Open-Source Morphology for Endangered Mordvinic Languages

Abhinav S Menon 2020114001

Overview

Introduction

Designing for a Reusable API

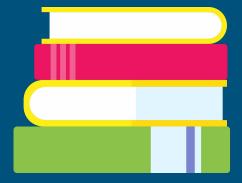
Erzya and Moksha Language Pair

Current State and Ensuring Maintainability

Conclusions

Introduction

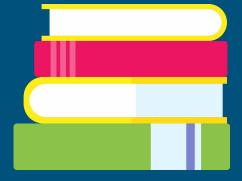
- Mordvinic languages are endangered
 - Erzya and Moksha
- In this paper
 - Open-source FST-based morphological analyser
 - "Highlight importance of design decisions"
 - Describe how to contribute
- Previous work on Erzya and Moksha



Designing for a Reusable API

- Open source is preferable
 - Different groups independently create same tools with different APIs

- GiellaLT framework
 - HFST
 - Supported by Apertium



Erzya and Moksha Language Pair

- Structure of Apertium
 - Morph analysis
 - Morpheme mapping
 - Morph generation

Paradigms differ in some cases



Current State & Ensuring Maintainability

Current State

- Left out
 - Declension of nominals
 - Orthographic compounds
- Method of development
 - o Agile:)
 - Unit testing

Future Contribution

- Anyone should be able to contribute
- XML-based dictionary
 - User-friendly UI
- Non-tech-savvy users can contribute

Conclusions

Core design principle: Open-source

Accessibility in usage & contribution

Views

(Almost) Not a linguistics/CL/NLP paper

Lots of fluff

- "Moksha has a symmetric paradigm in the core cases, nominative, genitive and dative (see Table 2), with distinct word forms for each case-possessor combination slot and an additional distinction for number of possessa when the possessor is singular, whereas Erzya makes virtually no consistent distinction for case or number in the nominative and genitive with one exception the specific nominative singular reading of the third person singular +N+SG+NOM+PXSG3 and option- ally in the first and second person singular of modern written literature (see Table 3)."
- Motivation, development process, API, interface
 - Sidesteps actual description of FSTs

References

Jack Rueter, Mika Hämäläinen, and Niko Partanen. 2020. Open-Source
Morphology for Endangered Mordvinic Languages. In Proceedings of Second
Workshop for NLP Open Source Software (NLP-OSS), pages 94–100, Online.
Association for Computational Linguistics.