

# Estonian Parliament Speeches as Source for Content Analysis



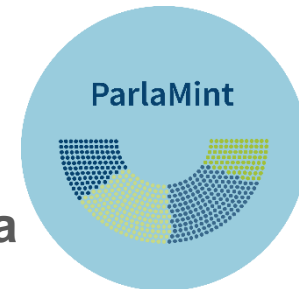
ParlaMint

Martin Mölder, Neeme Kahusk, Kadri Vider  
University of Tartu, Estonian Literary Museum

**8TH ESTONIAN DIGITAL HUMANITIES CONFERENCE**

**October 5-7, 2022, Tallinn University, Estonia**

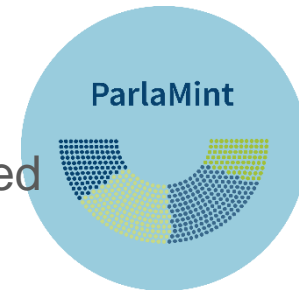
# ParlaMint corpora of parliamentary speeches



- [CLARIN-ERIC](#) initiative, contributes to the **creation of comparable and uniformly annotated multilingual corpora of parliamentary sessions**.
- [ParlaMint I & II](#) - data from **27 European parliaments**: Austria, Basque Country, Belgium, Bulgaria, Catalonia, Croatia, Czech Republic, Denmark, **Estonia**, Finland, France, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Slovenia, Spain, Sweden, Turkey, United Kingdom
- Not only text, but also **rich metadata** (mandates, sessions and speakers, their political party affiliations etc.).
- **Annotation layers** for linguistic and other (e.g content) data analysis (Named Entities, sentiment analysis etc)

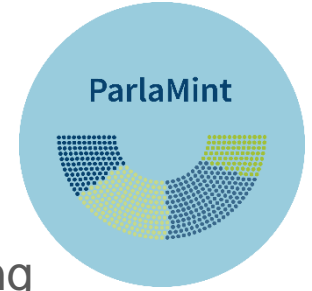
Existing alternatives: ParlSpeech 9 countries (Europe plus NZ), from 21 to 32 years.

# Estonian ParlaMint corpus



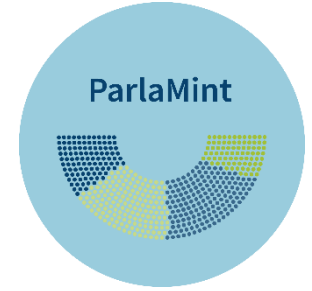
- Previous Estonian parliamentary speeches datasets mentioned in [CLARIN Resource Families](#): **13 million, tokenized, 1995-2001**, concordance via [KORP](#)
- Estonian ParlaMint corpus - a **full dataset** of Estonian parliamentary texts with meta-information about MPs **from 1992 onwards**.
- **Size**: 8 parliament compositions, 1283 unique speakers, more than 50 million tokens/words.
- **Meta-information**: speaker status (MP, government, guest), gender, birth, parliamentary group, election info (list, mandate, district, vote).
- **Annotation layers**: lemma, part-of-speech, morphological features, syntax, named entities

# Content in Many Formats



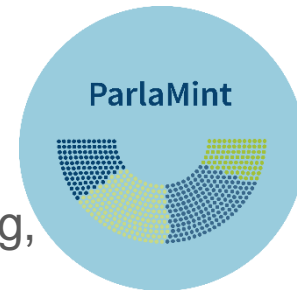
- Text only — the most human-readable form
- CoNLL-U — Morphological and Syntactic analysis according to Universal Dependency
- Vert — for corpus search systems like SkechEngine or Korp
- XML markup according to TEI — for archiving and further annotation purposes
- In addition to standard UD analysis, Named Entity layer is added.

# Workflow for Estonian Data



- Speech data from <https://stenogrammid.riigikogu.ee/et>
- MP data from parliament archives, election archives
- Government data form a separate government composition data set
- Json file: text with metadata
- From json to other formats
- Morphological analysis and syntax analysis with EstNLTK and Stanza, Named Entities with EstNLTK.

