European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2015/1091 programme.

European Language Resource Coordination
Connecting Europe Facility

Deliverable D3.2.18
Task 3

ELRC Workshop Report for Cyprus

Author(s): Natassa Avraamides-Haratsi
Dora Loizidou

Dissemination Level: Public

Version No.: <V1.0>

Date: 2022-01-04

Contents

1 Executive Summary

The present document is a report regarding the 3rd ERLC Workshop which was held virtually on 1st December 2021. It includes the final agenda of the event and brief summaries of the sessions which included speeches, presentations and discussion. The event with the title "Reshaping multilingual Europe: Language-centric Artificial Intelligence" was attended by researchers, academics, scientists, representatives of public services and of language technology industry, translators and students.

The workshop's highlight was to address questions about Automated Translation as a type of technology transforming the way we interact with our administration, the way we shop, work and travel but also systems that automatically aid sophisticated writing, digital assistants that transform our voices to text messages on mobile phones, bots that answer our calls to the bank or to our social security organization. The workshop sought to engage participants in a fruitful discussion on the status and prospects of Language Technology for Greek as well as the challenge of the digitization of the Greek Cypriot language. Researchers, developers, integrators and users of Language Technology, both from the private and public sector shared experiences, requirements and ways for transforming digital interaction in our multilingual Europe with Language Technologies. Last but not least, it was discussed how language data, i.e. texts and speech, can fuel development in Artificial Intelligence. The contribution from the official from the European Commission was informative and optimistic regarding developments in language digitization within the EU. The overview given by a senior representative from the Cypriot public administration was promising regarding the future of digitization and the importance of including language technologies in the digitization system of the public administration. The workshop's highlight was to underline the importance of digitizing the Cypriot Greek in order to achieve its survival in the age of digitization. Thus, the importance of collecting language resources in Cyprus was taken one step ahead with regard to the strengthening of the linguistic resources in the Greek language.

One of the main outcomes of the event is that the more language resources in Greek are collected the better the digitization of the Greek language will be. The Press and Information Office of the Republic of Cyprus, will continue to act as a national coordinator for language data collection in Cyprus in cooperation with the University of Cyprus and its network of public sector representatives from the national administration bodies.

The event was hosted in a dedicated webpage, https://lr-coordination.eu/cyprus3rd, where the presentations are also available.

2 Workshop Agenda

09:15 – 09:30	**Welcome and introduction**
09:30 – 10:00	**European Language Resource Coordination - Language Technologies and data for Greek** Stelios Piperidis, Institute for Language and Speech Processing / Athena R.C., ELRC
10:00 – 10:30	**The Automated Translation Platform** (Connecting Europe Facility) Anastasios Angelides, Head of Greek Translation Unit, Directorate-General for Translation (DGT)
10:30 – 11:00	**The potential of Language Technology and AI** Alberto Calzada, Lecturer in AI, European University of Cyprus
11:00 – 11:15	**Using Natural Language Processing and Machine Learning to Assess Language Disorders** Charalambos Themistocleous, Postdoctoral researcher in Computational Linguistics, Johns Hopkins University
11:15 – 11:30	*Coffee Break*
11:30 – 12:00	**Language Technologies by/for the public sector** Andriana Achilleos, Director of Department of Information Technology Services, Deputy Ministry of Research, Innovation and Digital Strategy Natassa Avraamides, Press and Information Office, Ministry of Interior, NAP Public Services
12:00 – 12:30	**Development of language data for the Cypriot Dialect: challenges and prospects** Spiros Armostis, Lecturer in Linguistics, University of Cyprus
12:30 – 13:15	**Discussion round** *Moderator: Dora Loizidou, University of Cyprus, National Anchor Point for Technology ELRC* Anastasios Angelides, Head of Greek Translation Unit, Directorate-General for Translation (DGT) Alberto Calzada, Lecturer in AI, European University Spiros Armostis, Lecturer of Linguistics, University of Cyprus Themos Stafylakis, Omilia
13:15 - 13:30	**Conclusions**

3 Summary of Content of Sessions

3.1 Welcome and introduction

Natassa Avraamides, ELRC NAP for Public Services in Cyprus, welcomed the participants and underlined the importance of the work done by the ELRC. She referred to the journey so far regarding the participation of the Public Services in the language data collection and its importance in supporting the digitization of the Greek language. Special reference was made to the contribution of the Press and Information Office (PIO) in collecting monolingual, bilingual and multilingual language data, such as press releases and publications of the PIO and in sharing them with the EC and the European LT community through the ELRC-SHARE repository. She mentioned the work done in the collection of language data by all countries represented in the ELRC network so far and the efforts made in creating a common understanding of the modus operandi with respect to language data collection and sharing, as these were described in the ELRC White Paper published in 2019. Natassa expressed the hope that the collection of big and quality language data in smaller languages, and in particular in Greek and Cypriot Greek, will be enable them to survive in the digital age.
Dora Loizidou, ELRC Technological NAP for Cyprus, presented briefly the previous ELRC workshops held in Cyprus in 2015 and 2018, as well as the new EU funding programme "Digital Europe Programme" (DIGITAL) which is focused on bringing digital technology to businesses, citizens and public administrations. Workshop presentations began after a short introduction to the workshop practicalities and agenda.

