European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2019/1083 programme.

European Language Resource Coordination
Connecting Europe Facility

Deliverable D3.2.11
Task 3

ELRC Workshop Report for Portugal

Author(s): Sara Grilo, António Branco, Paulo Vale

Dissemination Level: Public

Version No.: <V2.0>

Date: 2021-11-10

Contents

1 Executive Summary

The 3rd ELRC Workshop for Portugal took place as a virtual event on June 22, 2021. It was organized by the Agency for Administrative Modernization (AMA I.P.), the PORTULAN CLARIN Research Infrastructure for the Science and Technology of Language and the Lisbon office of the European Commission, Directorate-General for Translation, in collaboration with the European Language Resource Coordination (ELRC).

Language Technology (LT) is shaping our multilingual future. It has been transforming the way we interact with our devices and with each other, the way we shop, work and travel. It also reshapes our interactions with service providers, either public or private. Programs that automatically correct spelling errors, digital assistants that transform our voices into text messages on mobile phones, bots that answer our calls to the bank or to our social security services, systems that automatically translate from a foreign language are now part of our everyday lives, our businesses, and our administrations.

Developers, integrators, and users of LT, both from the private and public sectors, have shared experiences, requirements, and ways for transforming digital interaction in our multilingual Europe with Language Technologies. They have discussed how language data (texts and speech) can fuel the developments in Artificial Intelligence.

The workshop sought to engage participants in a fruitful discussion on the status and prospects of Artificial Intelligence and Natural Language Processing technology for Portuguese for Public Administration as well as for Small and Medium-sized Enterprises (SMEs). In a society working towards inclusiveness and where people move freely across borders, both physically and virtually, it is essential to support the development of digital and multilingual public or private services. The importance of translation and multilingualism cuts across society, with relevance in the political, social, academic, and many other fields. Besides broadening the potential customer base for digital services by reducing language barriers, the user experience is improved by giving them the freedom to interact in the language of their choice.

The workshop agenda was structured on six main topics: a) The potential of language technology and AI – where we are, where we should go; b) Language technologies for the Portuguese language; c) Digital Europe Automatic Translation Platform; d) eTranslation – Demonstration of the platform; e) eTraducao.gov.pt – Language technologies by and for the public sector; and f) Panel discussion – Creation, management and sharing of linguistic data: Existing practices and challenges.

The main finding of the workshop is that the lack of sharing of Portuguese language data prevents Portuguese from competing with the most widely spoken languages in the digital world. Both public and private sectors (e.g., translation companies) should license their data for reuse, so that not only machine translation systems can be improved but also digital public services. In addition, translators should start considering machine translation as their ally that can improve their work and their productivity, rather than their enemy.

The workshop was attended by 106 participants, from the public sector, from companies and from the research community.

2 Workshop Agenda

10:00-10:15	**Welcome and introduction** **Maria de Fátima Fonseca,** Secretary of State for Innovation and Administrative Modernisation
10:15-10:35	**The potential of language technology and AI – where are, where we should go** **Khalid Choukri**, Secretary-General of ELRA – European Language Resources Association
10:35-10:55	**Language technologies for the Portuguese language** **António Branco**, Director General of PORTULAN CLARIN Research Infrastructure for Language Science and Technology, and President of ELRA – European Language Resources Association
10:55-11:15	**Digital Europe Automatic Translation Platform** **François Thunus**, DGT – Machine Translation
11:15-11:25	Pause
11:25-11:45	**eTranslation – Demonstration of the tool** **Nelson Loureiro**, Deputy Head of Unit and Translator at the European Commission
11:45-12:05	**eTraducao.gov.pt – Language technologies by and for the public sector** **Paulo Vale**, PS-NAP for ELRC, Digital Transformation, AMA IP
12:05-12:35	**Panel discussion – Creation, management and sharing of linguistic data: Existing practices and challenges** **Moderator: Ana Lorenzo Garrido**, Representation of the European Commission in Portugal **Panel:** **João Coelho Lopes**, Head of the Portuguese Language Department of the Directorate-General for Translation of the European Commission **Fátima Castanheira**, Director of Traducta and Founder and Former President of APET – Portuguese Association of Translation Companies **João Curado**, Data Science Team, AMA IP
12:35-12:45	**Conclusion**

