AL-FARABI KAZAKH NATIONAL UNIVERSITY

Department of International Relations

Chair of Diplomatic Translation

**Translation business in the field of international and legal relations**

**“Translation of Scientific and Technical Documents”**

2024-2025 academic year, fall semester

**Lecture 5**

**Module 1: Introduction to scientific and technical translation**

**Lecture 5 Lexical features of scientific and technical texts**

Corpus-based terminology research

An electronic collection of texts, compiled according to some organ- izing principles, is called a corpus. In this section we focus on how we can use **corpora** (plural form of corpus) to carry out terminological and phraseological research, both monolingually and bilingually, in preparation for translation.

Corpus software or corpus tools help us to extract data of various kinds from a corpus. Typically, corpus users are interested in gathering data about patterns of language usage, frequencies of use and informa- tion about likely co-occurrence of words, terms or phrases. Likely co-occurrences may be categorized into **collocations**, the co-occurrence of two or more words, and **colligation**, the occurrence of grammatical structures with other structural patterns or words. Frequency and co-occurrence data about both SL and TL can be very useful to us in our translation preparation and translation production. If our corpus texts belong to a specialized genre, then we are likely to be able to extract usage data which will tell us something about the terminology and phraseology of that genre, as well as terminology related to the particular concepts belonging to that subject domain.

To illustrate this kind of corpus-based terminological research in very concrete terms, we will work through a specific example below. The following tasks have been completed using Sketch Engine, www. sketchengine.co.uk (Kilgarriff et al. 2014) which is available via monthly individual or annual institutional subscriptions, and offers a 30-day free trial. Sketch Engine provides an online corpus query interface and access to many pre-existing large corpora, monolingual and multilingual, as well as functions to enable you to create your own corpora. However, other corpus software could be used to perform some or all of these tasks; the principles of the terminological research are more important here than the choice of specific tools. Alternative corpus tools in widespread use include Wordsmith Tools (Scott 2015) and the freeware AntConc (Anthony 2014).

Before we start on our terminological research, it would be useful for you to explore and familiarize yourself with the main tools that corpus software can offer you. Concordancing is a key function of all corpus software, enabling users to find patterns of usage through a visual display. A concordance tool extracts all instances of a node or search word or phrase from the corpus, together with a short excerpt of the co-text to its left and right. It then displays the lines of text, with the node or search word or phrase highlighted, aligned and cen- tred on the screen, with its co-text either side of it. This is also called a **concordance** or a **KWIC** (Key Word in Context) display. If you have access to Sketch Engine, select any pre-existing corpus and use the ‘Concordance tool’ to generate data from that corpus, by searching for word forms of your choice. If you do not have access to Sketch Engine, search the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA) using the online concord- ancing tool at corpus.byu.edu. If you would like to learn more about how concordances can be used in linguistics research, consult Sinclair (2003). The other key tools of most corpus software are word lists and keyword lists, both of which we will use in the task below.

Let us assume that the text we have been commissioned to translate is a report issued by the European Union’s Joint Research Centre (JRC) in 2014, entitled ‘Science for Nuclear Safety and Security’, available at ec.europa.eu/jrc/en/publication/science-nuclear-safety- and-security. As noted in the accompanying blurb, the report

aims to give a comprehensive overview of the JRC’s work in rela- tion to nuclear safety and security. It highlights the JRC’s rele- vant scientific output in nuclear safety; nuclear security; reference measurements; materials and standards; nuclear knowledge management; training and education and, in the last chapter, innovation.