# Possible use cases of parliament speeches corpora



**Linguistics:** words, word forms, syntax, word sense and meaning, structure of edited speech...

**Discourse analysis:** dialogue structure, topic & comment, transitions, sentiment analysis...

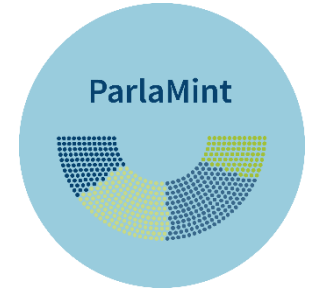
**Language technology:** lot of multilingual data for MT and language models, summarisation tools...

**Information extraction:** named entities (persons, places, organisations etc), timeline, topical classification, keywords...

**Political sciences:** populism, affective polarization, parliamentary speech as parliamentary behaviour

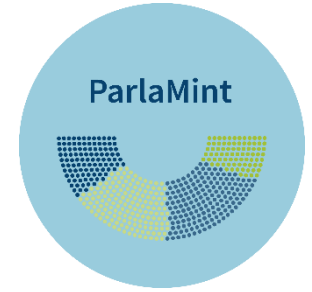
**=> Pile of textual data with interesting content <=**

# Measuring populist speech



- Populism - an essentially contested concept (and a very widespread political “weapon”)
- Ideational approach to populism:
  - A thin-centered ideology
  - Three components
    - Anti-elitism
    - People-centrism
    - Politics as a battle between good and evil
  - The opposite: technocratic pluralism

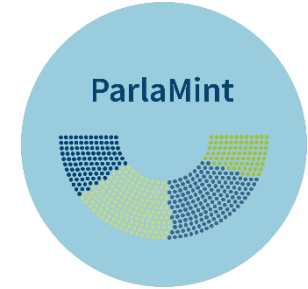
# Measuring populist speech



- A classification problem - a statement / text is either populist or not.
- Most comprehensive available data set - Global Populism Database
- Training data - Global Populism Database coding materials
- The EstBERT model - fine-tuned for this particular task.
- The “final initial” model: ~73% accuracy in classifying populist/pluralist sentences.