Before the sessions, the participants were asked to answer one poll question to identify the area in which they operate. The answers are reported in the following table:

Q1: Which of the following best describes the area in which you operate? (you can select up to 2)

	#answers	%
Public sector	1	1.7%
Research community/University	27	45%
Creator/Provider of language technology/artificial intelligence systems	5	8.3%
Translation services	12	20%
Small and Medium-sized Enterprise/Freelancer	8	13.3%
Other	7	11.7%

3.2 European Language Resource Coordination - Language Technologies and data for Greek

Stelios Piperides, senior researcher at the Institute for Language and Speech Processing / Athena R.C., and representative of the ELRC consortium, gave a brief overview on the development of translation technology and stressed the importance of contributing language data to the ELRC and to eTranslation. He also emphasised the importance of multilingualism in the EU and underlined the need to prevent digital extinction of smaller languages, such as Greek. Furthermore, Mr. Piperides presented some examples of AI applications used in our everyday lives (digital personal assistants, chatbots, smart cars) and explained how these are powered in the background by language technologies. Arguing that language-centric AI is a special case in the broad AI field, he underlined that language data are a special case of data, because of language's inherent polysemy, ambiguity, context-dependency etc. Mr. Piperides underlined the importance of linguistic data as the "fuel" of language

technologies, saying that the volume and quality of data as well as the variety of data are crucial for the further development of language technologies and Artificial Intelligence. Language data is also the "fuel" for economic growth towards the EU vision to "Unleash the potential of data with European common data spaces built on innovative secure and energy efficient cloud to edge technology" (Digital Europe Work Programme 2021-2027)". Within this framework the role of ELRC has been of vital importance as part of the Connecting Europe Facility (CEF), ELRC was able to promote eTranslation, to raise awareness of the value of language data, to collect and compile language resources which are provided through the ELRC-SHARE repository. Stelios also referred to the vital topic of "big versus small languages" stressing that the native languages of 140 million citizens are in the Language Technology Danger Zone, as only very few languages meet the required level of language technology support which can ensure their digital preservation. Mr. Piperides also made a reference to the European Language Equality project which aims to create a strategic agenda and a road map for achieving the full digital equality of all languages in the EU until 2030.

Q: Do you think that Machine Translation will ever be able to replace human translation?

A: Mr. Piperides highlighted that, as language evolves through time, machine translation systems need to be retrained and adapted. The COVID pandemic is an indicative example: if eTranslation had not been retrained with language data containing the new coronavirus vocabulary, it would have been impossible for us to be able to translate Covid-19 texts that emerged in 2020. Technology must strive to catch up with the evolution of language. In terms of creativity, he does not believe that machines will be able to replace human beings.

At the end of this first session, the participants were asked to answer three poll questions to investigate their experiences as users of language technologies. The results are presented in the following tables:

Q1: Which of the following language technologies do you use more? (you can select up to 3)

	#answers	%
Automated translation	29	25.7%
Speech recognition (e.g. dictation of messages on your mobile phone)	4	3.5%
Text to Speech synthesis (e.g. your mobile phone reads out your messages)	3	2.7%
Digital assistants (e.g. on mobile phones or call centers)	13	11.5%
Information search and retrieval technologies (e.g. web search)	31	27.4%
Spelling and grammar corrector	33	29.2%

Q2: If you use some of these technologies, in which language do you usually use them? (single choice)

	#answers	%
Greek	6	13%
English	28	60.9%
French	9	19.6%
Other language	3	6.5%

Q3: As a simple user, how satisfied are you with the quality and reliability of these technologies for Greek?

