

USING TOOLS OF CORPUS LINGUISTICS FOR INVESTIGATING AND TEACHING LANGUAGE FOR MEDICAL PURPOSES

Kateřina Pořizková

Masaryk University Language Centre, Medical Faculty, Brno - Czech Republic

Robert Helán

Masaryk University Language Centre, Medical Faculty, Brno - Czech Republic

Annotation

The aim of this paper is to demonstrate the implementation of a corpus linguistics approach in researching and teaching Latin and English languages for medical purposes. In the former, a corpus of authentic Latin clinical diagnoses was created from healthcare documentation, providing the instructors with valuable data concerning the actual use of Latin in professional medical writing. In the latter, a corpus of English published medical case reports was collected from two online journals, enabling structural and textual analyses of the genre, which were then used pedagogically in academic writing courses. Both corpora were uploaded to Sketch Engine, a tool creating summaries of a word's grammatical and collocational behaviour. It can be claimed that this approach, in which authentic and published materials are used as a basis for the development of teaching materials, greatly enhances the quality and relevance of language education in medical contexts.

Keywords: corpus linguistics, Sketch Engine, Latin diagnoses, English medical case reports

Introduction

Corpus linguistics as a relatively recent linguistic discipline provides modern methods of investigating natural language by means of digitally processed written and spoken authentic texts (Biber et al 1998, Āermák 2006). Various types of specialized electronic software are used for a linguistic analysis of language corpora from different points of view, e.g. frequency of terms, concordances, phrases, or lexico-grammatical constructions. A team of IT specialists at Masaryk University in Brno have collaborated on the development of originally British licensed Sketch Engine software. This software has become a valuable tool for LSP (language for specific purposes) instructors since it offers a fast, precise and effective analysis of language data which can then be applied for research, teaching, and testing.

The paper is divided into two interrelated parts. The first part describes the ways in which the Sketch Engine software has been used for data-driven investigations of authentic Latin diagnoses. The second part focuses on a corpus- and genre-based analysis of medical writing in English, particularly medical case reports. For these aims, the Sketch Engine software was used for a textual and structural analysis of the genre.

Authentic Latin Clinical Diagnoses

It can be rather difficult for the Latin medical terminology instructors to demonstrate to students the use of professional language in clinical settings. The main problem lies in the fact that medical documents are virtually inaccessible due to ethical considerations (i.e. patients' and doctors' personal data). The Latin section of the Masaryk University Language Centre initiated collaboration with teaching hospitals in Brno and Prague, resulting in the formation of a database of authentic Latin clinical diagnoses (Pořizková et al 2013). The database consists of 7 417 diagnoses, the majority taken from surgical case histories, and is currently being transformed into a corpus which is available to the Latin medical terminology instructors for their research or pedagogical applications (e.g. teaching and testing language for medical purposes).

A key problem in teaching clinical terminology is the fact that at present no internationally valid nomenclature exists in the area of medical terminology (with the exception of promising projects such as Snomed CT¹). Latin medical terminology instructors must rely on existing clinical terminology dictionaries, primary and secondary literature of individual medical fields, and samples of authentic healthcare documentation obtained incidentally from family members or friends. The chief purpose of developing a corpus of clinical diagnoses is to enable access to a large collection of

¹ <http://www.ihtsdo.org/snomed-ct/snomed-ct/>

authentic language data which would offer a complex and relevant information on how practising clinicians apply their language knowledge and skills. The following are the main aspects:

- a) frequency of terms
- b) typical collocations (of nouns and adjectives, in particular)
- c) typology of prepositional phrases
- d) linguistic means of expressing modality (probability)
- e) lexico-grammatical features of clinical diagnoses (see Table 1).

Table 1: A sample of Latin clinical diagnoses from Sketch Engine

S9200	Fractura calcanei l. dx. commin., sine dislocatione. Infractio calcanei l. sin.
S300	Contusio reg. lumbalis.
S860	Ruptura tendinis Achillei l. dx.
L031	Phlegmona pedis l. dx.
S0600	commotio cerebri
C19	Ca rectosigmoidei generalisatum cruens.
E115	Gangraena diabetica cruris l.dx.
C20	Ca recti, st. p. radiotherapiam
K509	M. Crohn, infiltratus pericoecalis, abscessus pelvis minoris.

