University of Warsaw
Faculty of Psychology

Kamil Tomaszek
Record book number: 432044

Middle Polish Dependency Treebank in Universal Dependencies format: Design, Implementation, and Analysis

Master's thesis
in COGNITIVE SCIENCE

The thesis was written under the supervision of
Dr. Alina Wróblewska
Institute of Computer Science
Polish Academy of Sciences
Dr. Grzegorz Krajewski
University of Warsaw

Warsaw, September 2025

Summary

This thesis presents a rule-based approach to converting the Middle Polish Dependency Treebank (MPDT), annotated in a Polish-specific scheme, into the Universal Dependencies (UD) format. After introducing the project motivation, data sources, and target standard, the thesis outlines general design assumptions behind the conversion, the mapping strategy, and the validation workflow. It reports overall outcomes of the conversion and sketches applications and extensions, including releasing MPDT-UD and implications for research in historical language processing within cognitive science.

Keywords

Middle Polish, dependency trees, treebank conversion, Universal Dependencies

The title of the thesis in Polish

Średniopolski Bank Drzew Zależnościowych w formacie Universal Dependencies: projekt, implementacja i analiza

Streszczenie

Praca przedstawia podejście regułowe do konwersji Średniopolskiego Banku Drzew Zależnościowych (MPDT), anotowanego w polskim schemacie, do formatu Universal Dependencies (UD). Po krótkim omówieniu motywacji, danych i standardu docelowego zaprezentowano ogólne założenia projektu, strategię odwzorowań oraz schemat walidacji. Przedstawiono ogólne wyniki konwersji oraz możliwe zastosowania i kierunki rozwoju, w tym udostępnienie MPDT-UD i znaczenie dla badań nad przetwarzaniem języka historycznego w kognitywistyce.

Słowa kluczowe

język średniopolski, drzewa zależnościowe, konwersja korpusu, Universal Dependencies

The title of the thesis in English

Middle Polish Dependency Treebank in Universal Dependencies format: Design, Implementation, and Analysis

Contents

Chapter 1

Introduction

1.1. Project Overview

1.2. Context and Scope

1.3. Objectives and Contributions

1.4. Thesis Outline

Chapter 2

Linguistic Background

2.1. Overview of Middle Polish

2.2. Grammatical and Syntactic Characteristics

2.3. Annotation Assumptions (High-Level)

2.4. Summary

Chapter 3

Conversion Approach

3.1. Design and Assumptions

3.2. Mapping Strategy (POS, Features, Relations)

3.3. Implementation Overview

3.4. Processing Workflow

Chapter 4

Validation and Results

4.1. Data and Procedure

4.2. Validation Setup

4.3. Overall Findings

4.4. Limitations

Chapter 5

Applications and Future Work

5.1. Resource Release and Integration

5.2. Use Cases

5.3. Extensions

Chapter 6

Cognitive Science Perspective

6.1. Relevance to Cognitive Science

6.2. Implications