

MINI-CORPORA IN THE LANGUAGE CLASSROOM: TITLE AND ABSTRACT MINI-CORPUS

© Robin L. NAGANO
(University of Miskolc, Miskolc, Hungary)

nyerobin@uni-miskolc.hu

Received: 14.08.2012; Accepted: 30.08.2013; Published online: 07.07.2014

Materials selection is always a challenging issue, especially when faced with learners of mixed areas of study and varying degree of motivation. One approach that allows student-specific material is to use collections of texts, or corpora. These are invaluable tools for exploring and analyzing texts. However, it is also necessary that students use and learn from entire texts, and one of the most efficient ways is to have students choose and work with texts that interest them personally. These can then be formed into a mini-corpus, so that both full texts and the lexical and syntactic information of the entire body of texts can be exploited for materials development. If students are ready to begin actively exploring and analyzing texts on a particular topic (generally level B2+), tools such as concordance software can also be employed. Here, two classroom-oriented mini-corpora are introduced, along with sample tasks and examples. Both are collections of titles and abstracts of research articles in areas of engineering. In one case, the focus is upon text analysis and academic writing, while lexical development is the focus of the second case. Student response to this trial activity has been very positive: they appreciate having the chance to test the generalized information taught in class on texts that were immediately relevant to their needs.

Keywords: corpus, EAP, LSP, titles, abstracts

Second language teaching in higher education faces several difficult tasks. Among them are dealing with a rather large number of not very enthusiastic learners, at different levels, with differing motivation and goals. In addition, the importance of language for specific purposes is growing. This paper discusses the possible role of small purpose-built corpora in teaching as an approach with advantages in facing the current challenges of language teaching.

Motivation in the language-learning classroom

Motivation for learning is often divided into two categories: integrative and instrumental (a similar notion is intrinsic and extrinsic motivation) (see e.g. Csizér & Dörnyei, 2005). An integratively motivated language learner is driven to learn the language for internal reasons: to better understand a culture, to feel a sense of belonging, for interaction with speakers of the target culture, etc. Motivation powered by instrumentality is external, perhaps more practical – to manage the shopping, to fulfill requirements at school, to be eligible for a job, to negotiate with business partners – and sees the language as a tool to be used for some purpose. Csizér and Dörnyei (2005), in a study of Hungarian secondary school students' motivation in language learning, found that integrative motivation leads to better performance. However, they redefine the term so that it encompasses some types of instrumental motivation when related to the 'ideal self' (internalized instrumental motivation), but not short-term motives such as avoiding punishment, which do not contribute much to the long-term effort needed to acquire a language. Gardner discusses also 'attitude towards the learning situation', where "it is the influence of the Educational Context on the individual's attitudes that influence the individual's level of motivation" (2007: 15).

Integratively motivated students are usually fairly self-directed, and do not need much outside stimulus. Students with primarily instrumental motivations will readily cooperate as long as they feel that they are progressing towards meeting their goals. However, when they feel that the material covered is not relevant to them and their needs, they become reluctant to expend time and energy on it.

In my experience, instrumentally-motivated learners who have reached general proficiency – say, those on level B2 of the Common European Framework of Reference for Languages – often lose motivation. Perhaps this is because one goal – maybe their only explicit goal – has been achieved as soon as they receive the intermediate certificate of language proficiency (useful for application to university and vital for graduation from university under the Hungarian system). Another factor may be that they see how much there is still to learn – like being in the middle of a large lake, surrounded on all sides by water and the shores so very far away – and do not know which direction to head in.

At this point, language teachers face two challenges. One is to help learners set new goals. In the context of higher education, this is often the goal of learning language forms and skills useful for their studies and (future) professions. While students know, vaguely and passively, that foreign language proficiency is a plus in the job market, they need to be convinced of this in some way. The second challenge is to provide them with material and to work on skills that are relevant to their new goals.

We now run into a further problem. In higher education contexts, language classes tend to consist of students from one faculty – economics, mechanical engineering, law, earth sciences, etc. – but each faculty has numerous branches, studying quite different areas. How do we choose relevant materials for all? Well, we look for the common denominator, for skills that all will find useful, vocabulary that is widely used, and we cycle through a variety of not-too-specialized topics, hoping to suit most of the students most of the time.