The findings from the corpus-based analysis of authentic Latin clinical diagnoses have been conducive to the creation of innovative, data-driven teaching materials for Latin medical terminology courses. The data obtained from teaching hospitals played a major role in the selection and organization of the lexico-grammatical phenomena in these teaching materials, resulting in a different model than the traditional one used in Latin medical terminology textbooks. Another aspect of these materials is the fact that they have been created within interdisciplinary collaboration with experts from the fields of anatomy and clinical medicine.

Published English Medical Case Reports

The genre of medical case reports (MCRs), a medical recount of a pathological condition in a single patient, can be categorized as one of the three genres of clinical case reporting, the other two being oral case presentations and written case histories. Although clinical case reporting has received some attention from several disciplines such as sociology (Anspach 1988), literary theory (Montgomery Hunter 1992) and linguistics (Taavitsainen and Pahta 2000), the research has been undertaken primarily from critical and diachronic points of view. In this study, carried out primarily for pedagogic purposes, the genre of MCRs is approached in terms of its structural organization and textual characterization. By demonstrating to medical students how the genre is structured and what language is used in the specific sections of the reports, we can help them become more aware of the many complex but often implicit linguistic, disciplinary, and cultural aspects of professional writing in medicine.

For the pedagogic purposes stated above, a main corpus of 40 current MCRs was created, totalling 46,160 words. The two online journals selected for the creation of the corpus were the Cases Journal and the Journal of Medical Case Reports, open-access medical journals specializing in publishing MCRs. Sketch Engine was used for generating lists of word frequencies and concordances from the corpus uploaded into the software from text files. For a more systematic analysis of the genre, four subcorpora of the respective sections were created out of the main corpus, i.e. introductions, case presentations, discussions, and conclusions.

For the analysis of the structure, Swalesian (2004) structural move analysis was used, dividing a given genre into specific rhetorical moves, which can further be subdivided into steps. The following table on the next page demonstrates the division of the MCR into specific moves with examples of typical language used in each move. The moves in the presentation section of the MCR reflect a logical progression of the problem-solution pattern. It must also be born in mind that certain moves and steps tend to recur repeatedly and might be structured in a different order than is suggested in the table, which demonstrates a rather ideal structure of the report.

Table 2: Structural move analysis of all the sections of the medical case report with typical language

MOVE	DESCRIPTION	EXAMPLE OF TYPICAL LANGUAGE ²
THE INTRODUCTION SECTION		
Move 1: Establishing a territory	Contextualizing the report by defining and describing relevant pathological conditions or special procedures	<i>Chronic idiopathic urticaria (CIU) is a common, benign condition...</i>
Move 2: Establishing a niche	Indicating a gap in clinical knowledge and the relevance of the report by invoking its uniqueness and a lack or absence of similar reports	<i>A recent literature search revealed no reported cases of injury...</i>
Move 3: Presenting the present work	Announcing the present report (optional move)	<i>We present a case of accidental carbon monoxide poisoning...</i>
THE CASE PRESENTATION SECTION		
Move 4: Presenting a problem	Providing a clinical identification of the patient by specifying the patient's demographic data and his/her (past medical, social, family, etc.) history	<i>A 24-year old Caucasian Irish male student sustained a laceration ...</i>
Move 5: Investigating the problem	Summarizing significant examination findings and/or results of investigative procedures and determining the diagnosis	<i>Endocrine testing showed that she had normal pituitary hormone levels...</i>
Move 6: Addressing the problem	Describing actions taken to treat the patient by recapitulating any surgical or pharmaceutical interventions	<i>The patient underwent laparoscopic appendectomy and adhesiolysis.</i>
Move 7: Evaluating the outcome	Stating the success or failure of the patient's treatment having resulted either in the patient's survival and recovery or death	<i>Despite eventual clearance of her fungemia, the patient died from ...</i>
THE DISCUSSION SECTION		
Move 8: Presenting background information	Describing (a specific aspect) of the case and citing clinical studies aiding in presenting this information	<i>It is now largely accepted that nephrectomy is the treatment of choice in most patients...</i>
Move 9: Reviewing literature pertinent to the case	Contrasting and comparing present and previous reports of similar cases (or stating a lack or absence thereof)	<i>Only a few reported cases show complete or partial reversal of ...</i>
Move 10: Drawing implications of the reported case	Invoking relevant features of the reported case and suggesting possible clinical, pedagogic, or research implications and recommendations	<i>The present case demonstrates the importance of diagnosing esophageal cancer early...</i>
THE CONCLUSION SECTION (optional section)		
Move 11: Summarizing the reported case	Recapitulating central features of the reported case	<i>We have reported a case of trapezium fracture associated with</i>
Move 12: Summarizing the implications	Discussing the most important implications and recommendations of the reported case	<i>This case highlights the need to be vigilant to other causes of hypoxaemia...</i>

