Jurgita Jaroslavienė/Albina Auksoriūtė

Innovations and challenges in the digital transformation of the Lithuanian language industry

Abstract (Lithuanian)

Pastaraisiais metais žinių visuomenė pereina į kokybiškai naują etapą, išsiskiriantį sparčia inovatyvių informacinių technologijų plėtra. Kalbos technologijos – tai tik viena, tačiau itin svarbi informacinių technologijų dalis siekiant užtikrinti visavertį lietuvių kalbos vartojimą skaitmeninėje terpėje, įtvirtinti ir palaikyti lietuvių kalbos statusą informacinėje visuomenėje, sudaryti galimybes kitakalbiams integruotis į Lietuvos visuomenę ir pan.

Lietuvių kalbos institutas, atsižvelgdamas į informacinės visuomenės lūkesčius ir poreikius, prie informacinių technologijų plėtros prisideda gausindamas skaitmeninius lietuvių kalbos išteklius, kurdamas inovatyvius internetinius mokymosi įrankius ir pažangias viešąsias elektronines paslaugas, plėtodamas internetinę lietuvių kalbos išteklių infrastruktūrą (http://lki.lt/skaitmeniniai-lietuviu-kalbos-istekliai/).

Dabartinę Lietuvių kalbos išteklių informacinę sistemą (http://lkiis.lki.lt/) sukūrė ir 2015 m. visuomenei pristatė Lietuvių kalbos institutas kartu su Lietuvių literatūros ir tautosakos institutu, Vilniaus universitetu ir tuomečiu Lietuvos edukologijos universitetu pagal programos *Lietuvių kalba informacinėje visuomenėje* projektą. Lietuvių kalbos išteklių informacinės sistemos svetainėje šiuo metu galima naudotis suskaitmenintais 16 vienakalbių ir dvikalbių žodynų, suskaitmenintomis 6 *Lietuvių kalbos žodyno* ir tautosakos paveldo kartotekomis, 4 elektroninėmis lietuvių kalbos išteklių duomenų bazėmis. Šioje sistemoje vartotojas ras ir lietuvišką šriftą *Palemonas*, ir kelias lietuvių kalbai mokytis skirtas priemones (pvz., mokomuosius kalbos žaidimus, skatinančius mokytis savarankiškai, įtvirtinti lietuvių kalbos rašybos, kirčiavimo įgūdžius, gausinti raiškos priemonių atsargas).

Atsižvelgdami į pasaulinę praktiką ir didėjantį naujų pažangių elektroninių paslaugų poreikį bei siekdami, kad lietuvių kalba visavertiškai gyvuotų elektroninėje erdvėje, Lietuvių kalbos instituto mokslininkai kartu su pažangiųjų informacinių technologijų kūrėjais pagal naują projektą, finansuojamą iš Europos Sąjungos struktūrinių fondų, nuo 2018 m. pradėjo Lietuvių kalbos išteklių informacinės sistemos plėtros darbus. Šiuo metu į modernizuojamąją Lietuvių kalbos išteklių informacinę sistemą, kuri vadinasi "E. kalba" (http://lkiis.lki.lt/ → https://ekalba.lt/public#/home/main), integruojami dar 3 reprezentatyvūs vienakalbiai lietuvių kalbos ištekliai, kurie iki šiol buvo prieinami tik atskirose interneto svetainėse: suskaitmenintas didysis 20 tomų *Lietuvių kalbos žodynas*, internetinis *Bendrinės lietuvių kalbos žodynas* su antraštynu ir internetinis *Lietuvių kalbos naujažodžių duomenynas*. 2020 m. kuriamas lietuvių kalbos žodžių prasmių tinklas ir naujos elektroninės paslaugos.

Lietuvių kalbos ištekliai internete ir kuriamos naujos galimybės itin aktualios, parankios ir pravarčios visiems lietuvių kalbą vartojantiems ar norintiems jos išmokti žmonėms, taip pat valstybės institucijoms, mokslo ir švietimo sistemai bei verslui.

Be to, Lietuvių kalbos instituto mokslininkai aktyviai dalyvauja įvairiuose nacionaliniuose ir tarptautiniuose projektuose. Vienas iš tokių projektų – *eTranslation TermBank* (finansuojama Europos Sąjungos). Pagrindinis šio projekto tikslas yra pagerinti automatinio vertimo kokybę Europos Komisijoje.

