

Vocabulary In Discharge Documents

The Patient's Perspective

Veronika Laippala¹, Riitta Danielsson-Ojala^{2,4}, Heljä Lundgrén-Laine^{2,4}, Sanna Salanterä^{2,4}, and Tapio Salakoski³

¹Department of French Studies

²Department of Nursing Science

³Department of Information Technology

20014 University of Turku, Finland

⁴ Southwest Hospital District, Turku, Finland

first.last@utu.fi

Abstract. Medical discharge documents are summaries written by a physician about the patient's condition and aim at transferring information to other health care personnel but also to the patient. According to the legislation, the patient should be able to understand the document. In practice, however, this has been shown to be problematic.

This paper studies discharge documents from the patients' perspective and examines how they fulfil the legislation's demands on understandability. Concentrating on the vocabulary of the texts, we analyse the frequency of domain-adapted terms, abbreviations and foreign words. The material consists of 23 528 heart patients' discharge documents (5 747 126 words). The analysis is performed with the morphological analyser FinTWOL (<http://www2.lingsoft.fi/cgi-bin/fintwol>).

Altogether, FinTWOL analyses 24% of the corpus as unknown or foreign words, abbreviations or medical terms. The most common category, unknown words, includes misspellings and medical terms, such as *l.dex*. Of these, 100 most common cover for 43% of the total. These terms thus seem to be relatively fixed. Of the words analysed as abbreviations, some are common also in standard language, but others are still very domain-specific, such as *I.V.* (*intravenous*). Also the used abbreviations are very fixed: the 100 most common ones cover for 94% of the total. This, however, does not help the patient who probably reads only one document. Similarly, even though misspellings are globally infrequent, they still occur more than once per document.

In order to place the obtained results in a context, we performed a similar analysis on general Finnish university newspaper text from Turku Dependency Treebank. In comparison with the 24% obtained with the discharge documents, from the total of 10 687 words, 8,6% were given a special tag.

The results show that that terms and abbreviations are considerably more used in discharge documents than in general newspaper text. It is clear that a text with such a vocabulary is domain-specific and distinct from the language that the patient is used to. Also e.g. the varying use of upper and lower case letters (*dg* and *DG* for diagnosis) emphasize the particularity of the language. In

standard language texts such writing would not be acceptable. Standard writing would, however, help the patients to better understand the texts.

Keywords: Discharge Documents; Medical Language; Medical Terms; Patient's Comprehension

1 Introduction

Medical discharge documents are care summaries written by a physician about the patient's condition, medication, progress of the illness and continuation of care. They are a part of electronic patient record and aim at transferring information to other physicians and health care personnel but also the patient. According to the legislation (www.finlex.fi), the patient should be able to read and understand the document; the information should be explicit and intelligible, and only generally known and accepted abbreviations should be used.

However, previous studies have shown problems in the functioning of discharge documents: they do not necessarily reach their users on time or their quality may be poor.^{1,2} As many texts written in a medical context, discharge documents seem to be dense, telegraphic and contain frequent medical terms, abbreviations and misspellings.³⁻⁵ Even though these features probably do not present problems for physicians, they are difficult for the patients, who are not familiar with the domain.

This paper studies discharge documents from the patients' perspective and examines how they fulfil the legislation's demands on understandability. Concentrating on the vocabulary of the texts, we analyse the frequency of domain-adapted terms, abbreviations and foreign words in the documents in order to discover the proportion of medical language used.

2 Materials and Methods

The material consists of 23 528 heart patients' medical discharge documents from the years 2005-2009 from a Finnish hospital, covering 5 747 126 words, punctuation excluded. In order to study the vocabulary, the material was analysed with the morphological analyser FinTWOL (<http://www2.lingsoft.fi/cgi-bin/fintwol>). In addition to the morphological reading(s), the tool identifies abbreviations and (some) foreign words. This is particularly useful in the present study, as these words are potentially problematic for the patient.

3 Results

First of all, FinTWOL analyses 10% of the discharge documents' words as unknown. These include misspellings and (often abbreviated) medical terms, such as *l.dex*, *l.sin* and *Dgn*. In fact, of these, 100 most common cover for 43% of the total. It thus seems that these terms are relatively fixed. On the other hand, only 6% of the unknown words occur only once and are the most potential misspellings.