3 Summary of Content of Sessions

3.1 Welcome and introduction

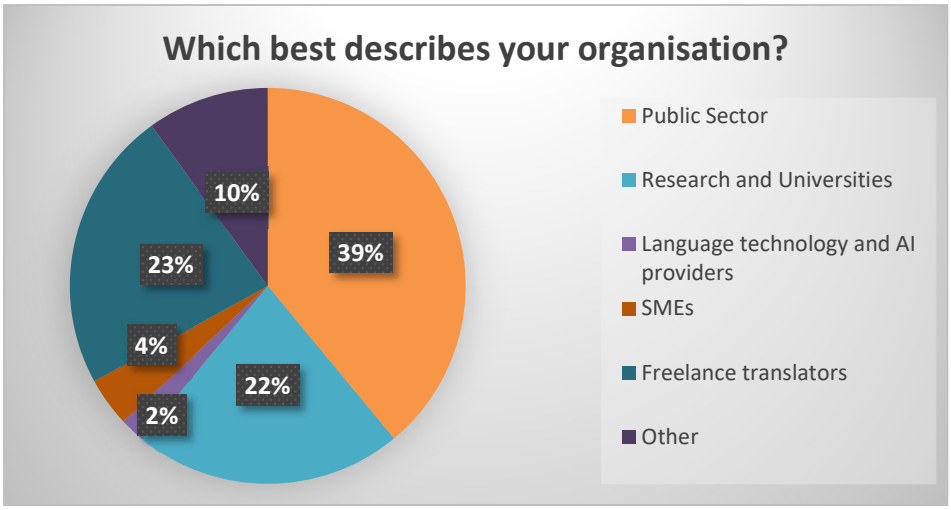
The Portuguese workshop was entitled "Artificial Intelligence and Natural Language Processing Technology for Public Administration and SMEs".

After a brief introduction to the goals of the workshop and its agenda by Ana Garrido, from the European Commission, the workshop started with Maria de Fátima Fonseca, Secretary of State for Innovation and Administrative Modernization, outlining the theme and framework of the event.
In an era where digital public services are increasingly integrated, inclusive, and accessible, the need to improve the quality of machine translation to develop multilingual digital public services is very important. And so is the sharing of language data. Linguistic technology has an impact on our daily interactions, with spell checkers, digital assistants, or machine translation (MT) services. MT also contributes to increase the productivity of translation professionals.

In Portugal, progress is being done in the pursuit of this goal, as seen in the eTraducao.gov.pt repository.

One of the goals of AMA is to promote the implementation of the regulation defining the Single Digital Gateway (SDG) and to provide online access to accurate and up-to-date information, assistance and problem-solving procedures and services. SDG is available in all official EU languages.
It is worth mentioning the Recovery and Resilience Plan, which consists of a set of AMA actions and the identification of services that will be renewed as digital and multilingual services.

At the end of this introduction, we asked the participants to answer a poll question to investigate their employment status.



3.2 The potential of Language Technology and AI – where we are, where we should be heading

Khalid Choukri, Secretary-General of ELRA (European Language Resources Association) and representative of the ELRC consortium, presented the evolution of language technology, including machine translation, and their applications. Where we are standing, particularly when we combine Artificial Intelligence (AI) and Language Technology (LT) and where we are heading/should try to head were the main topics of his presentation. He presented the evolution of the machine translation system from the statistical to the neural paradigm.

Language technologies (LT) were presented into the following categories:

- Speech Technologies
- Translation technologies
- Terminology technologies
- Localisation technologies
- Natural language understanding (NLU) technologies
- Text analytics technologies
- Multilingual and semantic search technologies
- Optical Character Recognition (OCR).

Khalid also presented trends and long-term prospects, such as UNESCO's decade of activities in indigenous languages (with the contribution of ELRA's LT4ALL initiative) and the European Language Equality (ELE) Project, which includes non-official European languages. He said that now a lot of focus is being put on social networks and other media for hate speech detection and media monitoring with the help of language technology.