Šiame straipsnyje trumpai aprašoma ir *Lietuvos Respublikos terminų banko* plėtra (šio banko informacinės sistemos duomenų administratorius ir tvarkytojas yra Valstybinė lietuvių kalbos komisija, žr. http://terminai.vlkk.lt/) bei kiti inovatyvūs skaitmeniniai sprendiniai Lietuvoje, išvardijami pagrindiniai iššūkiai, su kuriais Lietuvoje susiduriama kuriant kalbos technologijas, internetinius įrankius ir viešąsias elektronines paslaugas.

Abstract (English)

In recent years, the knowledge society has entered a qualitatively new stage, distinguished by the rapid development of innovative information technologies. The digital transformation of the Lithuanian language industry is just one of the extremely important roles of information technologies to help ensure proper use of Lithuanian in the digital environment, to establish and maintain the status of Lithuanian in the information society and to create opportunities for non-native speakers to integrate into Lithuanian society, etc.

Considering the expectations and needs of the information society, the Institute of the Lithuanian Language contributes to this successful digital transformation and manages the development of online infrastructure (http://lki.lt/skaitmeniniai-lietuviu-kalbos-istekliai/).

The current information system (online infrastructure) of resources for Lithuanian (http://lkiis.lki.lt/) was developed and launched in 2015 by the Institute of the Lithuanian Language in association with the Institute of Lithuanian Literature and Folklore, Vilnius University and the University of Educational Science of Lithuania as part of a project in the programme called "The Lithuanian Language in an Information Society". Their website currently offers the following resources: 16 digitalized monolingual and bilingual dictionaries, six digitalized catalogues of the great *Dictionary of the Lithuanian Language* and folklore heritage, four electronic databases of Lithuanian language resources. The website also provides the Lithuanian font *Palemonas* and several tools for learning Lithuanian (for example, educational language games to encourage independent learning, to promote Lithuanian spelling and stress patterns and to enrich the stock of means of expression).

In view of global practice, the increasing demand for new state-of-the-art electronic services and the endeavour for Lithuanian to have a full life in electronic space, researchers at the Institute of the Lithuanian Language joined forces with developers of advanced information technology and in 2018 began working to develop the information system of resources for Lithuanian within the framework of a new project funded by European Union structural funds. By the end of 2020, the project is expected to achieve the following goals: the integration of another three representative monolingual Lithuanian language resources that are currently only available on separate websites into the information system of resources for Lithuanian, which is undergoing an overhaul (http://lkiis.lki.lt/ → https://ekalba.lt/public#/home/main), namely digitizing the great *Dictionary of the Lithuanian Language* (20 volumes) as well as adding the digital *Dictionary of the Standard Lithuanian Language* with a list of titles and the digital *Database of Lithuanian Neologisms*; designing a semantic WordNet of Lithuanian and adding some more new e-services for users.

The Lithuanian language resources available online and the new possibilities that are currently under development are highly relevant, handy and useful to everyone who uses or wants to learn Lithuanian as well as governmental bodies, the education system and business.

In addition, the researchers at the Institute of the Lithuanian Language are involved in various national and international projects, one of them being the *eTranslation TermBank* project (funded by the European Union). The aim of the project is to improve the quality and scope of machine translations for the European Commission.

Finally, we present the latest developments on the *TermBank of the Republic of Lithuania* (the data administrator and manager of the information system of the *Lithuanian TermBank* is the State Commission of the Lithuanian Language, see http://terminai.vlkk.lt/ as well as information about some other innovative digital resources, public services and language tools in the Lithuanian language industry, including various challenges.

1.   Introduction

Due to the small number of speakers, the Lithuanian government and other state institutions support several programmes to promote a range of linguistic research and dissemination. The following documents and programmes on language technology (LT) policies in Lithuania are still valid or in the process of being updated.[1] The "Guidelines for Lithuanian LT Development 2014-2020", issued by the State Commission of the Lithuanian Language, are currently being updated (to appear in 2020 on www.vlkk.lt/). Its main priorities are research and development for language technologies, machine translation, speech analysis, dialogue systems, automatic summarization, semantic technologies, advanced text analysis, a compilation of language resources, etc. 'The Lithuanian Information Society Development Programme 2014-2020' approved by the Government of Lithuania is promoting the Lithuanian culture and language through ICT by creating digital content based on Lithuanian written and spoken language interfaces and by developing digital products and services. Accordingly, a national programme "The Lithuanian Language for Information Society (2014-2020)" was approved to ensure funding

