

Causation in a Learner Corpus of Hungarian-English Summaries of Disease Descriptions

Anita HEGEDŰS

Department of Languages for Biomedical Purposes, Medical School, University of Pécs

Abstract: *A learner corpus is a collection of language data stored electronically. This paper undertakes to investigate how Hungarian L2 learners of English use causal relationships in a learner corpus of summaries of disease descriptions submitted by Hungarian medical students at level B2 (upper intermediate). The 'UPMS Learner Corpus of Summaries' has been expanding since March, 2022. The data were processed using Sketch Engine. At present, the corpus contains 108 summaries, 24,321 tokens, 21,397 words and 1,244 sentences. In the corpus, the different forms of the lexical item 'cause' were found to be by far the most common ways of expressing causal relations. It was the 29th most common lexical item and the 10th most common content word in the corpus. Other common ways of expressing this function were the lexical items 'result' and 'reason', with 17 and 14 appearances, respectively. However, all the occurrences of 'reason' were used incorrectly in semantic terms and in 8 cases there was a syntactic mistake, too. The implication of the study for teaching medical English is that more emphasis should be put on the semantic distinction between lexical items expressing causation as well as on their syntactic patterns.*

Keywords: *causation; learner corpora; summary; disease descriptions; Sketch Engine.*