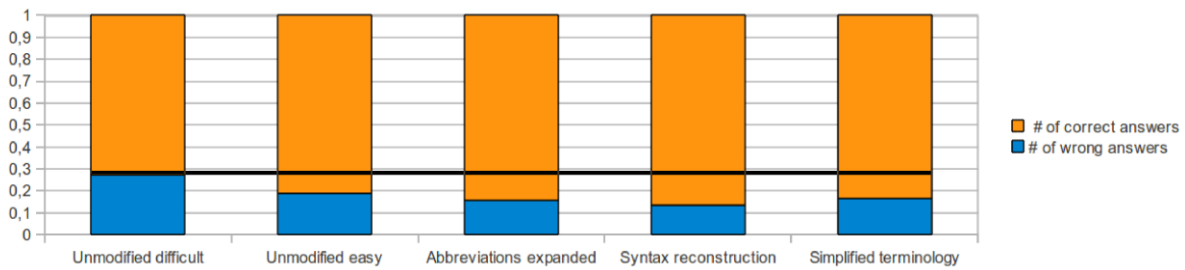
*Figure 1: Summary of comprehension results.*

| Measure | Value | Comment |
|---|---|---|
| BLEU | 20.85 | A very interesting comment |
| NIST | 5.95 | Some other interesting comment |
| METEOR | 45.62 | Yet another one |

Table 2. Table caption of the big table

## 4. References in text

All references within the text should be placed in parentheses containing the author's surname followed by a comma before the date of publication (Wierzbicka, 1987). If the sentence already includes the author's name, then it is only necessary to put the date in parentheses : "Wierzbicka (1987) claims that …". When several works are cited, those references should be separated with semicolons: (Wierzbicka, 1987; Artstein & Poesio, 2008; Stymne, 2008). When the reference has three or more authors, only cite the name of the first author followed by et al. (Megyesi et al., 2008).

Bibliographical references should be listed in alphabetical order at the end of the article. The title section, "References", should be an unnumbered level 1 heading. The first line of each bibliographical reference should be justified to the left of the column, and the rest of the entry should be indented. The following examples illustrate the basic format of references.

## Acknowledgements

The paper may contain an unnumbered acknowledgement section.

## Reference

R. Artstein and M. Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34(4): 555-596.

B. Megyesi, B. Dahlqvist, E. Pettersson, S. Gustafson-Capková, and J. Nivre. 2008. Supporting research environment for less explored languages: A case study of Swedish and Turkish. In J. Nivre, M. Dahllöf, and B. Megyesi, editors, *Resourceful Language Technology: Festschrist in Honor of Anna Sågvall Hein*, Studia Linguistica Upsaliensia 7, pages 96-110. Acta Universitatis Upsaliensis.

S. Stymne.. 2008. German compounds in factored statistical machine translation. In A. Ranta and B. Nordström, editors, *Proceedings of GoTAL, the 6th International Conference on Natural Language Processing*, pages 464-475, Gothenburg, Sweden. Springer.

A. Wierzbicka. 1987. *English Speech Act Verbs: A semantic dictionary*. Academic Press, Sidney.