

Corpus concordancing in teaching academic discourse writing to medical students

Ewa Donesch-Jezo

E-mail: ewadi@op.pl
Jagiellonian University (Poland)

Abstract

In the academic world of science, a world that has become highly globalized and competitive, the ability to write a research paper in English is of increasing importance. This ability is essential for the dissemination of scientific achievements. Therefore, the teaching of the skills necessary to describe a research procedure according to the conventions of academic writing, has become an indispensable component of the syllabi of ESP for university students.

Working in response to the assumption that the analysis of a small corpora, containing samples of research articles combined with appropriate learning tasks, is a useful resource for developing fluency in academic literacy, I conducted a comparative study of two corpus-based methods of instruction in the teaching of academic discourse writing.

The purpose of this paper is to present a classroom-based study, comparing the effectiveness of two methods for the teaching of academic literacy to medical students. The two methods were: (1) concordancing-related corpus analysis, and (2) non-concordancing-related corpus analysis. Both of the studied methods were based on the analysis of the corpora, comprising research articles in the field of medicine, and both of them aimed at enhancing the students' understanding of the role of metadiscourse in academic texts.

1. Introduction

ESP (English for Specific Purposes) has evolved rapidly since the 1960s, originating from registers of specialist languages used in science and technology [1]. Nowadays, ESP and its branches, EAP (English for Academic Purposes) and EOP (English for Occupational Purposes) are major approaches to English language teaching adjusted to the learners' needs in their academic and workplace settings. The demands of the competitive contemporary world of research and scholarship combined with the growth of English as the leading language for the dissemination of academic knowledge [2] have bound the careers of academics and students to a relative proficiency in English language use in a university context. The ability to give lectures, make presentations at meetings and at international conferences, and, above all, to conduct and publish research in international scientific journals, are demands of the academic community imposed on academics and students, alike. The key genre used for the dissemination of scientific knowledge is the research article (RA), also called research paper (RP) [3, 4]. In response to these demands, the development of the necessary competence, allowing to write an RA which is linguistically appropriate and adjusted to the disciplinary-specific conventions of academic writing should be included in the ESP curriculum for both undergraduate and graduate students.

Trying to find, at a university undergraduate level, an effective method for the development of academic literacy, I conducted an empirical study comparing two methodological approaches to the acquisition of academic writing competence: (1) concordancing-based genre analysis, and (2) non-



concordancing, traditional genre analysis. Both of these methods involve the analysis of specially constructed specialist corpora, aiming to increase the students' awareness about the rhetorical structure and the role of metadiscourse in academic texts [3, 5, 6].

This paper provides the evidence from the above-mentioned study and displays the ways in which a concordancing programme, combined with a specialist corpus and language-learning tasks, can help to develop students' competence as writers within a specific academic domain.

2. Corpora and corpus concordancing in ESP

Corpus-based learning and teaching of lexico-grammatical and stylistic features of a foreign language require a sufficient number of language samples, which are ensured by the use of a language corpus. A corpus is a collection of authentic texts (written or spoken transcripts) gathered according to specific criteria [7]. These texts are usually stored in an electronic form, which allows their compilation in large amounts, comprising thousands or even millions of words. The size of a corpus depends on its application. Up till now language corpora have been used in many disciplines, such as lexicography, descriptive linguistics, sociolinguistics and foreign language pedagogy [8]. Corpora may be composed of texts from various sources (e.g. a reference corpus such as the British National Corpus or the Bank of English), or from a particular genre (a specialist corpus). Since corpora contain real-life language, and not the language specially invented for pedagogic purposes, they provide linguistic material for an evidence-based approach to foreign language teaching. The authenticity of language resources makes this approach ideal for studying and learning the features of academic genres and for developing genre competence in university students [8, 9, 10, 11]. For an analysis of a particular genre and teaching its writing within ESP a small corpus is sufficient, but it should be large enough to allow to make valid generalizations about language patterns in the text [12]. Using text analysis programmes, commonly known as concordancers or concordancing software (e.g. WordSmith, Mono Conc), students can examine grammatical, lexical and stylistic features of large amounts of texts within a relatively short time. A concordancer makes a list of all the occurrences of a search word or phrase in a corpus with its immediate context (some text to its left and right side) [10]. The output is displayed in a KWIC (key word in a context) format. In language teaching, concordancers have many applications: they can produce frequency lists, key words, alphabetically ordered lists, word clusters (multiword units), etc. Key words allow rapid identification of the content lexis of a text and provide insights into syntactic patterns in a text which could otherwise remain unnoticed, or would take a long time to discover [8].

3. Study design

3.1 Material and method

Designing the present study, I hypothesized that writing instruction based on resources from small specialist corpora along with concordancing-related tasks could facilitate the students' understanding of academic written discourse, and could be effectively used in teaching academic writing. The methodological framework that I used was related to corpus-based genre analysis as proposed by Tribble (2002), Baker (2006) and Hyland (2006). The genre analysis conducted by the students under my supervision involved the use of some lexico-grammatical items, including interpersonal metadiscourse devices, and rhetorical features of the text. "Metadiscourse" is a term denoting linguistic items which help relate a text to its readers, and also those which assist the writer to express his/her viewpoint and communicate with readers [3, 6]. In the present study the students investigated the use of such metadiscourse devices as hedges (which tone down the writer's assertiveness), boosters (which emphasize the writer's assertiveness), evidentials (which refer to information from previous researches) and self mentions (I, we) [3].

