

УДК 811.111 '42

**Очеретович Ольга***(м. Острог, Україна)**Національний університет «Острозька академія»***CONSTRUCTION AND PECULIARITIES OF USE OF BIATHLON  
TEXT CORPUS WITH ANTCONC CONCORDANCE TOOL**

*У тезах йдеться про особливості процесу розробки текстового корпусу у програмі AntConc, що стало основою для укладання базового глосарія терміносистеми біатлона.*

**Ключові слова:** біатлон, глосарій, конкорданс, термін, текстовий корпус.

*В тезисах рассматриваются особенности процесса разработки текстового корпуса в программе AntConc, который стал основанием для комплектования базисного глоссария терминосистемы биатлона.*

**Ключевые слова:** биатлон, глоссарий, конкорданс, термин, текстовый корпус.

*The research deals with the peculiarities of the construction process of the text corpus obtained with the help of AntConc software, which served as a foundation for compilation of the basic glossary of biathlon terminology system.*

**Key words:** biathlon, glossary, concordance, term, text corpus.

The topicality of the chosen topic consists of the fact that despite a considerable number of glossaries, dictionaries, books, encyclopedias, articles, and other sources of sports vocabulary, biathlon terminology requires a higher degree of systematization of terminological units by developing electronic corpus of biathlon. The text corpus of the selected topic serves as the foundation for the research of biathlon terminology, while implementation of modern technologies significantly simplifies the process of in-depth study of this terminology system. The object matter of the research is a text corpus of biathlon of English

language; the subject matter – corpus technologies of representation of the language sports system.

In our previous studies, we have created a preliminary conceptual map of the semantic field through a selection of concepts from available lexicographical and textual sources and built an electronic corpus of texts biathlon and analyzed it. In the presented research, a previously prepared biathlon text corpus served as the foundation for biathlon terminology glossary compilation. This time, however, we did not solely rely on our own language introspection but employed a specially designed program for the text corpora – AntConc. This computer program was developed by Professor Laurence Anthony at Waseda University in Japan for conducting researches in the field of corpus linguistics. The peculiarities of working with AntConc concordancer imply simplifying the processing of a large volume of electronic data. Moreover, this concordancer has a very user-friendly interface and the following basic parameters: Concordance, Concordance Plot, File View, Clusters/N-Grams, Collocates, Word List, and Keyword List. Each time the work with the piece of software was started with uploading the annotated texts that construct our developed biathlon corpus in the program. The functional possibilities of AntConc are demonstrated below.

Concordance allows to learn a word in its natural environment due to the possibility to sort the context both on the right and on left at three levels. For example, a sorted concordance of the term '*shooting*' will look as shown below in Fig. 1. As one can observe, this term has a tendency to be placed before a noun, as for instance, *shooting ability* and *shooting accuracy*. In addition to this, by pointing the mouse cursor on the highlighted term, we automatically go to the document of the annotated text or its fragment.

*Figure 1 The most frequent environments for the term shooting in the corpus*

Having reviewed the functionality of the AntConc program, we proceeded to the stage of compiling our own glossary (sport terminology record) – the registration of the terms in a separate table, adding the most frequently used collocations, definitions, frequency of occurrence of the terminological units and their use in the context – ter-

minological entries. The above-mentioned process was carried out in several stages. First, the analysis of the keyword list was prepared in AntConc. Given the frequency and appropriateness, the terms were included in the table. Owing to the groups /N-Grams, the information about the environment of a given word or phrase was recorded. For example, the term *'finish'* showed such collocations as: finish line, finish record, finish referee, finish camera, finish order, finish time, the photo-finish, and to cross the finish. Second, regarding their frequency, some of these compounds were registered as separate terms, so they were also included in the table despite of the fact that normally the terms have only a one-component structure. Third, each term received a definition. The sources of definitions were mostly dictionaries as, for example, «American Heritage Dictionary of the English Language» [1]. However, if the term or terminological was not found in any of the dictionaries – we applied the processed texts and knowledge gained in the research process to create an approximate definition. As a result, a table of biathlon terms contains 173 various terms that are different according to their semantic, morphological, and structural composition: there are notions that belong to various layers in sports terminology; simple and compound terms (one-, two-, and three-component terms); abbreviations, as well as words comprising different parts of speech and being formed in different ways.

Using the development of the corpus linguistics, we managed to create our own biathlon terms glossary containing 173 terminological units. The glossary looks like a table with the following sections: Term, Definition, Source of definition, Context, Source of context, Collocations, and Frequency. A Specialized Corpus of Biathlon (SCB) served as the context in the vast majority of cases [2].

Having processed the text corpus of Biathlon terms (SCB) with a specially developed piece of software, we succeeded in compiling an electronic glossary of terminological units that appear as a list of terms with terminology records – full dictionary entries for them that explain each term separately and in the sentence, and that also shows the most frequent word combinations and a number of occurrences in a terminology system of biathlon. Such developments may be of the

use not only to researchers or sportsmen but also to people interested in winter sports.

In conclusion, the AntConc concordance tool, as one of the significant achievements of corpus linguistics, has greatly simplified the data processing of lingual data and contributed to the more comprehensive and accurate analysis of biathlon terms, their contextual environment, and their connections within the terminology system.

### **Література:**

1. American Heritage® Dictionary of the English Language, Fifth Edition, 2011. – 2084 p.
2. Encyclopedia Britannica [Electronic Resource]. – Mode of access : URL : [www.britannica.com](http://www.britannica.com). – Title from the screen.
3. IBU Congress (2012). IBU Event and Competition Rules. Rules adopted by the 2012 10th Regular IBU Congress. International Biathlon Union. Retrieved, 2014.–110 p.
4. Lakoff G. Women, Fire, and Dangerous Things. What Categories Reveal about the Mind / G. Lakoff. – Chicago and London : University of Chicago Press, 1987. – 614 p.
5. Terminology management. Text corpus SPORT [Electronic resource]. – Access mode : <http://moodle.oa.edu.ua/mod/data/view.php?id=3650>. – Title from the screen.