	#answers	%
Excellent	0	0%
Good	17	36.2%

Fair	21	44.7%
Poor	7	14.9%
Very Poor	1	2.1%
I don't know/No answer	1	2.1%

3.3 The Automated Translation Platform (Connecting Europe Facility)

Anastasios Angelides talked about the EU multilingualism policy as provided in Regulation 1/58 of the EU and Article 20 of the EU Convention. He referred to the role of translation within the bodies of the EU. Mr. Angelides gave a picture of the road to Automatic Translation from its very beginnings (SYSTRAN) and explained the advantages and disadvantages of the first translation systems leading to the creation of Machine Translation (Original MT@EC, based on legislative EU style language), the current system of eTranslation, launched in 2017, is a state-of-the-art neural machine translation system supported by CEF. Public services and their websites can connect to the eTranslation service. Mr. Angelides moved from the statistical MT paradigm to neural machine translation. He also underlined the benefits of NMT, i.e. excellent accuracy and low cost without the use of a pivot language (usually English) and also the drawback for less-resourced languages. Mr. Angelides referred to the usefulness of Cypriot Greek in enriching and diversifying the standard Greek terminology giving examples of words that were introduced as Cypriot Greek and are now being used in the standard Greek terminology. DGT supports the work done in the field of language technology and multilingualism and also works closely with the ELRC in order to provide the necessary resources to advance language technologies.

Q: eTranslation is now available to European SMEs. Do you think it would be useful to also make it available to academic organisations?

A: Angelides strongly believes in universities and the work they produce saying that they are even more organized. He added that providing access to academic institutions is possible.

At the end of this session, the audience was presented with three poll questions on their experiences with machine translation systems. The following tables summarise their answers:

Q1: Have you ever used an automated translation system?

	#answers	**%**
Yes	35	94.6%
No	2	5.4%

Q2: If yes, which one? (single choice)

	#answers	
eTranslation of the European Commission	4	10.8%
Other commercial or freely available system (e.g. Google translate)	25	67.6%
All of the above	7	18.9%
I don't know/No answer	1	2.7%

Q3: Based on your experience with the above translation systems, how satisfied are you with the translation quality to/from Greek?

	#answers	**%**

Excellent	0	0%
Good	8	28.6%
Fair	14	50%
Poor	4	14.3%
Very Poor	1	3.6%
I don't know/No answer	1	3.6%

3.4 The potential of Language Technology and AI

The keynote speaker Alberto Calzada, Lecturer in AI at the European University of Cyprus, started his presentation by providing the definitions of AI and Machine Learning (ML). As explained, AI can be grossly described as an algorithm that solves a task and illustrated this by giving the example of digital assistants, like Siri and Alexa. He clarified the link between AI and ML explaining that ML is a branch of AI which focuses on providing machines with learning capabilities. Learning is achieved with supervised, unsupervised or reinforced learning methods. He then referred to specialised areas of knowledge and applications that exploit AI and ML, such as Natural Language Processing (NLP), Voice Recognition, Robotics and Computer Vision.

Secondly, Mr. Calzada explained in a simplified manner various NLP tasks; he classified them in three levels: low-level, mid-level and high-level tasks. He described that the low-level tasks are simpler tasks such as spell-checking, basic information retrieval or next-word suggestion. He then referred to mid-level tasks considered as of increased complexity, for instance topic detection, document classification, extraction of information about named entities and creation of embeddings. As regards high-level tasks, he described that we deal with complex problems, such as automated text summarisation, machine translation and chatbots. He outlined that these tasks are starting to be used in industry and specified that low and mid-level tasks are in many cases the base for more complex tasks. He added that therefore, NLP can be regarded as a collection of "building blocks" which we can combine to create more complex systems.

Thirdly, Mr. Calzada made a historical overview of AI in order to explain its evolution since the 1950s. AI has evolved through four distinct eras: 1) procedural era in the 1950s-1960s, 2) statistical era in the 1970s-1990s with language models, 3) ML era since the 2000s till nowadays with document classification (for example spam and no spam) and word embeddings, and 4) deep learning era that just began in 2020s with transformers. He clarified that transformers are used to solve many tasks, such as machine translation, automatic text summarisation, text generation and named-entity recognition.

A number of questions were addressed to Alberto Calzada, followed by discussion:

Q: The systems are programmed to reply to tasks. To what extent is it easy to distinguish language varieties, for example to confuse/combine Cypriot Greek with Standard Modern Greek?

A: Mr. Calzada replied that language recognition has been achieved in a certain extent, but a lot of languages have common words and this is very complicated, for example Google does not always correctly distinguish Spanish from Catalan or Portuguese. He emphasised that the origin of the problem is lack of data. He explained that when Cypriot Greek is standardised, taught in school and written in the newspapers etc., the models will automatically imitate, but since now there are not enough Cypriot Greek data, such systems cannot be trained to distinguish the Standard Modern Greek variety form Cypriot Greek.

Q: Comment on DAR (Digital Audio Recording). This will be now used in various cases / court hearings in order to record and store them electronically. The problem is that, if a Cypriot speaks in Cypriot

dialect, how will it be translated? Not only for Cypriots but also in other countries with different dialects, there is this difficulty. How will artificial intelligence deal with such problems?