---

[1]   The development of Lithuanian language resources and technologies can be divided into three stages: from 2004 to 2012, from 2012 to 2015 and from 2016 to 2020. The period between 2012 and 2015 was labelled as the breakthrough (Utka et al. 2016; Utka/Dadurkevičius/Schnur 2019; cf. also Vaišnienė/Zabarskaitė 2012; Guidelines for Lithuanian LT development 2014, forthcoming). The breakthrough was achieved thanks to three major actions: the implementation of the national programme "The Lithuanian Language for Information Society", the preparation of political guidelines for the further development of language technologies for Lithuanian 2016 to 2020 and international collaboration between LT communities and infrastructures. However, since there is only a small market for language technologies in Lithuania, the private sector does not show much interest when it comes to developing LT (Utka/Dadurkevičius/Schnur 2019). During the period of just four years (from 2012 to 2015) quite a few important language resources and tools for Lithuanian were developed that significantly narrowed the gap between Lithuanian and well-resourced European languages (Utka et al. 2016; cf. also Kasparaitis/Skersys 2017; https://semantika.lt/ etc.).

from European Structural funds and the State. The following five projects using EU Structural Funds in Lithuania are worth mentioning: 1) "Services Controlled by the Lithuanian Speech" (LIEPA-2) (2017-2020). 2) "The Development of the Public Electronic Services of the Lithuanian Language Syntactic and Semantic Analysis Information System" (SEMANTICS 2) (2017-2020). 3) "Development of Information System of Integrated Lithuanian Language and Writing Resources" (RAŠTIJA 2)[2] (2018-2020). 4) "Improvement and Development of Machine Translation Systems and Localization Services" (2018-2021). 5) "Development of the Information System of Resources of the Lithuanian Language (E-language)" (E-LANGUAGE) (2018-2020). In addition, in 2015, the "Lithuanian Roadmap for Research Infrastructures" was updated with more than 20 initiatives of national significance. More than €17 million have been allocated for five language technology projects. These projects focus on developing various crucial LT services or resources including automatic speech transcription, digital public language services, machine translation and localization (for more information see Utka et al. 2016; Utka/Dadurkevičius/Schnur 2019; EC Digital Government Factsheet Lithuania 2019; ELRC Workshop Report for Lithuania 2019; AI 2019; Guidelines for Lithuanian LT development 2014; forthcoming, Rehm et al. 2020; etc.). For example, one of those projects focuses on the modernization of previously developed machine translation systems and is developing a machine translation system for Lithuanian, English, French and Russian to be used nationwide.[3] This project is being implemented by Vilnius University. Indeed, Lithuanian universities have long been a great source of research as well as tech talents and experts. The Institute of the Lithuanian Language is also playing a key role being one of the most important centres for research and dissemination of the Lithuanian language.

This contribution describes some significant innovations and challenges in the digital transformation of the Lithuanian language industry.

2.   Information system of Lithuanian language resources and e-services

Considering the expectations and needs of the information society, the Institute of the Lithuanian Language contributes to successful digital transformation and manages the development of online infrastructure (see http://lki.lt/skaitmeniniai-lietuviu-kalbos-istekliai/[4]). Most of the various resources are available on the

---

[2] See www.raštija.lt/, which includes information about previous and current (ongoing) projects.

[3] In Lithuania, there is still no centralized translation service in public administrations. Consequently, translation practices vary from institution to institution, ranging from decentralized outsourcing to in-house translation (Utka/Dadurkevičius/Schnur 2019; cf. also Tilde Translator 2017-2019).

[4] A number of different databases, corpora, dictionaries and other online resources are available there.

current information system of resources for Lithuanian on http://lkiis.lki.lt/. This platform was developed and launched in 2015 by the Institute of the Lithuanian Language in association with the Institute of Lithuanian Literature and Folklore, Vilnius University and the University of Educational Science of Lithuania as part of "Developing ICT Solutions and Content to Help Preserve the Lithuanian Language in the Public Space and Enabling the Use Thereof", a project in the programme called "The Lithuanian Language in Information Society". The website currently offers the following