There is another option, which involves using student-specific material. Texts can be selected by the teacher or by the student, within certain guidelines. Tasks can be designed that are general enough for use with a variety of texts, so that different students can be doing similar tasks based on different texts. This serves as an alternative – or more likely, a supplement – to a common set of texts and topics.

In addition to task-based work with these texts, the electronic form of the selected texts can be formed into a digital collection – a corpus – and the lexical and syntactic information contained within them can be utilized using corpus tools such as a concordancer.

Language corpora

While the practice of collecting texts into a body and studying them is by no means a new technique, advances in information technology have made it possible to collect, store, and search digital collections very quickly and efficiently. Large corpora have been widely used by lexicographers in writing dictionaries and by linguists interested in word use and more recently, syntactic use. Corpora are used to gather information on language use by writers of teaching materials (e.g., *Thurstun & Candlin, 1997*), grammar books (e.g., *Biber et al., 1999*) and in language testing (e.g., the TOEFL Corpus and its analysis by *Biber et al., 2004*). Teachers and learners of language, either as a first or second language, can refer to corpora to resolve issues of usage, including frequency, range, register and mode, and to identify collocations, or words that tend to be used together.

Large general corpora are very careful to collect texts in a way that represents usage in general, choosing samples from different text types and genres, different populations, different time periods, etc. (*Sinclair, 1991; Szirmai, 2005*). Some include samples of both spoken and written language, and others focus on one source (a corpus formed of articles from one newspaper, for example). There are open corpora, which continue to be added to, and closed corpora such as the well-known British National Corpus. Some examples of general English corpora are listed in Table 1.

More specialized but still rather large corpora have been formed to meet more specific purposes. For example, one corpus (*Hyland, 2000*) of English-language textbooks in various fields of study was formed in order to study features of academic language use, and used to produce the influential Academic Word List (*Coxhead, 2000*), which attempted to identify words that occur frequently across different fields of study and would thus be useful for students of English as a Second Language (ESL) studying in English-medium institutions or needing English for Academic Purposes (EAP). Another academic corpus was collected that included, along with written text, many samples of spoken language in academic situations (lectures, student-professor consultations, student conversations, administrative encounters, etc.). This was used extensively when designing tasks for the TOEFL exam (*Biber et al., 2004*), a test for assessing the language proficiency of ESL applicants to higher education programs in the United States. In order to produce statistically valid results, research corpora have to reach a certain size and be extremely carefully designed, since they must be representative (*Sinclair, 1991*).

Table 1. Examples of relatively recent large general English corpora available on-line

corpus	words	language / dialect	time period
Corpus of Contemporary American English	425 million	American English	1990-2012
Corpus of Historical American English	400 million	American English	1810-2009
TIME magazine corpus	100 million	American English	1923-2006
Corpus of American Soap Operas	100 million	American English	2001-2012
British National Corpus (BNC)	100 million	British English	1980s-1993

Note: Free on-line access to all corpora listed in the table is available at corpus.byu.edu, thanks to Prof. Mark Davies of Brigham Young University.

Another approach in corpus-based research and teaching is to make the corpus up from learner-produced texts. Learner corpora allow research into typical features of interlanguage usage, and which can also be useful for teaching purposes. These include certain learner populations (Japanese learners?) or particular situations, such as academic environments. Sharpling (2010) reports on the use of the British Academic Written English (BAWE) corpus in designing a proficiency test at Warwick University. Römer and Wulff (2010) analyze student writing using the Michigan Corpus of Upper-level Student Papers (MICUSP). Alternatively, a learner corpus can be compared with an expert corpus, as was done by Hüttner (2010), who compared student writing (university students' first full papers in English) with published academic writing, specifically examining the conclusions of research papers. Yet another trend in corpus-based language studies is the use of parallel corpora – typically, dealing with the same topic but in different languages – to aid translation and translation studies.