A: Mr. Calzada agreed and added that the same problem appears in other languages too. He explained that the problem is not only due to lack of data and resources, but a greater problem of historical stigma for certain dialects, for example in Italian or in Spanish where if someone uses a dialect is considered as an old person, a villager or with no education.

A: Stelios Piperides added that the issue raised by the previous speakers was important. He said that they do not want Cypriot Greek to be the "neglected" child. This is the reason why in collaboration with the European Language Equality program, they have tried to record language resources for Cypriot Greek. Technology can help with the issue of court hearings raised by an attendee. There are two options: either to develop a system specifically trained with Cypriot Greek - English data or to adapt an existing Standard Greek-English system to Cypriot Greek, given the availability of data. He concludes that automatic translation systems can hardly integrate external knowledge but a great deal of research is being done on it.

Q: (chat): Transformers (and their variations) have been the de-facto architecture for quite a few years. Are you aware of any new emerging architectures?

A: Stelios Piperides replied following his answer above. He said that there are emerging architectures but not de-facto industrially used. The main issue is how to perform the convolution of transformer models with what we call factual / external / declarative knowledge coming from external sources (i.e., Wikipedia, wikidata and so forth). There are lots of things that are hidden in the data, but the knowledge has been extracted and represented in a declarative nature.

A: Alberto Calzada added that the speech to text recognition does not do a good job in recognising intonation, for instance irony or angriness. So, if someone is saying something negative, the way he says it changes the message.

A: Stelios Piperides agreed and added that an additional perplexity is due to the language varieties, beyond the standard and dialects. For instance, LT needs to be able to process all the varieties of English as spoken by non-native speakers around the world. In an experiment during the META-FORUM, the ELITR system for real-time transcription of conference presentations was tested. When the speaker changed from native to non-native, the degradation of the performance of speech recognition was remarkable. He adds that we will need more data to try to either train or adapt systems with non-native speaker's data.

A: Alberto Calzada replied that there are corpora which are being developed (Wikipedia, there is the "Simple English") for non-native speakers which might be done for other languages too.

A (chat): The Cypriot dialect is multifaceted and complex. To achieve all this will require the contribution of several linguists and more.

A (chat): Most of these tasks can be accomplished today with domain adaptation. For instance, a universal Greek model can be trained in which the Cypriot text can be marked and then during inference we can instruct the system to use the Cypriot dialect. We just need some amount of Cypriot text.

3.5 Using Natural Language Processing and Machine Learning to Assess Language Disorders

In the next session, Charalambos Themistocleous[1], a postdoctoral researcher in Computational Linguistics at the Johns Hopkins University, presented Natural Language Processing (NLP) and Machine Learning (ML) in Language Disorders. Charalambos currently works with two groups of patients: 1) primary progressive aphasia (non-fluent, semantic and logogenic variant) and 2) mild cognitive impairment (MCI, incipient memory difficulty in remembering events and situations, problems in decision making, planning, and finding their way in familiar environments etc.). He pointed out that to this day, there is no cure for dementia, but early-stage treatment can delay the progression of MCI and thus, the development of reliable tools for identifying early cognitive changes is of great importance.

Mr. Themistocleous explained that diagnosis, prognosis, and evaluation of patients' condition requires substantial effort and manual analysis of speech data. It is time-consuming and it requires substantial expertise, for instance speech transcription, annotation of the linguistic characteristic, measurements and scoring. He also referred to data elicitation and types of communication, such as free style conversation, map tasks and picture description tasks. He clarified that perceptual identification of speech characteristics (just by listening) is subjective, and varies on how clinicians process, interpret, and judge acoustic, grammatical, etc. properties of speech and language.

Mr. Themistocleous presented his research aims: 1) provide easy, quick, and automated diagnosis, prognosis, and ultimately improve therapy decisions using NLP and ML, 2) assist medical providers by identifying subgroups of patients by subtyping patients into variants, 3) evaluate the effectiveness of treatment methods by quantifying their effects on speech and language and 4) augment current treatment and evaluation batteries for speech and language pathology using NLP and ML. He then presented tools that they developed in order to identify patients with MCI through discourse analysis (transcription and segmentation, audio processing, morphosyntactic analysis). He also mentioned the use of tools for scoring spelling, phonological and semantic errors. Finally, he referred to a research done on linguistic varieties in order to identify a speaker of Cypriot Greek and Athenian Greek from a single vowel or consonant.

Q: In what languages can these tools be used and do you know any relevant research for Greek?

