The New Navigators: from Professionals to Patients
R. Baud et al. (Eds.)
IOS Press, 2003

About the Language of Hungarian Discharge Reports

György Surján[a] Gergely Héja[b]

[a] *National Information Centre and Library of Medicine*
[b] *Budapest University of Technology and Economics*

Abstract

Concerning use of terminological systems translated to any particular language or computer assisted translation of patient records the daily medical language has to be taken into account rather than the pure academic language. The aim of this study to investigate the language of the diagnoses of the Hungarian hospital discharge reports. Word of a collection of discharge diagnoses were categorised according to the language and form (normal words, abbreviations, acronyms). It was found that Latin is still a dominant, but the language is rather a mixture, where local jargon, non standardised abbreviations occur frequently. Some signs show, that discharge reports are not used primarily as a tool of communication of physicians, rather serve local administrative purposes.

Keywords:
language, discharge report, diagnosis, patient record

1. Introduction

What is the language of the discharge report diagnoses used in Hungarian hospitals? This apparently simple question has some practical and a theoretical aspect. One of the practical aspects is related to the translation of various vocabularies, coding systems thesauri etc. In such situations we are faced with the problem that a pure Hungarian translation does not meet with the language used by the Hungarian medical community in the daily practice. This was the case e.g. with the translation of ICD-10. The discrepancy between the living language and the official Hungarian ICD causes difficulty in building computer assisted coding systems. Another practical aspect is related to the coming accession to the European Union, where the free mobility of people generates a need to make available and intelligible patient information in foreign language environment. In this situation again the daily medical language should be the source (or target) language, not the pure Hungarian.

From theoretical aspect we can say, that any language is product of a community characterised by a shared part of knowledge and by the frequent information exchange within the members of the given community or population. Without shared knowledge lingual communication would be impossible. Without frequent communication the language starts to diverse. (Like Portuguese in Brazil and Portugal. The problem is discussed by Umberto Eco [1]) For that reason, studying the language of a subpopulation, like the Hungarian medical community, may clarify their common knowledge and communication patterns. The language of course depends on the actual way of communication also: we use different language for written and oral and also for colloquial and official communication. For the recent study we selected a rather specific layer of communication: we studied the language of medical discharge reports, more particularly the language of clinical diagnoses as they appear on the discharge reports. Our original goal was to enhance our computer assisted coding tools, but the study – as we try to show it in the followings – revealed other surprising results.

2. Material and method

A collection of 3079 clinical diagnoses collected from discharge reports from a hospital (the former Haynal University of Health Sciences, recently National Healthcare Centre) was separated into words (i.e. space delimited strings). We got a list of 2940 different strings. In this list 54 string were symbols, like numbers, hyphens etc. The remaining 2886 words were analysed. All words were categorised according to their language and type. The type categories were the followings:

- Ordinary words (e.g. "anaemia")
- Abbreviations (e.g. "tons." Which stands for "tonsillits")
- Acronyms (e.g. "COLD" which stands for "Chronic Obstructive Lung Disease")

The difference between abbreviation and acronym is that the later comprises more than one word. We extracted the list of acronyms and counted the number of words they refer to. The majority of acronyms comprises three words (WPW, SLE etc.) the minimum was two (by definition) and the maximum was 6. Their average was very close to three (3.047).

In the language categorisation we used the following categories:

- Latin
- Hungarian
- English
- German
- French
- Mixed
- Personal
- Technical
- Indefinable
- Unknown

Words written in Latin orthography were considered as Latin, regardless to the fact, that some of these words are Greek in origin (e.g. "ischaemia").Proper names (like Hodgkin, Tawara) were treated as a separate category (Personal), regardless to the nationality of the person. Scientific names of micro-organisms, chemicals, drugs were also classified into a separate category, technical terms, since such expressions can not be translated (e.g. "Borrelia"). We set up a "mixed" language category for words with Latin word stem and Hungarian suffix or combined words with components of various languages. (E.g. "adhaesiók" =adhesion in Latin+'k' sign of plural in Hungarian; or "aortaív" for aortic arch.) The distribution of the used different words (disregarded the frequency of each word) and the distribution of tokens (the number of occurrences of each word) was studied as well.