There is also a group of smaller corpora where the corpus builder chooses a more restricted category of texts. These tend to be specialized in various ways: either in topic or in genre/text type or in source, or even in some combination of the above. Such corpora are often used for research into the use of language for specific purposes (LSP). Some examples include Hyland (2000), who investigated research articles, abstracts (Samraj, 2002) and titles (Soler, 2007; Nagano, 2010).

Compared to the million or more words of a general corpus, Bowker and Pearson (2002) recommend at least 10,000 words for a specialized corpus used in LSP situations, such as building useful glossaries, term extraction and use as a writing guide or translation resource. This size is still reasonably reliable, they state, as long as the texts are restricted to one focus. Clearly, it is necessary to identify needs and form guidelines for choosing texts (topic, genre, sources, date of publication, length in words, etc.).

In this paper, though, I am discussing mini-corpora for classroom use. I have chosen the term to emphasize the small size of the corpus. This necessarily means that the corpus cannot produce statistically valid results for, say, lexical frequency lists. However, that is not the purpose of (my conception of) a mini-corpus. A mini-corpus is suitable for classroom activities, working with students on basic text analysis, and for gaining lexical and syntactic information using corpus tools, while ensuring that the texts are directly relevant to the learners.

Since a mini-corpus is of a more manageable size, it is feasible to gather, say, 20-30 texts on a certain topic, or of the text type being focused upon. Suitable texts are usually available on the Internet, already in digital form. They can easily be imported into a concordancer, a kind of software for analyzing corpora. In this paper, I will be showing examples using the free concordancer TextSTAT, of the Free University of Berlin, since it is user-friendly and easily accessible, although more sophisticated software is available on the market.

The focus of corpus users is usually upon corpus tools that draw on the full collection of texts. One useful feature is the ability to produce frequency lists of all the words appearing in the corpus. Another is to look at concordances: a search word or phrase is given, and the screen displays all examples occurring in the texts. This is displayed with the words occurring to the left and to the right of the target, in a way called key word in context, or KWIC. These can be sorted in various ways to highlight different features. This allows users to observe commonly occurring lexical partners (collocations) and the syntactic frameworks the target word/phrase is used in. It is possible to see selected items in context in the original text, as well.

However, one should not rely only upon corpus tools at the expense of looking at full texts! These allow us to work on larger-scale issues such as cohesion and coherence, information content and patterns of expression, and other textual features. However, issues that arise during text analysis may lead back to a search of the corpus to investigate other instances. Examples shall be shown below, as I describe two variations on a mini-corpus and their potential applications in a university classroom.

Title and abstract mini-corpora

The possible texts for a corpus of written texts ranges from newspaper articles to blogs, business letters to research articles, term papers to text messages. I have chosen to concentrate upon the genre of the research article because my interests are in academic writing, and because I feel it is in the interest of university students to become familiar with journal articles, as a step in their education and potentially in their further studies and/or career. However, full articles are rather unwieldy for classroom use. They are long, complex, demand background information, and working with them is time-consuming. On the other hand, abstracts are (generally) short and are stand-alone texts. Anyone researching an academic topic is likely to read many more abstracts than full articles, since abstracts help potential readers select relevant articles. This is even truer of titles. One survey found that for every full article read by researchers, approximately 20 abstracts and 100 titles are read (Mabe & Amin, 2002). Abstracts have often been chosen as a teaching focus, for reading (Morton, 1999; Nagano, 2004) or for writing (Cox, 1995) in second-language classrooms, but also as a means of initiating native speakers into the academic discourse of an area of study (e.g., Mendelson, 1987). Labassi (2009) makes a case for starting with titles when teaching second-language students how to process information in academic papers quickly and efficiently. Therefore I have chosen to focus upon titles and abstracts of research articles as an introduction to both Languages for Specific Purposes (LSP) and English for Academic Purposes (EAP). Here I present two text collections for different courses and focuses.