A: Charalambos Themistocleous replied by underlining that the tools they make belong to a system which consists of various components. The goal is to get a one-minute recording and be able to enter information about a patient's condition. In the recording, they can get information from the audio but also from the audio's transcription in written text. The complete system is in English. Some of the components are in Swedish and others in Greek. However, with some adaptation they can be developed for Greek too. Research done on people with mild diagnostic disorder or progressive aphasia is in English. Now, for Greek there are far fewer studies on mild diagnostic disorder and progressive aphasia. He hopes through the cooperation of Cyprus's, Greece's and his own university to develop this space with the analysis of natural language and machine learning.

Q: According to your experience, how reliable are these tools for different language varieties?

A: It is an ongoing research question and analysis that has already recognised the importance of dialects in America. The tools he prepares are tools that can take the information, analyse and understand it through audio, text and dialect speakers. In terms of reliability, he cannot surely say that it is stable due to the fact that the tools are based on machine learning. It depends of course on the data the tools have been trained with. If they have an adequate body of speakers of a particular

[1] Due to different time zone, Charalambos Themistocleous speech was recorded and we played the video during the online workshop.

language, then they can increase both the accuracy and reliability of the tools and the analysis. To some extent it is the volume of data and to another degree it is the configuration and type of tools.

A (chat): Example of relevant research for the diagnosis of language disorders using language technology for Greek: https://www.ilsp.gr/projects/plan-v-el

Q: The next question concerns the research for the distinction of linguistic varieties of Cypriot Greek with Athenian Greek. Can you tell us some of the results?

A: It includes a data collection of 45 speakers, through which he analysed various audio and production characteristics of speakers from Athens and Cyprus. For example, the trained algorithms can distinguish a speaker through a consonant or a vowel.

Q (chat): When the systems give a diagnosis how easily is this accepted by the doctors and/or patients? Do they ask for an explanation of why?

A: Dora Loizidou said that this question will be addressed to Charalambos Themistocleous by e-mail and he will contact directly the person that asked it. She also reminded that the audience can send questions to Charalambos Themistocleous via e-mail at cthemis1@jhu.edu or contact him on his website https://charalambosthemistocleous.com

3.6 Language technologies by/for the public sector

Andriana Achilleos, Director of the Department of Information Technology Services, Deputy Ministry of Research, Innovation and Digital Strategy, presented the government's general strategy on digitization and the work done so far in the public sector regarding the facilities for the citizens in eGovernment (eJustice, SYsnet) as well as the handling of digital payments during the pandemic COVID-19. Mrs. Achilleos referred to the mission of her department to upgrade the quality of life of Cypriot citizens and the quality of services offered to citizens and businesses. Making special reference to the period during the pandemic, she stressed the usefulness of tools such as the vaccination portal, COV-Tracer, the EUDCC, SMS as well as the Cyprus Flight Pass service as digitalized facilities offered.

Mrs. Achilleos mentioned that more and more government agencies are making their services available digitally and this is achieved by adjusting the EU policies in Cyprus. Areas such as health, economy, infrastructures, eGovernment serve the citizens enormously on a daily basis. The government's goal is to achieve accessibility to more services by 2025, thus saving more time and increasing administrative costs as well as expanding digitization in an increasing number of public services. As regards Cyber Security the aim is to create a secure digital environment for financial development and promotion of services in which Cyprus maintains a high position (commercial shipping, financial services).

Q: Would you consider any possibility for inclusion of Language Technology in the complex and multifaceted tasks of Deputy Ministry of Research, Innovation and Digital Strategy, in a more specialized way?

A: Achilleos agreed on that since LT belongs to Artificial Intelligence. She then said that they have enough material in their hands to start and see the issue more closely and in detail.

Natassa Avraamides, Press and Information Office, Ministry of Interior and ELRC-NAP for Public Services discussed the development of language data collection in the Cypriot Public Sector since 2018. Mrs. Avraamides initially referred to the appointment of her as NAP for the Public Services thus engaging her organization, the Press and Information Office, in the procedure of LR collection for the ELRC-SHARE repository. The PIO was until 1990 the governmental organization responsible for Translations using a system of outsourcing to partner translators. Unfortunately, most of the

documents had not been digitized, thus minimizing the volume of currently available Language resources. Nevertheless, the PIO organized the second ELRC Workshop in Cyprus in 2018 together with the ELRC Technology NAP from the University of Cyprus and deposited a considerable amount of Language Resources of monolingual and bilingual texts including Press Releases from the PIO-Newsroom and also its own Publications. Mrs. Avraamides expressed her Organization's intention to continue contributing to the ELRC-SHARE believing strongly in the development of Language Technologies and the strengthening of multilingualism in the EU and the survival of smaller languages in the EU like Greek but also Greek Cypriot.

At the end of this session, the audience was asked to answer five additional poll questions, to investigate data sharing practices and challenges. The results are presented in the following tables:

Q1: As a citizen, would you trust a digital assistant to communicate with public services and request information?

