

# Development and Research in Lithuanian Language Technologies (2016-2020)

2020

## AUTHORS:

Andrius Utkā  
Vytautas Magnus University

Jurgita Vaičėnė  
Vytautas Magnus University

Monika Briedienė  
Vytautas Magnus University  
Baltic Institute of Advanced Technology

Tomas Krilavičius  
Vytautas Magnus University  
Baltic Institute of Advanced Technology

## AIM

To survey Lithuania's involvement in international and national LT related initiatives and research infrastructures, and overview the advancements in language resources and technologies, as well as key projects and research

## PARTICIPATION IN EUROPEAN LT INITIATIVES

- European Federation of National Institutions for Language (EFNIL)
- European Language Resource Coordination (ELRC),
- European Open Science Cloud (EOSC)
- FAIR data
- EU Declaration on Cooperation on Artificial Intelligence/Lithuanian Artificial Intelligence Strategy
- COST actions (IC1408, IC1207, CA18209, IC1002, CA18209)

## LARGE SCALE PROJECTS

"The Lithuanian Information Society Development Programme 2014 2020":

- Lithuanian Language Speech Services (LIEPA 2 <https://raštija.lt/liepa-2/apie-projektq-liepa-2/7666>)
- The Information System for Syntactical and Semantic Analysis of the Lithuanian Language (SEMANTIKA 2 <https://semantika.lt>)
- Integrated Information Systems of the Lithuanian Language and Language Resources (Raštija 2 <https://raštija.lt>)
- Modernization and Development of Machine Translation Systems and Localization Services
- Lithuanian Language Resources (E.kalba <https://ekalba.lt>)

## MAIN RESULTS OF LARGE SCALE PROJECTS

- syntactically annotated corpus (ALKSNIS)
- morphologically annotated corpus (gold standard) (MATAS)
- 1,000 hour speech corpus
- Internet corpus (BIT)
- modernized corpora and lexicons
- neural MT system (Lithuanian, German, Polish, French, and Russian language pairs)
- automatic transcription of speech files
- speech recognition of computer commands
- Lithuanian language synthesizer
- advanced internet search
- automatic text summarization
- hate speech detection

## LANGUAGE TECHNOLOGY INFRASTRUCTURES

CLARIN LT  
<http://clarin.lt>

Lithuania is a full member of CLARIN ERIC since 2014. Consortium members:

- Vytautas Magnus University (Centre of Computational Linguistics)
- Kaunas Technology University
- Vilnius University
- Mykolas Romeris University
- Baltic Institute of Advanced Technologies.

ELG  
<https://www.europeanlanguagegrid.eu>

- In 2019, the Institute of Lithuanian Language became a member of EU funded project European Language Grid (2019-2021)

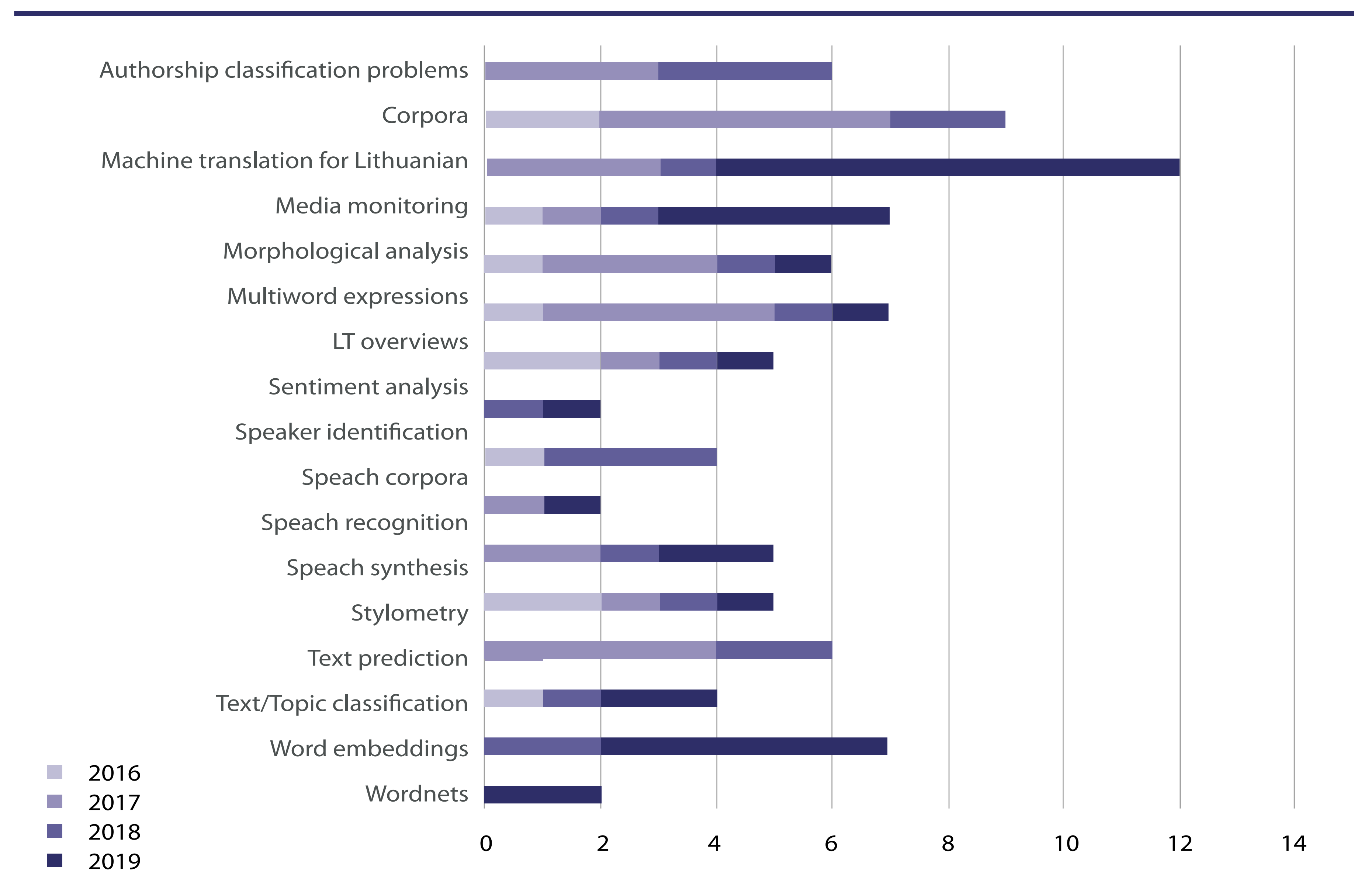
## National Infrastructures

- Raštija LT <https://raštija.lt>
- LKSSAIS <https://semantika.lt>
- E.kalba <https://ekalba.lt>

## RESEARCH

The analysis of scientific papers of Lithuanian language technologies shows that machine translation and speech technologies were the most trending research topics in 2016-2019.

Full Bibliography of Research Papers  
<https://clarin.vdu.lt/xmlui/handle/20.500.11821/38>



## CONCLUSIONS

Developments contributing to the integration of Lithuanian as less resourced language in the digital environment are as follows:

- inclusion of language technology problems into strategic documents and legislation
- development of national and international language technology infrastructures

- promotion of the international visibility of language resources or technologies adapted to or created for the Lithuanian language
- implementation of Lithuanian language services in large scale projects,
- offering knowledge sharing services to interested parties
- development of open access possibilities of language technologies and resources

"[...] there is currently a widening technology gap between well resourced languages and less resourced languages"

The European Parliament resolution on "Language Equality in the Digital Age" (09/11/2018)