Example 1: EAP focus

One module of a special three-semester course for engineering students involves academic writing. One of the target genres is the research article, and we spend some time going over different sections, the information contained in them, and the language used. As an in-class consolidation activity, we carry out tasks using titles and abstracts that are closely related to the participants' research topics. These are collected either by students or by the teacher, based on key words given by students. The objectives of the tasks are to classify titles, and to analyze abstracts in terms of moves and discourse features, using a worksheet. While each student is working with a different set of texts, all are carrying out the same tasks. We then gather in groups, compare notes, and form generalizations. When questions occur, we turn to reference sources, including the on-line corpora listed in Table 1. As the course goes on and the number of collected texts grows, we can form our own reference corpus of abstracts in mechanical engineering.

Shown in Figure 1 is a filled-in worksheet for title classification and analysis. When students compare notes, they will find that quite long titles with one part (i.e., no subtitle) are typical of engineering, and that noun phrases are the most conventional structure, often modified by a couple of propositional phrases, and sometimes with added features (as in the title shown here). They will find that topic is always present but that other types of information may vary depending on their field. (For more information on titles in the hard sciences, see e.g. Haggan, 2004; Soler, 2007).

Figure 1. Worksheet for title form and content

<p>Title: Effect of tool holder geometry and cutting condition when milling nickel-based alloy 242</p> <p>1. How many parts does it have? one</p> <p>2. What structure does it have? noun phrase + relative clause</p> <p>3. How many words does it have? 13-14</p> <p>4. How many prepositional phrases does it contain? one</p> <p>5. What information does it give?</p> <p style="padding-left: 2em;">Topics (tool holder geometry, cutting conditions) + material + process</p> <p>6. What type of paper do you think it will be?</p> <p style="padding-left: 2em;">experimental: milling at different geometries and under different cutting conditions; comparing results</p>
--

Worksheets for abstracts also allow students to work their way step-by-step through an analysis, synthesizing topics covered in class with language in use in their own research topics. This helps students recognize the relevance of the tasks and information.

Example 2: LSP and vocabulary focus

As one example of a mini-corpus and its classroom applications, I shall introduce a mini-corpus with a double focus: it consists of titles and abstracts of research articles, and is thus restricted in genre and text type, while the topic is restricted to an engineering topic, the energy performance of buildings (EPB). The corpus includes 25 titles and abstracts of articles published in nine journals between 2000 and 2011. The majority of articles are from the journal *Energy and Buildings* (Elsevier). The abstract corpus contains 4,562 running words (an average of 182 words per abstract), while the total title word count is 322 (an average of 12.88 words per title).

The class that this corpus was collected for is a group specialized in the Energy Performance of Buildings within the Master of Science in Energetics program in the Faculty of Mechanical Engineering and Informatics at the University of Miskolc, Hungary. The students' English proficiency varies, with some who have already spent a semester abroad studying in English, and others who are closer to B1 level and struggle to communicate. The one-semester language course takes place in the third semester of their four-semester program, when they are beginning their research projects and often required to use English-language sources. The course objectives are to improve vocabulary and proficiency in technical English and especially in this particular field and to serve as an introduction to the language of research and its presentation. The final task of the semester is to write a progress report and give a PowerPoint presentation on their research up to that point.

The mini-corpus serves mainly as a supplementary source of information, and the texts contained within it are used for full-text tasks as well as with corpus tools. Naturally, the corpus is a useful resource for the teacher when preparing tasks, but here I shall focus on student contact with the specialized corpus. The objectives here are to identify frequently occurring words, look at common collocations, and examine issues in professional language use.

The students' first direct contact with information from the corpus is a word list drawn up from the frequency list provided by TextSTAT. After removing function words such as articles, prepositions, and pronouns, a list of content words is produced. The list is cut off at some reasonable frequency, such as five occurrences. Going over the list is a take-home task that is of special benefit to the weaker students. An abbreviated sample is shown in Table 2.

