

**Dmitry Yu. Gruzdev, Dmitry O. Kodzhebash**  
**Military University, Russian Ministry of Defense**  
**Moscow, Russian Federation**  
gru@inbox.ru

## **POS-powered queries for neat and lean concordances in ad-hoc corpora analysis**

### **Abstract**

The paper addresses the ways of boosting the query effectiveness in ad-hoc, a.k.a. DIY corpora. Based on the established trend of abandoning KWIC approaches in favor of pattern searches in the practice of professional translators, we hypothesize that advanced methods, including regular expressions and annotation, have a potential of accommodating them. The former has already proven an efficient text analysis tool for programmers while the latter takes credit for today's NLP technologies. It is the aspect that can be harnessed for increasing the efficiency of performing translation tasks. The obvious problem with ad-hoc corpora is that they rarely mature beyond the raw corpus status. However, tagging automation makes further improvements desirable. To grasp the potential of a tagged DIY corpus, we picked a random collection of news texts published in 2019–2020 and subjected it to automatic POS-tagging, acting as a computing lingua franca for more comprehensive and less constraining queries. The primary goal was to retrieve patterns based on the parts of speech of constituent words rather than anchor the search efforts to specific words. Taking advantage of existing tools, namely AntConc and TagAnt, we test the hypothesis. This confirmed the relevance of POS tagging of ad-hoc corpora for template extraction of linguistic data by translators. The resulting concordances, more concise and relevant, go a long way in keeping the analysis time lower, hence better overall efficiency.

**Keywords:** tags, POS, regular expressions, information retrieval, query, annotated corpus

© Gruzdev D. Yu., Kodzhebash D. O. 2023

**For citation:** Gruzdev, D. Yu., Kodzhebash, D. O. (2023). POS-powered queries for neat and lean concordances in ad-hoc corpora analysis. *Teoreticheskaya i prikladnaya lingvistika* [Theoretical and Applied Linguistics], 9 (4), 35–48. [https://doi.org/10.22250/24107190\\_2023\\_9\\_4\\_35](https://doi.org/10.22250/24107190_2023_9_4_35)