– six digitized monolingual dictionaries: *The Dictionary of the Modern Lithuanian Language*, *The Dictionary of Synonyms*, *The Dictionary of Antonyms*, *The Dictionary of Phraseology*, *The Dictionary of Comparisons*, *The Systemic Dictionary of the Lithuanian Language*;
– ten digitized bilingual dictionaries: *The Lithuanian–Latvian Dictionary*, *The Latvian–Lithuanian Dictionary*, *The Lithuanian–German Dictionary*, *The German–Lithuanian Dictionary*, *The Lithuanian–Polish Dictionary*, *The Polish–Lithuanian Dictionary*, *The Lithuanian–English Dictionary*, *The English–Lithuanian Dictionary*, *The Latin–Lithuanian Dictionary*, *The Ancient Greek–Lithuanian Dictionary*;
– three digitized catalogues of *The Dictionary of the Lithuanian Language*: *the Main Catalogue*, *the Catalogue of New Additions*, *the Catalogue of Dialects*;
– three digitalized catalogues of folklore heritage: *the Catalogue of Riddles*, *the Catalogue of Post-war Guerrilla Songs*, *the Catalogue of Folk Beliefs*;
– four electronic databases of Lithuanian language resources: *The Database of Historic Toponyms*, *The Database of Surnames*, *The Geoinformation Database of Lithuanian Toponyms*,[5] *The Dialects Archive*;

---

[5] The online geoinformation database of Lithuanian place names is geared primarily towards scientific research and application purposes. It is the first such database in Lithuania to present information about linguistic units – place names – and geographical units – geographical objects. Every linguistic/geographical unit has its coordinates mapped and is accompanied by attributes pertaining to the place name and the geographical object that it denotes, namely variations of the place name in the living language, its stressed main form, accentuation paradigm and other cases stressed, a scientific explanation of its origins and formation, the time it was recorded in the living language, the time it was first mentioned in historical sources, its pronunciation in all cases, the population of towns and villages, parameters and pictures of geographical objects, their administrative affiliation, etc. On top of that, it is the first database of Lithuanian place names to cover both the existing names of towns, rivers and lakes and the extinct proper names of these features as well as land names that were recorded from the local populace in the interwar (1935-1940) and post-war periods. The development of this database makes all the data contained in it (the authentic place name forms, accentuation, clarifications of formation and origin, object descriptions etc.) easily accessible to Baltic linguists and geographers in Lithuania and elsewhere, allowing them to carry out deeper and more expansive research.

– this system also provides the Lithuanian font *Palemonas* and several tools for learning Lithuanian. For example, educational language games to encourage independent learning, to promote Lithuanian spelling and stress patterns and to enrich the stock of means of expression. Besides, it is an excellent way for parents and children to spend time together, broadening their horizons of cultural history and language, because this educational material is presented in the form of engaging stories. The games may also help the Lithuanian diaspora to remember Lithuanian and what it sounds like.

In view of global practice, the increasing demand for new state-of-the-art electronic services and the endeavour for Lithuanian to have a full life in electronic space, researchers at the Institute of the Lithuanian Language joined forces with developers of advanced information technology and in 2018 began working to develop the information system of resources for Lithuanian within the framework of the *"*Development of the Information System of Resources of the Lithuanian Language (E-language)", a project funded by European Union structural funds. The new online system (infrastructure) will be called E. KALBA (*E-language*). By the end of 2020, the online infrastructure *E-language* (http://lkiis.lki.lt/ → https://ekalba.lt/public#/home/main) will integrate another three representative monolingual Lithuanian language resources that are currently only available on separate websites into the information system of resources for Lithuanian, which is undergoing an overhaul: the digital *Dictionary of the Standard Lithuanian Language* with a list of titles, the digital *Database of Lithuanian Neologisms* and a digitalized version of the great *Dictionary of the Lithuanian Language* (20 volumes).

The Dictionary of the Standard Lithuanian Language is an electronic resource (also available on http://bkz.lki.lt/) for the general public. It is a normative dictionary that reflects the real-life vocabulary of the language and the correct use of common words. Optional variants are also provided. The resource is constantly updated (for more information, see Liutkevičienė 2015, 2016).

The *Database of Lithuanian Neologisms* (also available on http://naujazodziai.lki.lt) started to be developed in 2011 as part of the META-NORD project funded by the European Commission and is constantly updated (for more information see Miliūnaitė 2018). The goal is to have a database for storing new lexical items in Lithuanian, eventually revealing a systemic image of this lexis as well as presenting users of the language with hands-on information that is not otherwise available in other lexicographic sources. The new lexis – new coinages or neologisms – makes its way into the language through borrowings (translation) from other languages, creating new words based on the language's own formational principles or by giving new meanings to existing words. The database is expected to give a picture of the extent of new coinages in Lithuanian over the past few decades; once more data are available, it should eventually capture the dynamics

of new words including when, and in what sources, new words appeared, how much they spread in usage, and in what way, and the new models of word formation that become evident. At present the database contains around 5,500 new words from different domains of modern Lithuanian, primarily printed and web media, administrative language and fiction. The database gives the definition of a coinage, indicates the area where it is used, shows its semantic and formational ties with other new words in the database, illustrates its usage with examples, and provides normative comments for language users if necessary.