Table 2. Most frequently occurring content words in EPB abstract corpus, with number of occurrences

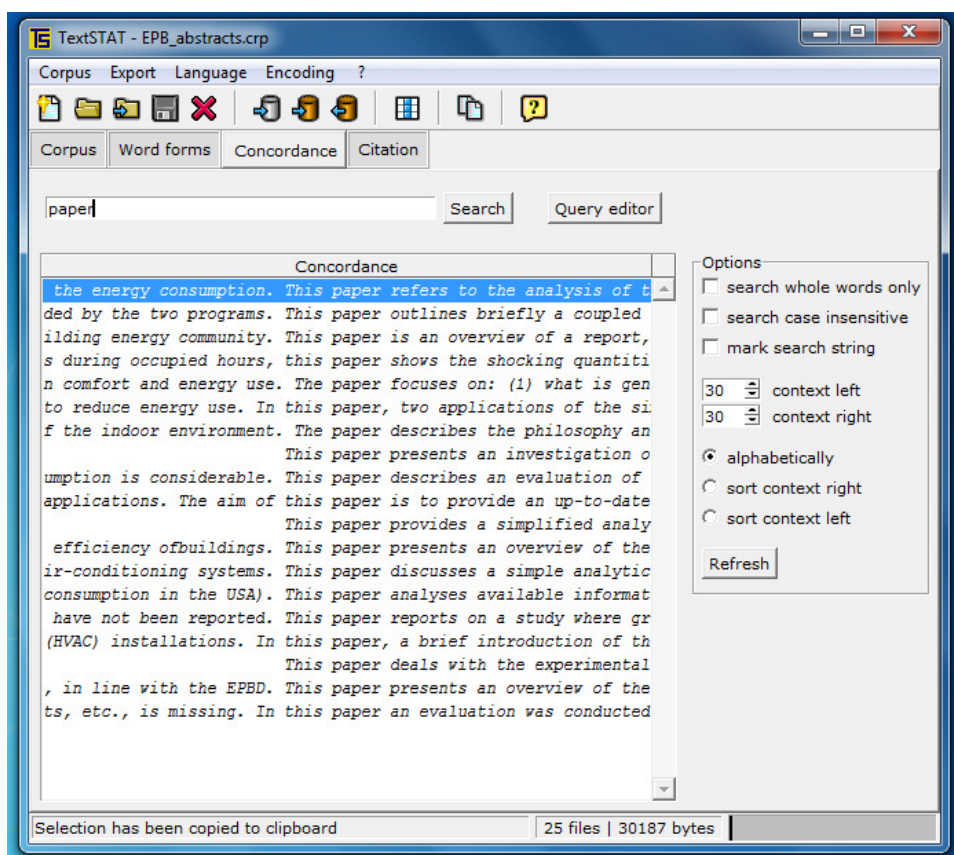
word	number	word	number	word	number
energy	111	European	13	application(s)	8
building(s)	97	climate	12	approach(es)	8
load(s)	21	roof(s)	12	different	8
consumption	27	value(s)	12	evaluation	8
cooling	20	environment	11	green	8
simulation(s)	20	reduction	11	lighting	8
paper	19	residential	11	reduce	8
performance	19	savings	11	annual	7
thermal	19	systems	11	comfort	7
tree(s)	19	window(s)	11	conditioning	7
air	17	analysis	10	demand	7
indoor	16	design	10	environmental	7
method(s)	16	results	10	existing	7
heating	14	office(s)	9	forecasting	7
data	13	potential	9	HVAC	7
emission(s)	13	radiation	9	influence	7

From the frequency data listed in the figure, it is evident that the choice of texts has a large influence on the data. “Tree(s)”, for example, is probably not such a major item of interest in the subject area as would be implied by its position in the list, but is ranked high here because two of the 25 abstracts deal directly with the effect of tree shade on building energy consumption.

Different senses of a word can be focused upon using the concordance feature. For instance, we can see at a glance that ‘paper,’ as used in this collection of abstracts, refers not to the material but to the article under discussion (Figure 2). A search for ‘while’ in the abstract corpus finds that it is used four times, each time in the sense of a contrastive transitional phrase, not in the sense of ‘during’ (e.g., *French and Polish buildings have the highest production of airborne emissions, while Polish buildings have the highest emissions of solid wastes*). Other examples and ideas are given in Tribble and Jones (1997).

Technical terms can also be identified using the concordancer. A search of occurrences of ‘green’ shows that it is present only in the combinations *green building* and *green roof*, clearly not referring to color. The adjective ‘thermal’ brings up a number of technical terms, including *thermal transmittance values*, *thermal properties*, *thermal capacity* and *thermal inertia*. Bowker and Pearson (2002) offer several suggestions for term extraction.