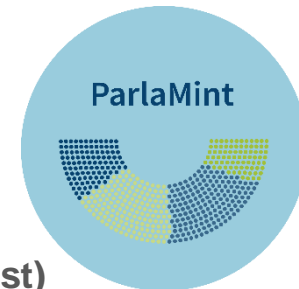


# Measuring populist speech



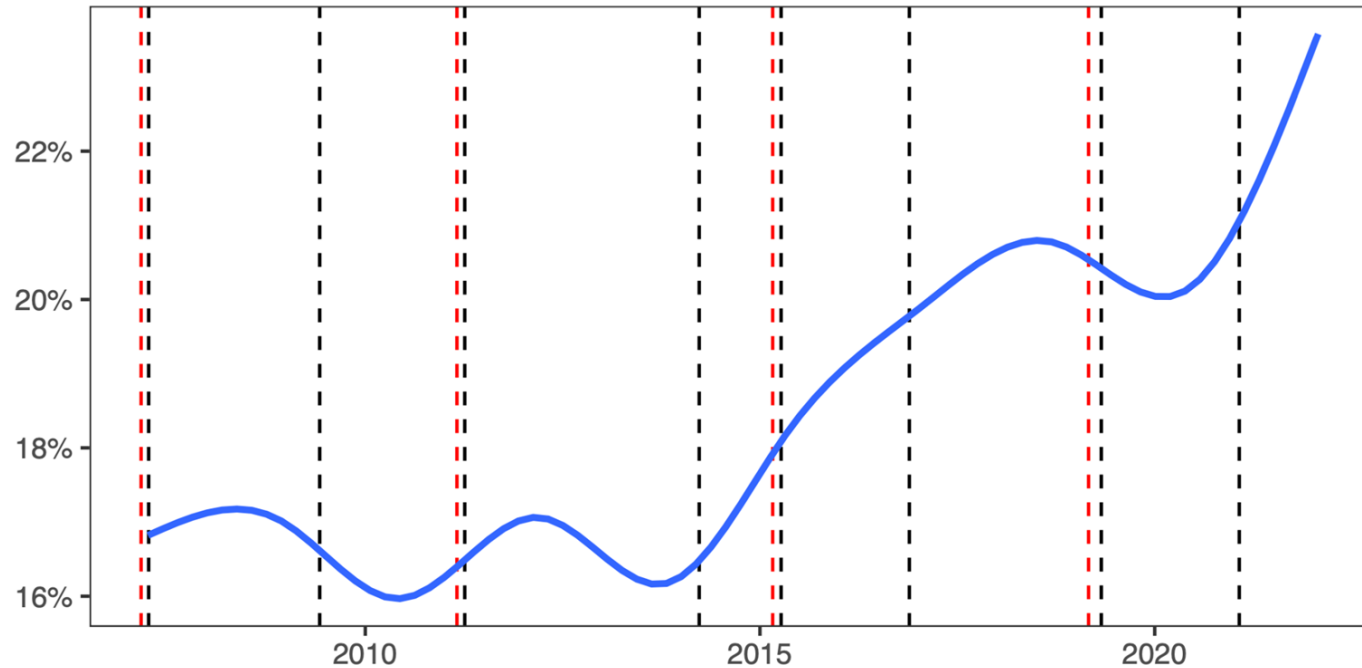
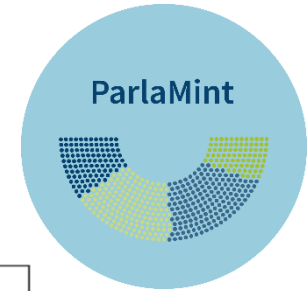
- *“Rahvas eelkõige! Kui peab valima poliitika ja rahva vahel, tuleb valida rahvas; kui peab valima enda ja rahva vahel, tuleb valida rahvas; kui peab valima rahvaste vahel, tuleb valida oma rahvas, austades teiste rahvaste samasuguseid õigusi!” - 98.8% populist*
- *“Sellise olukorrani on viinud riigijuhtimise ebademokraatlik tsentraliseerimine, meedia monopoliseerimine, võimukandjate ennastkiitev demagoogia, korrumpeerunud kartellipoliitika, majanduslangus ja Eesti kapitalile kuuluva majanduse järkjärguline väljasuretamine, Eesti muutmine odavtööjõu maaks, välisfirmade kasumi maksuvaba väljaviimine Eestist, omavalitsuste tulubaasi kärpimine, kohaliku elu väljasuretamine.” - 99.8% populist*