	#answers	%
Yes	14	56%
No	4	16%
I don't know/No answer	7	28%

Q2: Would you trust a machine translation system to communicate with public services in a language you don't speak?

	#answers	%
Yes	12	48%
No	12	48%
I don't know/No answer	1	4%

Q3: Does your organisation have language resources/collections of translated texts in digital form?

	#answers	%
Yes	5	22.7%
No	7	31.8%
I don't know/No answer	10	45.5%

Q4: Does your organisation use a data management plan, i.e. guidelines and standards, to make the data generated by the organisation findable, accessible, interoperable and reusable?

	#answers	%
Yes	6	28.6%
No	1	4.8%
I don't know/No answer	14	66.7%

Q5: Which of the following are the main problems that may prevent the sharing of language data? (you can select up to 3)

	#answers	%
Legal issues (e.g. copyright, personal data)	8	33.3%
Insufficient language data management	2	8.3%
I/my organisation do not/does not see any value in sharing language data	1	4.2%
We do not possess language data	4	16.7%
It is not my responsibility/I need my boss's permission	2	8.3%

Other	1	4.2%
I don't know/No answer	6	25%

3.7 Development of language resources for Cypriot Greek: challenges and prospects

In this session, Spyros Armostis, Lecturer in Linguistics at the University of Cyprus, focused on challenges and prospects in developing language resources for Cypriot Greek. Firstly, Mr. Armostis presented the languages, linguistic varieties and dialects spoken in Cyprus: Modern Standard Greek (official language of the Republic of Cyprus), Cypriot Greek, Western Armenian, Cypriot Arabic (both recognised as minority languages) and Cypriot sign language (different from the Greek sign language). He also added that existing language resources are mostly for Modern Standard Greek and that there is a limited number of language resources for Cypriot Greek. He then presented the socio-linguistic situation of Cyprus, the lack of language standardization and the lack of spelling rules for written Cypriot Greek.

With respect to existing linguistic resources for Cypriot Greek, such as dictionaries, keyboards, spelling checkers and text-to-speech, he provided historical information about Cypriot Greek dictionaries that bloomed in the last decades, since the 80's, with a number of online Cypriot Greek dictionaries appearing in the last few years. Greek Cypriot keyboards are available for Windows, Macintosh, GNU/Linux and Android (Gboard). He also referred to spelling checkers for Cypriot Greek in Google sheets and for iPhones. He explained that the latter converts Cypriot Greek and "Greeklish" to Greek script. He also presented a Cypriot Greek text converter to International Phonetic Alphabet, useful especially to speech therapists.

Thirdly, Mr. Armostis listed a number of applications which are simply not available for Cypriot Greek, such as open vocabulary databases, text annotation tools, translation tools, text-to-speech, speech recognition, as well as clinical applications (speech analysis tools by speech therapists for diagnostic purposes). He also explained the challenges that the Cypriot Greek language faces: lack of standardisation (grammatical structure and vocabulary, as well as non-codification of the writing system) and lack of research in phonetics and phonology. He also mentioned sociolinguistic factors (negative attitudes, lack of language policy, lack of corpus due to limited written use). Mr. Armostis concluded by presenting the prospects of Cypriot Greek: change of attitudes, development of databases, preparation of teaching material for teaching Cypriot Greek for non-native speakers, as well as clinical tools specialised in the analysis of Cypriot Greek language data.

Special reference was made to old poetry texts found in the Cypriot and Cretan dialects of Greek as well as the texts of the Assises, where Cypriot Greek text is found. It is worth noticing that in Cypriot Greek there are many ways of saying the same word as this has developed through time. Efforts are made to create codes for keyboards for the Cypriot Greek. Armostis gave the example of the development of online dictionaries and namely the use of wikipriaka which is digitally accessible using the Cypriot Greek language. Furthermore, the result of research was the creation of keyboards in Cypriot Greek.

Q (chat): Cypriot Greek could be considered as an official language, since it has a rich vocabulary, grammar and a different phonology. What is your opinion? If this could happen, what is the procedure for a language to become official?

A: Mr. Armostis mentioned that any language form could become an official language. He insisted that the problem is socio-political. Some people do not want or do not see the need for Cypriot Greek to become an official language and above all there is no political will. He thinks that, in order to protect the Cypriot Greek dialect, we need to change our attitudes towards it. He concluded by saying that there are also some negative effects of making Cypriot Greek an official language, as being the only official language will raise the need for Cypriots to go to Standard Greek schools in order to certify

their proficiency in Standard Modern Greek. He believes that it is important that we continue learning both varieties of Greek while trying to reward Cypriot Greek use.

Q: In translation in other languages, linguistic varieties are acceptable. For example, in languages like German (German in Germany, Austria and Switzerland), French, Dutch etc., there is a local linguistic variety in which the translation can be done.

