

PRODUCTIVE SYNTACTIC MODELS

Kedrove G.E.¹, Potemkin S.B.

Lomonosov MSU Philological faculty tel +79459164723 e-mail prolexprim@gmail.com

¹Lomonosov MSU Philological faculty tel +79459164723 e-mail kedr@philol.msu.ru

While assessing the effectiveness of the educational material for the foreign students within the framework of the "Russian as the foreign language" training course, it is important to find out which syntactic structures proposed for study are most commonly used in daily communication.

The source material is a Russian-Chinese phrase book, which is a set of words and sentences in two languages (in our case, Russian and Chinese). Tokenization and lemmatization of the Russian text and definition of POS and grammatical features (gender, number, case) of each word form was carried out. Then a superficial syntactic analysis of each sentence was performed on the basis of Nivre's algorithm. The resulting syntactic links are sorted according to the frequency of their occurrence. Expectedly, the most frequent structure is the subject-verb-object (SVO) form. Nominal groups are mainly represented by such forms as noun (nominal) – noun (genitive); possessive pronoun – noun; adjective – noun.

Each syntagma is substituted into the request to the National Corpus of Russian Language (NCRL). In response, a list of all sentences containing a pair of words of this syntagma at a given distance (from 1 to 3) and the total number of such sentences is obtained. The next task was to determine the productivity of each structure. Then we will replace the word contained in the most common syntagma with the other word from the phrase book. The syntagmas obtained as a result are also checked for the use of the NCRL. The sum of frequencies for each such syntagma determines its productivity and, consequently, the appropriateness of its inclusion in the educational material.