Figure 2. Concordance from abstract corpus for 'paper'



The syntactic role of a word can also be investigated: we find, for instance, 'potential' used as a noun seven times (*the global potential for...*) and as an adjective three times (*...offer potential savings*). It turns out that 'use' is more common as a noun in this type of text: it appears 34 times as a noun (*the use of..., annual air conditioning energy use*) and 17 times as a verb (in 12 instances in the passive voice). 'Load(s)' appears only as a noun, in collocations such as *peak load* and *cooling load*.

The corpus can also be used to answer questions when students sit down to write. For instance, what verb should be used with 'measurements'? Writing it into the search field will bring up concordances including *measurements were made*. What other non-concrete things can be made? Searching for 'make' brings *improvements, an evaluation, comparisons, and an impact* (usually in the passive voice). And does 'data' take a singular or plural verb? According to this small corpus (but remembering the lack of statistical validity), in this field abstract writers use both: there are four instances using the singular and two using the plural.

The next student contact with the corpus is with titles, as I ask students to give a tentative title to their research project report. We first discuss the various activities and approaches (computational, numerical simulation, experimental, analytical, theoretical, etc.) and then take a look at the list of 25 titles. A shorter sample is given in Table 3. I ask them first to work in pairs or small groups to try to classify the studies by approach and type of information included in the title, using a simplified version of the worksheet in Figure 1. I also request each student to select four or five titles that sound interesting (and keep a record of this for later use).

Table 3. Sample titles, EPB corpus

1.	Shade trees reduce building energy use and CO2 emissions from power plants
2.	Contrasting the capabilities of building energy performance simulation programs
3.	The dark side of occupants' behaviour on building energy use
4.	Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings.
5.	An analysis of future building energy use in subtropical Hong Kong
6.	Optimal electrical circuiting layout and desk location for daylighting controlled spaces
7.	The philosophy behind EN15251: Indoor environmental criteria for design and calculation of energy performance of buildings
8.	Study of the influence of roof insulation involving local materials on cooling loads of houses built of clay and straw

Interestingly, even the small sample of titles here shows some variety, with Title 1 in the form of a sentence (with a finite verb) stating the main findings, Titles 2 and 4 beginning with '-ing' constructions (which lends a practical, 'how-to' effect), and a two-part title in Title 7. Compared to the broad area of mechanical engineering, the specialized field of building energy performance may be more applied and interdisciplinary, which has been shown to have an effect on title structures and forms (Gnutzmann, 1988).

Work with full-text abstracts follows, but examples are omitted here. A multi-layered approach is taken, as recommended by Devitt (1997) for dealing with authentic texts. First the focus is on the types of information contained in the abstract. Next students deal with words, including multi-word terms and acronyms (if used). Then they are asked to look at sentence-level features such as the use of tenses, the passive voice, and transitional phrases. Questions on usage can be followed up by concordance searches.

Conclusion

The purpose of this paper was to discuss the potential role that a small but purpose-built corpus can play in the classroom. Since texts are chosen for (or by) particular groups of learners, they are relevant to the interests and/or needs of the learner, thus presumably leading to a higher degree of motivation, even in instrumentally motivated students.

With a collection of short texts such as those dealt with here – titles and abstracts of research articles in engineering fields – it is possible to use the full texts in the classroom. One advantage of a purpose-built mini-corpus is that it can act as a library of varied texts for classroom use – with sufficient care, worksheets can be designed that can be used with any text, and therefore different learners can be using different texts to carry out the same tasks. Not only does this allow learners to work with texts relevant to themselves, but they can then compare notes with others to discover whether their findings are generally true or not.

The use of corpus tools also allows all of the lexical and syntactic information in the texts to be accessed. Frequency lists can be modified into vocabulary lists, and technical terms can be extracted. Concordances allow teachers and learners to observe patterns of lexical items in use, including syntactic constraints on their usage, and common collocations can be identified. Ideally, the corpus and the concordancer can be distributed to students for their private use, and perhaps for future expansion. Being able to

build their own corpus and use corpus tools can potentially be an important step on the road to being independent and proficient users of the language, including language for specific purposes.