A: Mr. Armostis agreed. He mentioned that the same thing could happen for the Cypriot Greek. Greek could be a polycentric language with Cyprus being the home country of one of its varieties.

Q: The Ministry of Education should ensure that the Greek language is taught correctly in schools. I have noticed that in Cyprus children are not taught Greek correctly in schools, maybe that is why they wish that Cypriot Greek will become an official language.

A: He replied that those who ask and discuss about Cypriot Greek becoming an official language, do not do so because they are not proficient in Standard Modern Greek or because they feel like they are lagging behind. Their opinion is based on ideological purposes. He concludes that the difficulty in the expression of Greek is not due to the presence of the Cypriot Greek, but it is mainly due to our attitudes towards the dialect.

3.8 Discussion round

The panel session of the Cypriot workshop sought to analyse the use and potential of AI and language technology especially with regard to Greek in Cyprus and Cypriot Greek. The panel was moderated by Dora Loizidou. She introduced the session by underlining the importance of AI in developing the digitization of Standard Modern Greek and Cypriot Greek. She then introduced the panelists:

- Anastasios Angelides, Head of Greek Translation Unit, Directorate-General for Translation (DGT)
- Alberto Calzada, Lecturer in AI, European University Cyprus
- Spyros Armostis, Lecturer in Linguistics, University of Cyprus
- Themos Stafylakis, Omilia

The panelists tackled subjects such as the readiness of Greek and Cypriot Greek for the AI era and the dependence on international players.

1. ***The lack of funding is one of the obstacles mentioned in Spyros Armostis presentation in the development of language technology. Are there any other the obstacles? Policies for LT? Lack of training data? Lack of infrastructure? Lack of expertise?***

Spyros Armostis agreed that obstacles in developing programs for Cypriot Greek digitization are due to lack of language policy and lack of interdisciplinary collaboration. He added that as texts in Cypriot Greek sparse, it is therefore difficult to move forward to a next step.

Anastasios Angelides referred to the general work done in digitizing different smaller languages as in the case of Irish which became an official EU language, whereas regarding the Luxembourgish, there has not been an application yet in order to make it an official language of the European Union.

2. ***What do you think is the next big leap in Language technologies development? What are your plans for future developments of your technology?***

With regard to AI era for Cyprus the panelist Alberto Calzada explained that in this area the craftsmanship is slowly being lost and large corporations have the power to promote developments as they offer “free” services, whereas smaller scientific communities do not have the financial capacity to move forward.

The researcher Themos Stafylakis mentioned the trend towards Spoken language understanding and the way to create more integrated systems of AI in language technology.

3. ***Is there a language technology industry and research in Cyprus or are we dependent on Greece or even on the big international players?***

As regards the digitization of Cypriot Greek, Spyros Armostis mentioned that a major obstacle is the lack of financial support but also the lack of language resources as they play the main role in developing language technology systems. Cyprus uses language technology systems from Greece mainly, which sometimes can be unclear in terminology matters, but also Cypriot terms sometimes are introduced to Modern Greek.

4. ***What are the obstacles in the development of language technology in the field of translation as it is experienced in the DGT?***

Anastasios Angelides mentioned that the developments in translation technology have been drastic in the course of 30 years with the beginnings being quite primitive, but the systems have been developed and today eTranslation has been incorporated in translation systems and the translators use it now, as it has become part of their everyday work. Translators have always feared the "replacement" of their work by machine translation, but today they have embraced the changes. Translators should not be afraid of the new changes in technology.

5. ***Can you describe a use case/scenario of your technology for the benefit of society/ for the benefit of public administration?***

Themos Stafylakis referred to a research (Marie Curie program) that he made in audio-visual voice recognition, where recognition is achieved through lip reading and not only through voice. The results of this research are used today in cases of persons who have lost their voice and they are now able to communicate, thanks to this technology. Mr. Stafylakis stressed that to development of such technologies is only possible when the appropriate language data are available. He also said that language technology can be of special use and benefit for the people such as the area of speech recognition or lip reading in the medical domain.

Alberto Calzada referred to code switching which (Greek-English) is very common in Cypriots' everyday life. This phenomenon often raises the question whether Cypriot Greek is further endangered. .

On this point Mr. Armostis pointed out that "translanguaging" and code mixing is something widely used and clarified that code mixing in Cyprus is not similar to Spanglish in America, but technology wise it might be said that mixed texts might appear in the future which is a problem as this might be described as noise in language data.

6. ***What are the mature services and tools for Greek (and/or Cypriot Greek) that can enhance digital services?***

Alberto Calzada replied that such mature tools are available for English and for some of the widely spoken languages like Spanish. He gave the example of a call centre, where the tool is becoming more intelligent by directing the person to a specific agent. This technology is using voice recognition and text-to-speech and he explained that the technology remains the same and he considers that it's a matter of time to do this in more languages.