Just like many other digital products that exist on the internet and reflect modern processes, the new coinages database is available to users as a non-finite, ever evolving and expanding compilation of data that can be improved by anyone willing to send their feedback and suggestions.

Digitizing the great *Dictionary of the Lithuanian Language* (20 volumes, also available on www.lkz.lt) is considered to be the most significant work of 20th century Lithuanian lexicography and Lithuanian philology in general. This dictionary describes the long process of compiling this enormous undertaking and lists the fieldworkers, compilers and editors, etc. who contributed to the lexicographical recording of half a million Lithuanian words. The *Dictionary of the Lithuanian Language* (in 20 volumes) and its digitized version, the accumulation of a digitized database of the Lithuanian lexicon, is very important not only for academics but also for the general public.

The online infrastructure *E-language* will provide a semantic WordNet of Lithuanian and some more new e-services for users. The Lithuanian WordNet and the following e-services are still in production:

– *Search across the semantic net* – to retrieve information on the lexis in the Lithuanian WordNet,
– *E-terms* – to obtain all of the information about a particular medical, financial or IT term available in the resources,
– *E-marketing* (consisting of the following e-services: *E-name*, *Analysis of user queries*, *Sentiment analysis*) – to use Lithuanian language resources for marketing purposes,
– *E-tips* (consisting of the following e-services: *Word formation guide*, *Language tips*) – to boost the productivity of the use of the possibilities that Lithuanian has to offer, create new words, obtain detailed information about the norms of the standard language and their variations and changes.

Thus the information system of digital linguistic resources (http://lkiis.lki.lt/ → https://ekalba.lt/public#/home/main) is undergoing further developments to embrace more levels of language; the system will also integrate more resources to supplement the set of data from each level. Any user (from an academic to a member of the general public) will be able to retrieve linguistic information and

its contextual data of their choice. The development of the Lithuanian WordNet[6] will be based on internationally recognized models and practices of WordNet data (such as the Princeton WordNet data model), and the original version of the WordNet will include the set of common base concepts as recommended by the Global WordNet Association, which was defined in the EuroWordNet and BalkaNet projects. Furthermore, the original version of the WordNet will be developed by integrating available lexicographical resources (the *Dictionary of the Modern Lithuanian Language*, the *Dictionary of Synonyms*, the *Dictionary of Antonyms*, etc.).

The Lithuanian language resources available online[7] and the new possibilities that are currently under development are highly relevant, handy and useful to everyone who uses or wants to learn Lithuanian, as well as governmental bodies, the education system and business.

3.   Terminological resources in Lithuania

The development of several terminological sources in Lithuania is presented below. First of all, the *TermBank of the Republic of Lithuania* is described, since this is the most reliable and universal source of Lithuanian terminology available on the internet (http://terminai.vlkk.lt/).

In 2003 the Language Commission, together with the Office of the Parliament of the Republic of Lithuania, initiated the creation of a national term bank. At the end of the same year, parliament passed the *Law of the Term Bank of the Republic of Lithuania* (2003). It regulates the creation, management, use and financing of the term bank as well as the duties, rights and responsibilities of individuals and institutions providing, managing and using data in this term bank. The structure, information units, rules of creation, ordering and usage, technical data, requirements and functions of the information system in the term bank are defined in the *Methodology of the Term Bank of the Republic of Lithuania*, stating as follows:

> The purpose of the Term Bank is to ensure consistent usage of the approved terms of the Lithuanian language, particularly those used in legal acts of the Republic of Lithuania, to create a common information system for various public authorities, which could be accessed by and receive data from other individuals and organizations and which could be used freely by specialists in different fields not only in Lithuania, but also in other countries. (see LTBM)

---

[6]   There is also another Lithuanian WordNet but it differs in the number and variety of linguistic resources (Vitkutė-Adžgauskienė/Dainauskas et al. 2016; Vitkutė-Adžgauskienė/Utka et al. 2016, etc.).

[7]   For more detailed information, see the external links in the references at the end of this article.