3. Results

The distribution of words is summarised in Table 1. As the traditional Hungarian medical language was based on Latin, it is not surprising, that the majority of words (70%) is still Latin. The high number of proper names and technical terms is also natural in such kind of text. But physicians today can not express themselves in Latin in a sufficiently flexible way. They use the Latin terms as they learned during their study, but they are not able to create new Latin expressions for the newly emerging concepts, and sometimes they have problem with the proper inflections. E.g. `"Granuloma filii mammae"` means *'granuloma of the son of the breast'* but the intended diagnosis was *'Granuloma fili mamme'* which means *'granuloma caused by surgical thread in the breast'*. This fact explains the

significant number of Hungarian and English terms. (English terms came mostly from the scientific literature). The lack of proper knowledge in Latin also explains the presence of the mixed words: Latin roots with Hungarian suffixes. So we can say that the language of Hungarian discharge diagnoses is a special mixture of languages with Latin dominancy. This fact must be considered in case of translation of medical thesauri (coding systems, controlled vocabularies etc.) or patient information from any European language to "Hungarian" Neither pure Latin nor pure Hungarian would not fit with the common medical language.

But here we are facing with an important question: is this language a standard mixture of various languages, or does it show local or individual variations? The same concept is expressed always in the same language or not?

	abbreviation	acronym	word	Total
mixed	5	31	177	**213**
German			1	**1**
English	1	13	55	**69**
French		1	5	**6**
Latin	454	33	1558	**2045**
Hungarian	9	6	373	**388**
Personal	1	3	70	**74**
technical	1	4	30	**35**
unknown	6	33	16	**55**
Total	**477**	**124**	**2285**	**2886**

Table 1 Distribution of words according to type and language

We found, that in many cases the proper Latin form of a Hungarian or "mixed" word is also present in our sample. Some of the English terms are never used either in Latin or in Hungarian (e.g.: `arrest, attack, dumping, kinking, outlet, sustained`), while some others are also present in Latin. These are usually words of Latin origin, written in English orthography: `cardiac, thoracic.` So there is a large heterogeneity in the usage of Latin and Hungarian forms. The usage of the English terms is more consistent but still not uniform. We have similar experience with the usage of abbreviations and acronyms. More than 20% of the total number of words are abbreviations or acronyms. Some of them are also used in their full form (abd. – abdomen; bicup. – bicuspidalis; diab – diabetes etc.), while others are always used in their short form only.

Most abbreviations are colloquial, like the notation of laterality: "l.d." for right, "l.s" for left side from the latin "lateris dextri, lateris sisnistri." We found that the same word may have many abbreviated form (`"prog."`; `"progr."`; `"progred."` for `"progrediens"`), one abbreviation might have more than one meaning (`"inf."` may stand for 'inferior', 'infectio' or 'infarctus') But the most surprising experience was, that we have found some abbreviations which are so incredible hotchpotch of terms and languages, that we wonder if anyone could understand it except the very local people. Let we present two examples:

`"coronbet"` stands for coronary disease, where "coron" comes from coronaia (Latin) while "bet" is the first syllable of the Hungarian word for disease ("betegség")

`"J.Twbl"` stands for "right Tawara's bundle block, where "J." is the Hungarian abbreviation of 'right', "Tw" is the abbreviation of 'Tawara" and bl is the abbreviation of the English word 'block'. While abbreviations and acronyms cover only approximately 20 % of

all words, nearly 43 % of sentences contain at least one abbreviation or acronym (beyond the most colloquial abbreviation of laterality).