References

- BOWKER, L., & PEARSON, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London and New York: Routledge.
- BIBER, D., JOHANNSON, S., LEECH, G., CONRAD, S., & FINEGAN, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- BIBER, D., CONRAD, S. M., REPPEN, R., BYRD, P., HELT, M., CLARK, V., CORTES, V., CSOMAY, E., & URZUA, A. (2004). *Representing Language Use in the University: Analysis of the TOEFL® 2000 Spoken and Written Academic Language Corpus*. TOEFL Monograph Series. Princeton, NJ: Educational Testing Service.
- COX, J. (1995). Analysing geotechnical engineering abstracts: Towards a pedagogical template. *ESP Malaysia*, 3 (2), 136-144.
- COXHEAD, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- CSIZÉR, K., & DÖRNYEI, Z. (2005). The internal structure of language learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal*, 89 (1), 19-36.
- DEVITT, S. (1997). Interacting with authentic texts: Multilayered processes. *The Modern Language Journal*, 81 (4), 457-469.
- GARDNER, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum*, 8, 9-20.
- GNUTZMANN, C. (1988). Aufsatztitel in englischsprachigen Fachzeitschriften: Linguistische Strukturen und kommunikative Funktionen. In Gnutzmann, C. (Ed.), *Fachbezogener Fremdspracheunterricht* (pp. 23-38). Tübingen: Gunter Narr.
- HAGGAN, M. (2004). Research paper titles in literature, linguistics and science: Dimensions of attraction. *Journal of Pragmatics*, 36, 293-317.
- HÜNING, M. TextSTAT version 2.9 [software]. Available from <http://neon.niederlandistik.fu-berlin.de/en/textstat/> [06.06.2012]
- HÜTTNER, J. (2010). Purpose-built corpora and student writing. *Journal of Writing Research*, 2 (2), 197-218.
- HYLAND, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow, Essex: Pearson Education.
- LABASSI, T. (2009). Reading titles of empirical research papers. *The Reading Matrix*, 9 (2), 166-174.
- MABE, M. A., & AMIN, M. (2002). Dr Jekyll and Dr Hyde: Author-reader asymmetries in scholarly publishing. *Aslib Proceedings*, 54 (3), 149-157.
- MENDELSON, M. (1987). Teaching the abstract as an introduction to technical writing. *The Technical Writing Teacher*, 14 (1), 1-10.
- MORTON, R. (1999). Abstracts as authentic material for EAP classes. *ELT Journal*, 53 (3), 177-182.
- NAGANO, R. L. (2004). Introducing academic English through abstracts. *Novelty*, 11 (1), 46-52.
- NAGANO, R. L. (2010). Question forms in journal article titles. *Proceedings of the 4th Applied Linguistics PhD Conference* (pp. 100-111). Budapest. Available at <http://www.nytud.hu/alknyelv/dok10/proceedings10.pdf> [06.06.2012]
- RÖMER, U., & WULFF, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research*, 2 (2), 99-127.
- SAMRAJ, B. (2002). Disciplinary variation in abstracts: the case of Wildlife Behaviour and Conservation Biology. In Flowerdew, J. (Ed.), *Academic Discourse* (pp. 40-56). Harlow, Essex: Pearson Education.
- SHARPLING, G. P. (2010). When BAWE meets WELT: The use of a corpus of student writing to develop items for a proficiency test in grammar and English usage. *Journal of Writing Research*, 2 (2), 179-195.

- SINCLAIR, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SOLER, V. (2007). Writing titles in science: An exploratory study. *English for Specific Purposes*, 26 (1), 90-102.
- SZIRMAI M. (2005). *Bevezetés a korpusznyelvészetbe*. Budapest: Tinta.
- THURSTON, J., & CANDLIN, C. N. (1997). *Exploring Academic English: A Workbook for study Essay Writing*. Sydney: NCELTR.
- TRIBBLE, C., & JONES, G. (1997). *Concordances in the Classroom: A Resource Guide for Teachers*. Houston, Texas: Athelstan.