The owner and manager of the Term Bank is the State Commission of the Lithuanian Language (data administrator and manager). The Chancellery of the Parliament of the Republic of Lithuania was responsible for information system administration until the end of 2018 and at the beginning of 2019 the State Commission took over this duty.

Ministries and their subordinate departments and services, government authorities, the Parliament of the Republic of Lithuania and other state institutions are authorized to store and manage terms in their fields in the Term Bank. Terminology Commissions have been established in different state institutions, initially responsible for making an inventory of terms of legislation in force, editing term entries and presenting them to the Term Bank; subsequently, the focus has moved to the coordination of draft legislative terms with the Language Commission and entering them in the Term Bank. The terms deliberated by the Terminology Sub-Commission and approved by the Language Commission are added to the Bank with the indication that they have been *approved*. The Law on the Term Bank of the Republic of Lithuania has increased the management and dissemination of the terms of laws and other legal acts (for more, see Umbrasas 2013).

In addition to legislative terms, dictionaries that receive a positive assessment from the Language Commission are entered in the Term Bank. Data from 46 terminological dictionaries are now available in the Term Bank. There are over 245,700 term entries in the bank, with around 20,000 entries added each year, 4-5% of them from legal acts. Currently, the Information System of the Term Bank is being updated, not only the window and interfaces visible to the user but also the internal Term Bank information system.

The Term Bank of the Republic of Lithuania is not the only terminological resource containing Lithuanian terms. Another four resources should also be mentioned:

– *The Term Base of the Lithuanian Standards Board* – since 2000, the Lithuanian Standards Board has maintained a database of standardized terms. Currently, it contains around 70,000 term entries from standards. This database is freely available for users on the website of the Lithuanian Standards Board (www.lsd.lt/index.php?-452282422).
– *Raštija.lt* – an integrated information system for Lithuanian language and writing resources with data from 20 terminological dictionaries (informatics, computing, medicine, chemistry, physics, etc., www.raštija.lt).
– *EuroTermBank* – 15 resources with Lithuanian terminology are available from the free centralized online terminology database *EuroTermBank* (www.eurotermbank.com).
– *IATE* (*Inter-Active Terminology for Europe*) – an inter-institutional European Union term database containing over 57,000 Lithuanian terms (http://iate.europa.eu/). (For more information, see Auksoriūtė 2016).

4. Challenges vs opportunities for successful digital transformation in the Lithuanian language industry

In Europe, the main challenges for the exploitation of language data for the purpose of developing language technology applications and creating platforms (infrastructures) are mainly related to the amount of language data needed, language data protection, copyright, the costs of preparation and annotation, privacy, anonymization, the interoperability of data infrastructures, metrics of quality and workflows, etc. Any solution determines the quality of the transformation of the language industry; higher quality data and good solutions always yield better results. Some main challenges (and needs) that have been met in recent years in the Lithuanian language industry are listed below: there is only a small market for language technologies in Lithuania;

–   a cloud problem still exists in Lithuania to store resources safely (the authorities are still working on this);
–   limited open data accessibility and use, and a lack of language resources;
–   a lack of skilled programmers, a lack of dedicated BA, MA and PhD programmes for language technologies;
–   rapidly aging technologies;
–   the cost of data preparation;
–   a lack of efficient centralized e-services, fragmentation;
–   and others. (For more challenges and needs as well as opportunities, see ELRC White Paper 2019; Telksnys 2017; Guidelines for Lithuanian LT development 2014; 2020; AI 2019; DEP LT Report 2020; Lietuva 2030; etc.).

To ensure the sustainable development of AI and to enable Lithuania to be competitive in the global market, a preliminary *Lithuanian Artificial Intelligence Strategy: A Vision of the Future* (AI 2019) was published in 2019. The strategy provides recommendations, with the goal described as being "To modernize and expand the current AI ecosystem in Lithuania and ensure that the nation is ready for a future with AI". The strategy includes the following sections: ethical and legal core principles for the development and use of AI, a breakdown of Lithuania's position in the AI ecosystem, the integration of artificial intelligence systems across all economic sectors, the national development of skills and competencies needed for a future with artificial intelligence, the growth of artificial intelligence research and development and a responsible and efficient approach to data.