To strengthen AI and LT in the European Union (EU), we need to continue:

- To identify strategic sectors with EU strength (e.g., multilingualism)
- To develop an EU-centric LT and data policies with:
    - international partnerships
    - not only market-driven
    - particular attention to non-official languages
- AI transparency (systems easy to understand, and AI regulation)
- Real funding for EU players (e.g., public procurements)

3.3 Language Technologies for the Portuguese language

António Branco, Director General of PORTULAN CLARIN Research Infrastructure for the Science and Technology of Language, and President of ELRA European Language Resources Association, began his presentation by recalling that Portuguese is the 4th most widely spoken language in the world (out of 7000), the 5th most used on the Internet, the 6th language most used on Twitter, and also one of the most spoken languages on Facebook.

With other authors, António Branco has published the White paper *The Portuguese Language in the Digital Age* as part of the META-NET project and from the references selected in top conferences, between 2010 to 2012, in all the languages represented, the authors drew the conclusion that Portuguese had less than 50 references (as opposed to English, which counts more than 900), ranking 11th among the languages with more references.

Having users and webpages in Portuguese is not enough to ensure prosperity in the digital age. It must be technologically prepared to face the technological shock of the digital age.

In Portugal, there are some companies whose primary goal is to deal with language technologies, as for instance *DefinedCrowd* or *Unbabel* (however, these are American companies funded by Portuguese scientists).

Research and innovation also play a key role in improving language technology and digital resources in Portuguese.

The PROPOR international conference takes place every other year, alternating between in Portugal and Brazil, and brings together a community of researchers from various fields (linguistics, language technology, speech processing and many others), contributing to the creation of Portuguese resources.

The Research Infrastructure for the Science and Technology of Language PORTULAN CLARIN has partners from all over the country and also from Brazil, from different scientific areas. It is one of the infrastructures that collects more tools in Portuguese. It integrates a repository, where it is possible to find many types of resources: corpora, lexical conceptual resources, language descriptions, tool services, etc. It also includes a workbench, which integrates several grammatical analysis services (syntactic, semantic, among others), a sentiment analysis service, etc. PORTULAN CLARIN encourages the sharing of language that contributes to improving language technologies for the Portuguese language.

It is also important to mention international integration. For example, the ELG project, a 3-year European project that compiles resources in several languages, including Portuguese, can be an entry point for those looking for language resources.

From 2015 to 2020, there was the ELRC Plan for the Advancement of Language Technology, which is fundamental, as far as the authorities do not want their language to be left behind.

Finally, António Branco presented a strategic challenge to improve digital services for Portuguese, considering that the Portuguese language is one of the pillars of Portugal sovereignty. To survive and thrive in the digital age, the language needs to be scientifically studied and technologically prepared. Only then can we have access to all the people, services, and goods that will become available in and through the information society and ensure that we have full citizenship in the society of the future.

There was a single question (twofold) from the audience, which António Branco and some other participants answered:

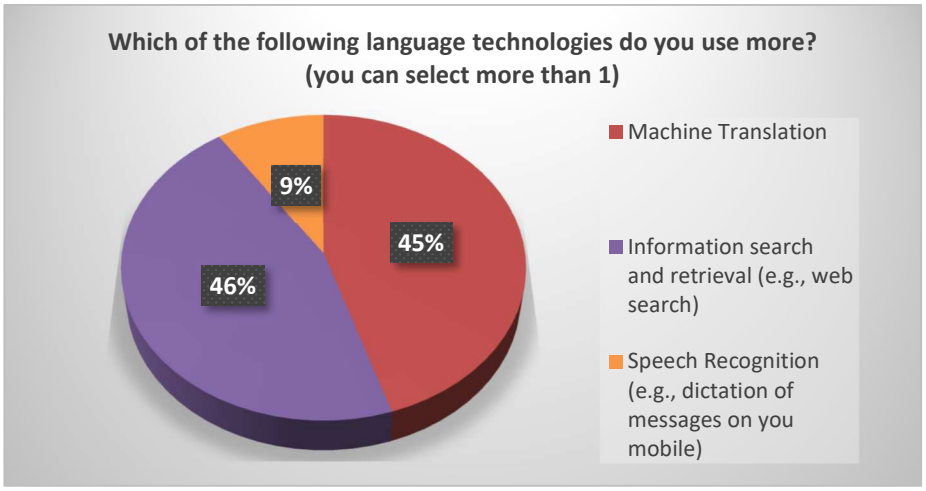
Q: There are several codes of the Portuguese language, the Brazilian variant being predominant. I wonder if you are working with organizations like Unicode regarding the standardization of Portuguese nomenclatures. Right now, PT is equivalent to the Brazilian variant, and Portuguese of Portugal appears as a subvariant of Portuguese, with the nomenclature being PT-PT.