Number of tokens	abbreviation	Acronym	word	Total
mixed	11	45	224	**280**
Germain			1	**1**
English	1	24	126	**151**
French		2	5	**7**
Latin	2823	94	4690	**7607**
Hungarian	11	7	658	**676**
Personal	1	5	93	**99**
technical	2	4	36	**42**
unknown	469	147	202	**818**
Total	**3318**	**328**	**6035**	**9681**

Table 2 Distribution of tokens (words with their frequency) according to type and frequency

Table 2 summarizes the distribution of tokens in the sample. The 2886 different words have 9681 occurrences in total. In most of the categories the average occurrence of the words is in the range from 1 to 4. The only German word was used only once, the French abbreviation was used twice, etc. What is surprising is the high frequency of the abbreviations. The 454 Latin abbreviations occur 2823 times, the 6 unknown abbreviations were used 469 times! It is evident that frequently occurring expressions are likely abbreviated, but the average occurrence of unknown expressions is about 14! Near 10% of the tokens are unknown, which means, that the intelligibility of discharge diagnoses for physicians working in different speciality and in different environment is seriously hindered by use of unclear abbreviations, acronyms and slang.

4. Conclusion and discussion

The authors did not find too much similar work in the literature. There are some studies dealing with proper use of words of various origin (Latin, Greek, etc) in medical documents. [3] Coombs and others studied the use of medical slang from the viewpoint of the individual. [4] Dawson et al. investigated the quality of discharge reports in a paediatric hospital, and state that use of discharge reports in communication was hindered by poor hand writing and the use of abbreviations.

[5]

The language of clinical diagnoses in Hungarian discharge reports is a mixture of various languages in which Latin still has a dominant role. However physicians today are far not fluent in Latin, and use a mixture of languages, perhaps more and more influenced by English. This mixture is not standardised and is essentially influenced by local medical jargon. As we mentioned in the introduction, the language is a social product of a set of people communicating with each other and sharing a common part of knowledge. As we were able to detect many expressions hardly intelligible for GP-s or physicians working in other institutes we may conclude, that discharge reports are not really used as a tool for communication between physicians, but rather as a local documentation tool. As the authors know, in Germany for instance, the discharge reports are written – and sent directly to – mainly the GP-s, traditionally, and only about a decade ago it was decided to hand out a copy to the patient. In Hungarian practice the discharge reports are given normally to the patient itself in most cases. It is the patient's role to give a copy to the GP and preserve an other. However some decades ago in case of serious diseases like malignant tumour – the

physiciants did not want to inform the patient about the real situation. So they found tricks to hide the reality. (Abbreviate the word tumour as "Tu." or use the not so well-known 'neoplasia' instead, or abbreviate as 'npl.') Today patients have the right to be informed about their disease, but still the language of the discharge reports is somehow obscure.

The high number of (English) acronyms presumably comes from the medical literature, where the usage of acronyms is rather colloquial. The problem is that in scientific papers it is usual to define the meaning of the acronyms used within the paper what is practically always missing from patient documentation. Therefore acronyms can be ambiguous or unintelligible for the reader.

In a previous paper [2] we pointed out, that the role of language in medical environment is not only the description and communication, but the direction and control of the local care process. Utterances are frequently used as triggers of certain action. In such cases the shortness of the utterance can be important in order to achieve quick reaction of the members of a health care team. Therefore the semantic correctness is sometimes sacrificed for the effectiveness. What we can see now is that such "shortcuts" appear in the patient documentation, were the role of the language would be mainly the description.

The analysed sample was collected from just one hospital. It would be reasonable to study collection from more institutes and analyse local variations more deeply.

References

[1] Eco U La ricerce della lingua perfetta nella cultura europea Gius Laterza & Figli Rome-Bari 1993 [Hungarian translation Atlantisz 1998 Budapest]

[2] Surjan G, Balkanyi L.On different roles of natural language information in medicine. *Stud Health Technol Inform.* 1997;43 Pt A:411-5.

[3] Rozenfeld LG, Kolotilov NN [Terminology problems in medical thermography] (Russian) Med Radiol (Mosk). 1988 Jan;33(1):56-9.

[4] Coombs RH, Chopra S, Schenk DR, Yutan E. Medical slang and its functions. Soc Sci Med. 1993 Apr;36(8):987-98.

[5] Dawson KP, Capaldi N, Haydon M, Penna AC. The paediatric hospital medical record: a quality assessment. Aust Clin Rev. 1992;12(2):89-93.