3.9 Conclusions

The 3rd ELRC workshop in Cyprus gave important messages for the participants to take home. Through the very interesting and important presentation of Stelios Piperides, Institute for Language and Speech

Processing / Athena R.C., ELRC, we were able to draw useful conclusions regarding the future of digitalization of the Greek language and gained insight into the newest developments in language-centric AI.

Rapid developments in translation technologies play a very important role, and this was highlighted by Anastasios Angelides, Head of Greek Translation Unit, Directorate-General for Translation (DGT), who guided us through the different stages of machine translation in the bodies of the EU and gave us an interesting picture of the importance of digitization both for the translators, who adapted to the new age, and the end-users. Mr. Angelides stressed that we should not be afraid of changes and always embrace positive developments in technology.

Alberto Calzada showcased the fields that make use of AI and/or ML like: Voice Recognition, Natural Learning Processing (NLP), Robotics, Computer Vision.

Mr. Themistocleous, Post-doctoral researcher in Computational Linguistics at Johns Hopkins University, gave important insights into the developments in Natural Language Processing and Machine Learning in Language Disorders and into the role of NLP for timely diagnosis of mild cognitive impairment (MCI): NLP and ML can facilitate quick, and automated diagnosis, prognosis, and ultimately improvement of therapy decisions. They can thus assist medical providers by identifying subgroups of patients, evaluate the effectiveness of treatment methods by quantifying their effects on speech and language and augment current treatment and evaluation batteries for speech and language pathology.

Andriana Achilleos, Director of Department of Information Technology Services, Deputy Ministry of Research, Innovation and Digital Strategy reinforced the political commitment for the digital transformation of the Cypriot public sector.

Natassa Avraamides, Press and Information Officer, ELRC NAP for the Public Services, stressed the importance of collecting language resources for ensuring the preservation of smaller languages in the digital era.

Focusing on relevant activities for the creation of Cypriot Greek resources, Spyros Armostis, Lecturer in Linguistics, University of Cyprus, re-emphasized the importance of digitally preserving smaller languages and dialects. It was underlined anew that lesser spoken languages are at risk of digital extinction, that Language Technology can facilitate digital interaction in multilingual Europe and also that Language Technology and Artificial Intelligence need sizable training data.

Furthermore, during the workshop, it was stressed that Language Technology has the potential to change governance and administration as well as commerce and common activities of everyday life.

Apart from the very important general messages, the highlight of the workshop was to show the efforts made by experts to develop language technology for Cypriot Greek. Throughout the workshop the need was stressed both by the National Anchor Points and the presenters for the collection of more and good quality data, thus contributing to the more efficient digitalization of the Greek language.

Last but not least, it was underlined that developments at the European level towards a wider development and uptake of language technologies are necessary.

4 Synthesis of Workshop Discussions

The 3rd Cyprus workshop consisted of two main pillars: LT development for Standard Modern Greek (one of the official languages of Cyprus) and for Cypriot Greek.

With regard to the collection of language data it was stressed that the need to enhance the network of contacts in the different ministries remains a challenge, as officers change their posts and the need for better collaboration still exists. As it was made obvious by the representative of the Deputy Ministry for Innovation the Cypriot government has made tremendous progress in digitalization of the public sector and it still continues to improve different areas and expressed the willingness to look into the integration of language technologies as part of the public services portfolio. Furthermore, it was made clear that it is important that the dialects of smaller languages like Greek and Cypriot Greek survive in the new era of digitization and not have the same fate as many other small languages/dialects in the past.

5 Country Profile: Language data creation, management and sharing

The National Anchor Point for the Public Services gave a general overview of the situation in Cyprus concerning the collection of Language Data. With regard to Cyprus' Country Profile nothing has changed since the publication of the 2019 White Paper. The Press and Information Office, where the NAP for Public Services is based since 2018, has been the main contributor of language data to the ELRC-SHARE repository. Since 23 March 2020, the European Commission's machine translation system, eTranslation, developed under the Connecting Europe Facility and widely used by European and national public administrations, is now available for all European small and medium-sized enterprises (SMEs), including SMEs based in Cyprus. As it was pointed out by the Public Services NAP, behind almost every digital product we use, language technologies are hidden. Language technologies such as machine translation are a means of overcoming language barriers and supporting linguistic diversity. It is important for the Greek language to survive in the new digital era and thus for the Greek culture along with the Cypriot culture to travel to the new digital world and this can be achieved by collecting as much language data as possible in order to improve the digitization of the Greek language and with it also the Cypriot Greek.