# Measuring populist speech

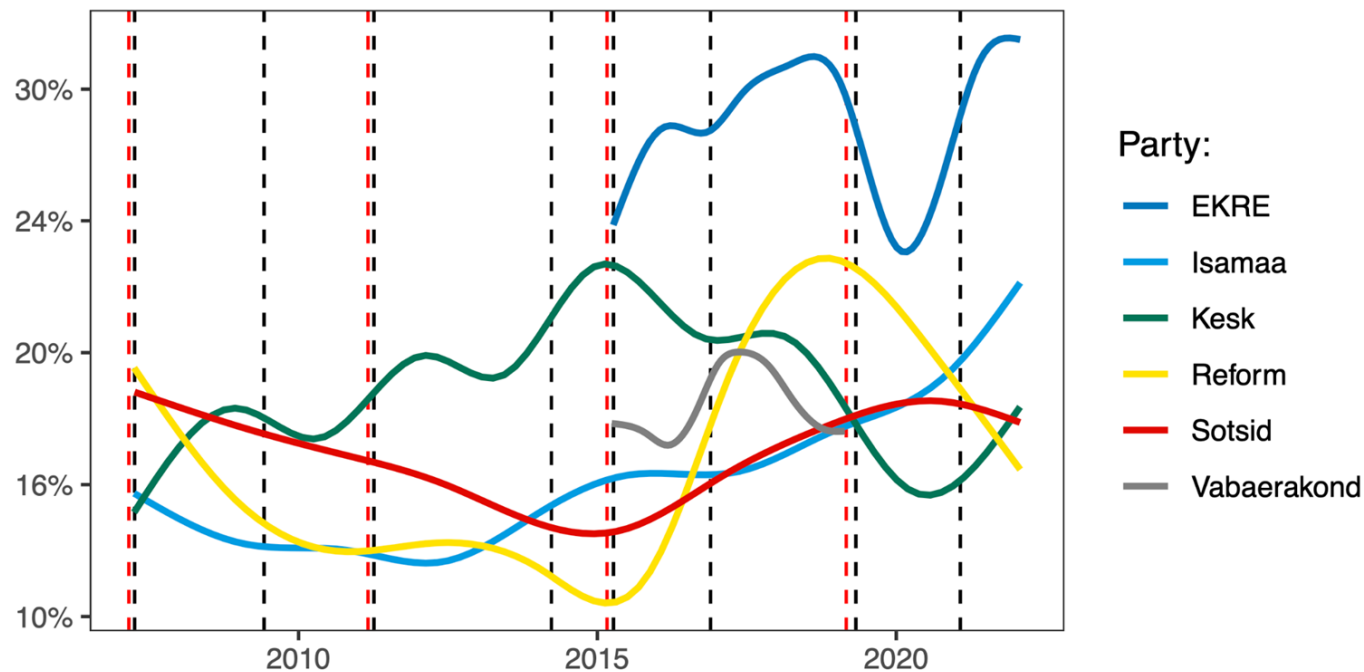
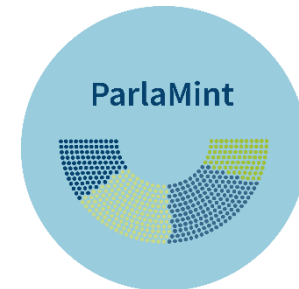


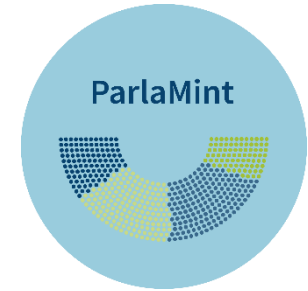
- “*Elektri hind sõltub...*”
  - *... nõudluse ja pakkumise tasakaalust.*” **99.8% non-populist (pluralist)**
  - *... rumalate poliitikute suvast.*” **95.2% populist**
  - *... kõrgelt haritud ja kompetentsete ekspertide hinnangust.*” **99.9% non-populist**
- Examples from the speeches of presidents:
  - “*Täname neid, kes sellest põrgust läbi tulid ja oma kodumaa vabadust südames kandsid, vaatamata pimedatele aegadele ja kurjadele jõududele.*” **99.9% populist ??**
  - “*Sest tegelikult olid rahva tahe ja vabadussoov nendeks käärideks, mis käristasid lahti võõra impeeriumi õmblused.*” **99.9% populist ??**
  - “*Võõra võimu öine kalk prõmmimine uksele määras saatuse.*” **99.9% populist ??**
  - “*Ühishanked ja ühised kaitseinvesteeringud võivad aidata, aga tasuks mõelda ka tavapärase Euroopa ümberjagamismudelite kasutamisele kaitsekulude valdkonnas.*” **0.0001% populist**
  - “*Intensiivsed üleminekuprotsessid, mida on läinud kümnendil läbinud samuti Kasahstan ja paljud teised riigid, lõppesid Eestis kuu aega tagasi Euroopa Liidu liikmeks saamisega.*” **0.0002% populist**

# Populist Speech in the Estonian Parliament, 2011 - 2022



# Populist Speech in the Estonian Parliament, 2011 - 2022





Thank you for attention!

[martin.molder@ut.ee](mailto:martin.molder@ut.ee)

[neeme.kahusk@ut.ee](mailto:neeme.kahusk@ut.ee)

[kadri.vider@eesti.ee](mailto:kadri.vider@eesti.ee)