5. Conclusion

As already mentioned above, the Lithuanian language resources available on the internet and the new opportunities that are currently being developed are highly relevant, handy and useful for everyone who uses or wants to learn Lithuanian, as

well as governmental institutions, the education system and business. Since the digital economy is vital for innovation, growth, studies, jobs and competitiveness, it is important to support national language research institutions and universities that are playing a key role in developing language technologies, such as online infrastructures (platforms) for digital and digitized linguistic resources, learning and training tools, innovative digital services (including a centralized translation service), semantic WordNets, various databases and applications, because it is not an easy task to create, combine, prepare and develop various kind of linguistic data, resources and research in combination with semantic WordNets, terminology and speech databases, etc. In the near future, access to all kinds of linguistic information would be extremely useful for an artificial intelligence program. In Lithuania, properly designed and developed language resources and linguistic data in line with international standards will bridge the gap between currently separate but related disciplines (Guidelines for Lithuanian LT development 2020). Considering all the challenges and needs described in the previous sections of this paper, it is vital to facilitate joint efforts on language resources and data creation, further developing online infrastructures (platforms), innovative e-services and applications. The magnitude of the digital skills challenge requires a long-term strategy and new partnerships between national, regional and international as well as private and public players, also including society, or the general public, as a whole.

External links and references

AI 2019: *Lithuanian artificial intelligence strategy: A vision of the future.* Cambridge, MA: Future of Life Institute. https://futureoflife.org/national-international-ai-strategies/.

Ar Lietuva liks kalbos technologijų paraštėse? 2018-10-04. http://alkas.lt/2018/10/04/spaudos-konferencija-ar-lietuva-liks-kalbos-technologiju-parastese-tiesiogine-transli-acija/.

Auksoriūtė, A. (2016): Current state of terminology in Lithuania: Scientific research, management and education. In: *Acta Baltico-Slavica* 40. https://ispan.waw.pl/journals/index.php/abs/article/view/abs.2016.007.

DEP LT Report (2020): DEP LT Stakeholder Consultation workshop report. https://ec.europa.eu/digital-single-market/en/news/language-technologies-digital-europe-programme-stakeholder-consultation.

ELRC White Paper (2019*): ELRC white paper. Sustainable language data sharing to support language equality in multilingual Europe: Why language data matters*. Second online edition. Saarbrücken: German Research Center for Artificial Intelligence (DFKI)/ European Language Resource Coordination, Connecting Europe Facility. www.lr-coordination.eu/sites/default/files/ELRC_Conference/ELRCWhitePaper.pdf.

ELRC Workshop Report for Lithuania (2019). European Language Resource Coordination, Connecting Europe Facility. www.lr-coordination.eu/sites/default/files/Lithuania/2019/ELRC%2B%20Workshop%20Report_Public_Lithuania.pdf.

EC Digital Government Factsheet Lithuania (2019): *European Commission: Digital Government Factsheet 2019 Lithuania*. https://joinup.ec.europa.eu/sites/default/files/inline-files/Digital_Government_Factsheets_Lithuania_2019_0.pdf.

Guidelines for Lithuanian LT development (2014): *Lietuvių kalbos plėtros informacinėse technologijose 2014–2020 m. gairės*. www.vlkk.lt/kalbos-politika/lietuviu-kalbos-pletros-informacinese-technologijose-gaires.

Guidelines for Lithuanian LT development (forthcoming): *Lietuvių kalbos plėtros skaitmeninėje terpėje ir kalbos technologijų pažangos 2021-2027 m. gairės*. [To appear on www.vlkk.lt/].

Kasparaitis, P./Skersys, G. (2017): Lietuviško balso sintezės dabartis ir perspektyvos. In: *Bendrinė kalba* 90. http://journals.lki.lt/bendrinekalba/article/view/160.

Lietuva 2030: *The strategy for Lithuania's advancement "Lietuva 2030"*. www.lietuva2030.lt/lt/apie-lietuva-2030.

Lietuvių kalba internete – vis daugiau žodynų ir elektroninių paslaugų. In: *tv3.lt*, 2018-10-29. www.tv3.lt/naujiena/lietuva/979281/lietuviu-kalba-internete-vis-daugiau-zodynu-ir-elektroniniu-paslaugu.

Lietuvių kalbs keliauja į internetą ir skaitmeninę erdvę. In: *15min.lt*, 2018-09-28. www.15min.lt/verslas/naujiena/galimybes/lietuviu-kalba-keliauja-i-interneta-ir-skaitmenine-erdve-991-1035100.