Further, FinTWOL analyses 7% of the words as abbreviations. Some of these are common also in standard language, such as *mm.* or *l.*, but others are still very domain-specific, such as *I.V.* (*intravenous*) and *MCC* (*Morbus Cordis Coronarius*). Similarly to unknown words, also the used abbreviations are very fixed: the 100 most common ones cover for 94% of the total. This, however, does not help the patient who probably reads only one document. Similarly, even though misspellings are globally infrequent, they still occur more than once per discharge document.

In addition, 5% of the words are tagged as proper names. These include doctors' names and drugs, such as *Furesis*, but also some Latin terms are given this tag, even though FinTWOL has also a special tag for foreign words. This FORGN tag is given to another 2% of the words, which are in our case Latin terms used to describe for instance the status or the diagnosis of the patient: *diabetes mellitus II* (*diabetes of the adulthood*).

The analysis shows that abbreviations, terms and Latin words are frequent in discharge documents; they cover on average 24% of the texts. In order to place these numbers in a context, we performed a similar analysis on general Finnish university newspaper text from Turku Dependency Treebank.⁶ From the (modest) total of 10 687 words, 0,2% were analysed as foreign, 0,4% as abbreviations, 3% were unknown and 5% were tagged as proper names.

4 Discussion

The results show that terms, abbreviations and Latin words are considerably more used in discharge documents than in general newspaper text. Even if some of these words did not present comprehension problems for the patient, it is clear that a text with such a vocabulary is domain-specific and distinct from the language that the patient is used to. In addition, the varying use of upper and lower case letters in abbreviations and spelling variants, such as the use of several abbreviations for one term (*Dg.* and *Dgn* for *diagnosis*), emphasize the particularity of the language. In standard language texts, where normative requirements must be followed, such writing would not be acceptable. In fact, this variation associates discharge documents with an informal register that may to some extent be understandable to the patient but not necessarily appropriate considering the text context.

5 Conclusion

The results stress the fact that discharge documents are written fast and mostly aimed at other physicians, not at patients. Normative, standard language writing would help the patients to understand the texts and therefore increase their means to actively participate in their own care. In addition, it would prevent possible misunderstandings between professionals in different units.

This study has focused only on the vocabulary of the texts. Therefore, the results are obviously very limited as comprehension involves also other aspects, such as the structure and syntax of the text.⁷ In order to study this, we are developing a domain-

adapted parser similar to the one previously developed by Haverinen et al.⁸ for intensive care patient reports.

Finally, another obvious direction for future research would be the application or development of language technology tools to assist in the communication between the physician and the patient. For instance, term search and abbreviation expansion would be very helpful as our analysis shows that the used terms and abbreviations are relatively fixed.

Acknowledgements

We are grateful to Juho Heimonen and Antti Airola for technical assistance and to Lingsoft Ltd. for making FinTWOL available for us. This work has been supported by the Academy of Finland.

References

1. Tallgren M: Epikriisin sijaan potilaalle lyhyt ja ytimekäs hoitoyhteenvedo? *Suomen Lääkärilehti* 41(2007).
2. Kripalani S, LeFevre S, Phillips CO, Williams MV, Baker DW: Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. *JAMA* 2007(297):831–841.
3. Laippala V, Ginter F, Pyysalo S, Salakoski T: Towards Automated Processing of Clinical Finnish: Sublanguage Analysis and a Rule-Based Parser. *International Journal of Medical Informatics* 78(12):2009:e7-e12.
4. Hobbs P: The role of progress notes in the professional socialization of medical residents. *Journal of Pragmatics* 2004(36):1579–1607.
5. Tiirilä U: Auttajista lausuntoautomaateiksi? Lääkäreillä keskeinen rooli etuuksista päätettäessä. *Duodecim* 124(2008):896–901.
6. Haverinen K, Ginter F, Laippala V, Kohonen S, Viljanen T, Nyblom J, Salakoski T: A Dependency-based Analysis of Treebank Annotation Errors. *Proceedings of International Conference on Dependency Linguistics (Depling'11)*, Barcelona, Spain, 2011:115-124.
7. Herring SC: Grammar and electronic communication. In CA Chapelle (ed.): *Encyclopedia of Applied Linguistics*. Hoboken, NJ: Wiley-Blackwell, 2012.
8. Haverinen K, Ginter F, Laippala V, Salakoski T: Parsing Clinical Finnish: Experiments with Rule-Based and Statistical Dependency Parsers. *Proceedings of NODALIDA'09*, Odense, Denmark, 2009:65-72.
9. <http://www.finlex.fi/fi/laki/ajantasa/2009/20090298>