ParlaMint-RO: Chamber of the Eternal Future

Petru Rebeja\textsuperscript{1} Mădălina Chitez\textsuperscript{2} Roxana Rogobete\textsuperscript{2} Andreea Dincă\textsuperscript{2} Loredana Bercuci\textsuperscript{2}

Language Resources and Evaluation Conference
June 20, 2022

\textsuperscript{1}Alexandru Ioan Cuza University of Iaşi

\textsuperscript{2}West University of Timişoara
Introduction
Objectives

- Study the distribution of future tense forms across parliamentary groups,
- Perform a quantitative analysis of large corpus,
- Exemplify how the data can be used for interdisciplinary studies.
Data Collection: Parliamentary Records
Processing Transcriptions

- Transcripts from HTML format are converted to XML format required by ParlaMint schema using the lxml library\(^1\).
- On-the-fly normalization of text is applied to extracted data.

\(^1\)https://lxml.de
Corpus-Specific Metadata

• Names, gender, and picture of deputies were scraped from the website of the Lower House\(^2\)

• Scraping affiliation data was unsuccessful; as such, no affiliation data is included in the corpus.

\(^2\text{http://www.cdep.ro/}\)
### Basic statistics of ParlaMint-RO corpus

| Level                               | Value                  |
|-------------------------------------|------------------------|
| Number of transcribed sessions      | 1,832                  |
| Number of processed speeches        | 552,103                |
| Number of words                     | 109,304,196            |
| Period                              | 2000 – 2020            |
| Source                              | Lower House            |
Data Analysis
Data Sources of Future Tense Usage Analysis

- ParlaMint-RO transcriptions,
- List of verbs from https://dexonline.ro\(^3\),
- Inflected forms of verbs for *formal future tense* from https://conjugare.ro.

\(^3\)The list is restricted only to dictionary entries that are not under Intellectual Property rights.
The data sets were joined and normalized into two sets of tuples:

\[(\text{speaker}, \text{date}, \text{count of all forms}, \text{count of all words})\]
\[(\text{speaker}, \text{date}, \text{verb form}, \text{count})\]
Data Aggregation

The tuples were aggregated using *pandas*\(^4\), and *matplotlib*\(^5\) was used for visualization.

The Python scripts, alongside the collection of verb forms are available on Github\(^6\).

\(^4\)https://pandas.pydata.org
\(^5\)https://matplotlib.org
\(^6\)https://github.com/romanian-parlamint/future-tense-usage
Results
| Speaker                                | Count  | Percentage |
|----------------------------------------|--------|------------|
| Valer Dorneanu                         | 10,859 | 0.73       |
| Tudor Ciuhodaru                        | 7,842  | 1.55       |
| Emil Boc                               | 4,480  | 1.48       |
| Valeriu Ștefan Zgonea                  | 4,190  | 0.51       |
| Florin Iordache                        | 3,837  | 0.62       |
| Adrian Moisoiu                         | 3,775  | 0.98       |
| Doru Ioan Tărăcilă                    | 3,602  | 0.73       |
| Gheorghe-Eugen Nicolăescu              | 3,421  | 1.40       |
| Nicolae Văcăroiu                       | 3,398  | 0.85       |
| Bogdan Olteanu                         | 3,268  | 0.76       |
Top 15 Inflections In Future Tense Used In Lower House

| Form       | Meaning                              |
|------------|--------------------------------------|
| va fi      | (he/she) will be                     |
| vor fi     | (they) will be                       |
| va trebui  | will have to                         |
| va avea    | (he/she) will have                   |
| vor avea   | (they) will have                     |
| vom avea   | (we) will have                       |
| vom fi     | (we) will be                         |
| va putea   | (he/she) will be able to             |
| vor putea  | (they) will be able to               |
| va veni    | (he/she) will come                   |
| vor veni   | (they) will come                     |
| va vota    | (he/she) will vote                   |
| voi fi     | (I) will be                          |
| veşi observa | (you) will observe               |
| va legisla | (he/she) will legislate              |
| vor veni   | (they) will come                     |
Conclusions

• The top 15 inflections in future tense reveals only three forms in first person, which suggests an impersonal tone related to shifting responsibility onto others.

• Another sign of projection apparent in the deputies’ speeches is the frequent use of “ar trebui să”, a conditional tense that can be translated with “should”.
Thank you!