The study included 31 students in the fourth year of medical studies - the participants of two classes of their obligatory EMP (English for Medical Purposes) course at the Medical College of the Jagiellonian University in Krakow. These two classes constituted two study groups, A and B, 15-16 persons in each. The students' proficiency level in English was B2-C1. The duration of an experiment was 8 hrs, scheduled 2 hrs weekly.

The corpus for the study was compiled in advance by the teacher by selecting randomly 15 research articles from prestigious scientific journals in the field of medicine. All these articles were written by native speakers of English, and had been retrieved online from the Jagiellonian University library. As Introduction and Discussion sections of the RA are the most prominent and also the most difficult for the writer to compose – the former introducing a topic in the RA and justifying the necessity of its investigation, the latter presenting and discussing the results and expressing conclusions [10] – these two sections were chosen for the students' genre analysis. The corpus was, therefore, divided into two sub-corpora, one for Introduction sections and the other for Discussion sections and stored in ASCII format. The analytical tool was the concordancer MonoConc Pro 2.2 [13].

3.2 Study procedure

Before the analysis of the corpus, the students of both groups had been offered an introductory session in which they were familiarised with the IMRD (Introduction, Method(s), Results, Discussion) sections of the RA, and discussed the functions and content of each of them. Group B was additionally provided with information on the use of a concordancer, and some print-outs of the concordance pattern were shown. Then the students of both groups received the two sub-corpora with texts of Introduction and Discussion sections, and, working in groups of 3-4 persons each, performed genre-exploring tasks designed by the teacher (the author of this paper). The tasks aimed to heighten the students' awareness about the rhetorical organization and linguistic features of the RAs. The difference between groups A and B lay in the form of the sub-corpora and the analytical tool used. The sub-corpora for group A were in a printed form, and the students were equipped with coloured text markers to mark the searched items in the text, whereas the sub-corpora for group B were in an electronic form, and the analytical tool was a piece of concordancing software.

The learning tasks, which had been modelled on those proposed by Tribble and Jones (1997), Tribble (2002) and Bowker and Pearson (2002), included open-ended questions, gap-filling, matching the parts of sentences, etc. For group B some clues to guide students in concordancing were provided. Worksheets containing instructions for each stage of the analysis process and the tasks had been prepared by the teacher in advance and handed out at the beginning of the session.

Examples of questions used in the tasks are given below.

1. How do the writers state objectives in the RAs? (Examine the concordances for the words "study" and "investig*", where the asterisk stands for any letters.)
2. What tense is used for presenting the objectives?
3. What syntactic pattern can you find in the stating of the objectives by the RA writers?
4. What reference verbs are used to review earlier research? (As the authors usually include a reference to the year of publication in brackets or parentheses, use the search item "(*)".)
5. How do the RA writers explain the necessity for doing the current research? (Examine the concordance for the words "however", "nevertheless" and "not".)
6. Where do the writers mark their presence in the RA? (Search for the concordances of the key words "I" and "we".)
7. How are the findings expressed in the RAs? (Examine the concordances for the words "finding*" and "result*").)
8. Where in the text do the authors decrease the assertiveness of their statements? (Examine the concordances for the auxiliary verbs "may", "might", "can", "could", lexical verbs "suggest", "seem", "appear", and adjectives "possible", "likely".)

9. Where in the text do the authors increase the assertiveness of their statements? (Examine the concordances for the words "definitely", "obviously", "demonstrate".)
10. What tense is used to express findings, results and conclusions? (Expand the concordances for the words "finding*" and "result*" to three words to the right.)

The students were asked to write the answers and conclusions drawn from their observations, and in some instances to formulate a rule.

The teacher, walking around the classroom, supervised the students' work and provided corrective and/or explanatory feedback when necessary. Findings obtained by the groups were compared between them and presented in the whole class context, after which the teacher summarised the results of the students' analysis.