Lietuvių kalba netrukus taps lengvesnė: internete kuria patogią sistemą. In: *DELFI*, 2018-08-24. www.delfi.lt/projektai/suprasti-interneta/lietuviu-kalba-netrukus-taps-lengvesne-internete-kuria-patogia-sistema.d?id=78879667.

Liutkevičienė, D. (2015): *Bendrinės kalbos normų pateikimas* "*Bendrinės lietuvių kalbos žodyne*" (The dictionary of the standard Lithuanian language and the norms of the standard language). In: *Bendrinė kalba* 88. http://journals.lki.lt/bendrinekalba/article/view/182.

Liutkevičienė, D. (2016): *Antraštinių žodžių reikšmių pateikimas* "*Bendrinės lietuvių kalbos žodyne*" (Presenting meanings of the entry words in the dictionary of the standard Lithuanian language). In: *Bendrinė kalba* 89. http://journals.lki.lt/bendrinekalba/article/view/169.

LTBM: *Lietuvos Respublikos terminų banko metodika*. www.vlkk.lt/vlkk-nutarimai/nutarimai/del-lietuvos-respublikos-terminu-banko-metodikos-patvirtinimo.

Miliūnaitė, R. (2018): Naujažodžių pateikimas ir paieškos galimybės Lietuvių kalbos naujažodžių duomenyne (Neologisms presentation and search possibilities in the database of Lithuanian neologisms). In: *Bendrinė kalba* 91. www.bendrinekalba.lt/Straipsniai/91/Miliunaite_BK_91_straipsnis.pdf, http://journals.lki.lt/bendrinekalba/issue/view/5.

Rehm, G./Marheinecke, K./Hegele, S./Piperidis, S./Bontcheva, K./Hajic, J./Choukri, K./ Vasiļjevs, A./Backfried, G./Prinz, C./Gómez Pérez, J.M./Meertens, L./Lukowicz, P./ van Genabith, J./Lösch, A./Slusallek, P./Irgens, M./Gatellier, P./Köhler, J./Le Bars, L./ Auksoriūtė, A. et al. (2020): The European language technology landscape in 2020: Language-centric and human-centric AI for cross-cultural communication in multilingual Europe. In: *Proceedings of the 12th Language Resources and Evaluation Conference* (*LREC 2020*). https://arxiv.org/ftp/arxiv/papers/2003/2003.13833.pdf.

Telksnys, L. (2017): Rašytinė ir sakytinė lietuvių kalba įprastoje ir elektroninėje terpėse. In: *Bendrinė kalba* 90. http://journals.lki.lt/bendrinekalba/article/view/161.

Tilde Translator (2017-2019): "*Tilde*" *vertyklė*. https://translate.tilde.com/en.

Umbrasas, A. (2013): Lietuvos Respublikos terminų bankas: 10 metų po įstatymo priėmimo (Term Bank of Republic of Lithuania: 10 years after passing the law). In: *Terminologija* 20. http://journals.lki.lt/terminologija/article/view/475.

Utka, A./Amilevičius, D./Krilavičius, T./Vitkutė-Adžgauskienė, D. (2016): Overview of the development of language resources and technologies in Lithuania (2012-2015). In: *Human Language Technologies – The Baltic Perspective* I, 12-19. http://ebooks. iospress.nl/volumearticle/45508.

Utka, A./Dadurkevičius, V./Schnur, V. (2019): Country profile Lithuania. In: *ELRC White Paper. Sustainable language data sharing to support language equality in multilingual Europe: Why language data matters*. Second online edition. Saarbrücken: German Research Center for Artificial Intelligence (DFKI)/European Language Resource Coordination, Connecting Europe Facility, 108-112. www.lr-coordination.eu/sites/ default/files/ELRC_Conference/ELRCWhitePaper.pdf.

Vaišnienė, D./Zabarskaitė, J. (2012): *The Lithuanian language in the digital age*/*Lietuvių kalba skaitmeniniame amžiuje*. (= META-NET White Paper Series). Heidelberg: Springer.

Vitkutė-Adžgauskienė, D./Dainauskas, J./Amilevičius D./Utka, A. (2016): Lietuvių kalbos žodžių tinklas (Lithuanian Word Net). In: *Darbai ir dienos* 64, 101-114.

Vitkutė-Adžgauskienė, D./Utka, A./Amilevičius, D./Krilavičius, T. (2016): NLP Infrastructure for the Lithuanian Language. In: *Proceedings of the 10th Language Resources and Evaluation Conference* (*LREC 2016*). Paris: European Language Resources Association, 2539–2542.