

Data driven research in linguistics

Dr Wojciech Witkowski

Last two decades have witnessed significant advances in computational development. This has led to the shift in research practice that affected humanities as well. Due to a rapid growth of digital archives enabling quick and user-friendly access to language data, data driven research paradigms have been developed. This seminar offers training in data-based linguistics. Students will get familiar with various types of data, data sources and methods of collecting data. This will equip them with the knowledge and skills necessary for gathering and analyzing data. They will also learn how to explain the linguistic facts illustrated by the collected data. Moreover, students will be exposed to the methodology of original research in such areas of linguistics as translation studies and second language teaching.

Topic areas covered

- quantitative and qualitative research methodology
- applications of Polish-English language resources in contrastive language studies
- quantitative analysis of the relation between lexicon and syntax
- Corpus-based semantic analysis
- empirical and comparative linguistics
- using Web as a corpus – getting good data from the Internet
- Data Driven Learning
- machine translation
- Natural Language Processing
- language and data analysis software

Sample thesis topics:

- Idioms in translation: a PARALELA corpus based analysis of English animal based idioms in Polish translations.
- How well does the target language reflect the source language? Parallel corpus based study of selected linguistic features of Polish-to-English translations and English usage.
- Polish-English lexical gaps: Princeton WordNet and Polish WordNet based analysis of sports vocabulary.
- Creating a term base for Polish-English-Polish technical translation: data sources and methodological considerations.
- Are verbs *to roll* and *to break* more likely to be found in sentences such as *X rolled/broke Y* or *Y rolled/broke.*: A corpus based analysis

- How frequent are *build* verbs in clauses such as *John built Mary a house* and *John built a house for Mary*.
- How do verbs of emotions behave in Polish and English sentences: a comparative analysis of English and Polish verbs' argument realization patterns.
- *To turn, to turn the page* and *to turn the tides of war* – how are these verbs different?: A COCA corpus based analysis.
- Comparison of FrameNet and Levin's verb classes: linguistic theory and language resources.
- Surfing and Łomzing: a Web based analysis of English *-ing* suffix occurrence in Polish.
- Teaching phrasal verbs to Polish learners. Applications and limitations of Data Driven Learning approach.
- Expanding students' lexical skills: a practical implementation of corpus data in language teaching.

Credit:

1st semester: submitting research problem, research hypotheses and thesis plan

2nd semester: submitting the framework chapter of the thesis

3rd semester: submitting the methodological chapter

4th semester: submitting the analytic chapter