4. Results and discussion

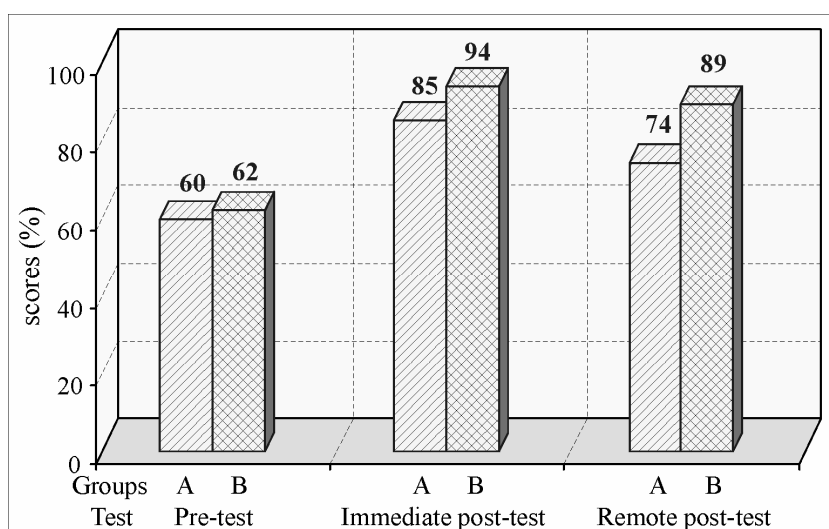


Fig. 1. Results of the tests

The results of my investigation showed (see Fig. 1), that both methods of raising the students' genre awareness was sufficient for them; firstly, to notice the rhetorical conventions of the RA and its lexicogrammatical items, and, secondly, to enable them to incorporate these features into their own discourse writing. The scores obtained by group A as compared to group B seem to indicate that concordancing-based exploration of specialist corpora can be effectively utilised in the teaching and learning of linguistic and stylistic features of an academic genre. The study of these features performed by the students guided by concordancing activities had all the qualities of data-driven learning (DDL). In DDL, the learning process is based on the learner's own discovery of rules using authentic language data [14, 15]. This type of text analysis enabled the students to make the following generalisations:

1. First person pronouns are used when the writers describe their own procedural choices in their research.
2. Expressions mitigating the force of a proposition (hedges) are more frequent than those expressing certainty (boosters).
3. The preferable tense for outlining the objectives of a study is the past tense.
4. The most frequent verbs introducing the purpose of the study are "identify", "investigate" and "determine".
5. The syntactic pattern for introduction of the purpose is "the*of*study was to*".
6. Reference verbs are "find", "measure", "perform", "report", "show", however they are rare (the writers tend to use reference numbers).

7. The writers explain the necessity for their current research using negative forms "is not known", "is not clear"; sentence connectors "however", "nevertheless", and verbs "need", "require".
8. Conclusions are presented in impersonal forms, e.g. "in conclusion", "it is concluded that", "lead to the conclusion that".

It should be noted that the group of students which used corpus concordancing software arrived at these generalisations far more quickly than those students who performed corpus analysis in a traditional, "manual" way. Consequently, group B completed the analysis of the corpus (15 Introductions and Discussions) within the allocated time, whereas group A was able to analyse only 10 Introductions and Discussions within the same time. Moreover, students' opinions about the employed methods of teaching of academic literacy were different: students from group B perceived it as "interesting" and "exciting" while in the opinion of students from group A, it was "boring" and "tiring".

5. Conclusions

Concordancing-based genre analysis can, it may be argued, be successfully used in ESP classes for the development of students' academic genre competence. The main advantages of this methodology are:

1. By indicating high-frequency words, phrases and syntactic structures, it shows how they are used in a given genre.
2. It favours discovery-based learning by encouraging the students to discover knowledge by themselves.
3. It helps teachers to better understand the generic form of the subject matter they teach.
4. It shows the students the choices they can make when writing a discourse in order to adjust it to the conventions of a given academic community.
5. The print-outs of the concordances can be used in the preparation of teaching materials and tests.

References

- [1] Halliday, M.A.K., McIntosh, A., Strevens, P. (1964). *The Linguistic Sciences and Language Teaching*. Oxford: Oxford University Press.
- [2] Hyland, K. (2004). *Disciplinary Discourses: Social interactions in academic writing*. London: Pearson Education Ltd.
- [3] Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum, pp. 48-58.
- [4] Swales, J. M. (1990). *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- [5] Hyland, K. (2004). *Genre and Second Language Writing*. Michigan: the Michigan University Press, pp. 202-205.
- [6] Hyland, K. (2006). *English for Academic Purposes: An advanced resource book*. London, New York: Routledge, pp. 58-64
- [7] O'Keeffe, A., McCarthy, M., Carter, R. (2007). *From Corpus to Classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- [8] Tribble, C. (2002). *Corpora and Corpus Analysis: New windows on academic writing*. In J. Flowerdew (Ed.) *Academic Discourse*. London: Pearson Education, pp. 131-149.
- [9] Tribble, C., Jones, G. (1997). *Concordances in the Classroom: A resource guide for teachers*. Houston: Athelstan Publications.
- [10] Bowker, L., Pearson, J. (2002). *Working with Specialized Language: A practical guide to using corpora*. London, New York: Routledge.
- [11] Hyland, K. (2009). *Academic Discourse: English in a global context*. London, New York: Continuum.
- [12] Baker, P. (2006). *Using Corpora in Discourse Analysis*. London, New York: Continuum.
- [13] Barlow, M. (2003). *Concordancing and Corpus Analysis Using MP 2.2*. Houston: Athelstan Publications.



- [14] Johns, T. (1991). *Should you be persuaded – two examples of data-driven learning materials*. English Language Research Journal (4). University of Birmingham, pp. 1-16.
- [15] Johns, T. (1993). *Data-driven Learning: An update*. TELL&CALL 1993 (2), pp. 4-10.