The above-presented structure and language of the genre of MCRs can aid medical students in familiarizing themselves with the ways the genre tends to be rhetorically organized (i.e. via moves) with conventional language being used to express those specific moves. Many genre-awareness activities can be developed in relation to the structural move analysis and conventional language use. Students can themselves analyze a MCR with regard to the typical moves. The move analysis may function as a writing guide to students who are not acquainted with the genre and its conventions. Alternatively, instructors may prepare a so-called 'skeleton' to facilitate the writing process for students. An example of a skeleton for a case presentation would be the following: *A (age)-year-old (race) man/woman presented with a (number)-month/year history of*

² Examples taken from the Journal of Medical Case Reports (<http://www.jmedicalcasereports.com/>) and the Cases Journal (<http://www.casesjournal.com/>)

(pathology/sign/symptom). His/her past medical history included (list of diseases). He/she was treated with (drugs) and subsequently underwent (surgery). Upon discharge, she complained of (symptoms). His/her physical examination revealed (signs). (Testing) was performed. Therapy with (drug) was initiated (when).

Conclusion

In this paper, we intended to demonstrate the usefulness and relevance of a corpus linguistics approach to language education in the areas of professional and academic writing. Not only do corpora enable evidence-based and data-driven language teaching, but they also make the courses more relevant and motivating to students, leading to language instruction which is enhanced in terms of both language and content.

References

1. Anspach R. R. (1988) Notes on the Sociology of Medical Discourse: The Language of Case Presentation. In: Journal of Health and Social Behavior. Vol. 29, No. 4. 357-375.
2. Biber D., Conrad S., and Reppen R. (1998). Corpus Linguistics. Investigating Language Structure and Use. Cambridge: Cambridge University Press.
3. Čermák F. (2006) Korpusová lingvistika dnešní doby (Corpus Linguistics Today). In Korpusová lingvistika: Stav a modelové přístupy (Corpus Linguistics: Its State and Model Approaches). Ed. by Čermák F. and Blatná R. Praha: Nakladatelství Lidové noviny, 9-19.
4. Helán R. (2012) Analysis of Published Medical Case Reports: Genre-Based Study. PhD Dissertation. Brno: Masaryk University.
5. Montgomery Hunter K. (1992) Remaking the Case. In: Literature and Medicine. Vol. 11, No. 1. 163-179.
6. Pořízková K., Artimová J., and Švanda L. (2013) Latinská lékařská terminologie ve světle moderních výukových metod (Latin Medical Terminology in the Light of Modern Teaching Methods). In: ACC Journal, Liberec: Technická univerzita, XIX, 3/2013/Issue C, s. 134-139.
7. Swales J. M. (2004) Research Genres: Explorations and Applications. Cambridge: Cambridge University Press.
8. Taavitsainen I. and Pahta P. (2000) Conventions of Professional Writing: The Medical Case Report in a Historical Perspective. In: Journal of English Linguistics. Vol. 28, No. 1. 60-76.

Summary

USING TOOLS OF CORPUS LINGUISTICS FOR INVESTIGATING AND TEACHING LANGUAGE FOR MEDICAL PURPOSES

The paper presents a corpus-based research into language for medical purposes. Specifically, it demonstrates the ways practising clinicians use Latin in clinical diagnoses and write English medical case reports for publication. The research draws on corpora of authentic language, analyzed via Sketch Engine – corpus linguistics software at Masaryk University in Brno. The paper concludes with important pedagogic applications stemming from such an approach, in which corpora are used as a basis for language teaching.