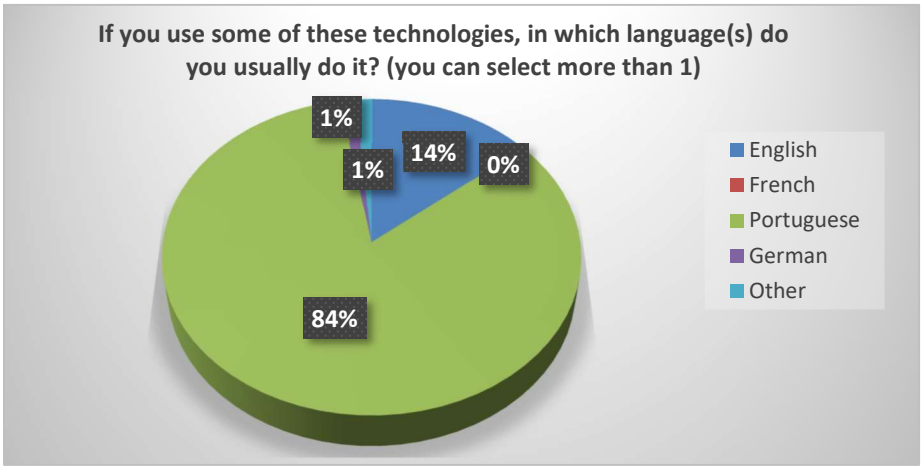
A: According to the Camões Institute, there are 271 million speakers of Portuguese – with 10 to 11 million in Portugal. The promotion of Portuguese must accommodate all variants, including the one with the largest number of speakers (Brazil). Portuguese is not less important with PT-BR being the largest variant. That's 210 million speakers versus 10 million.

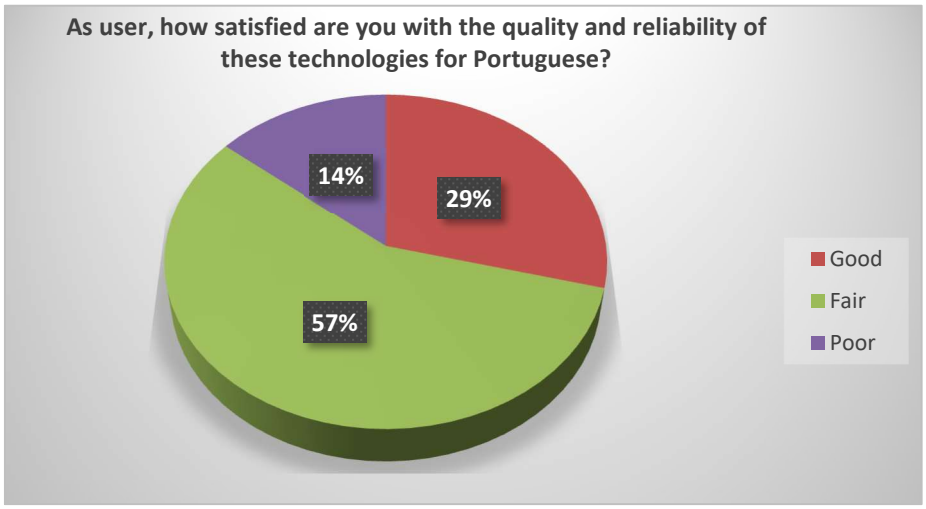
Q: Although there is no doubt that Brazil is mainly responsible for the spread of the Portuguese language worldwide, this is no reason to explain the attribution of PT to Brazilian Portuguese.

A: There always must be a chance to choose PT-PT. If English and Spanish have all the variants available, why not Portuguese? In practice, what the guide indicates is that the content identified by the ISO-639-1 code is suitable for the most common national variety, expressed by the ISO-639-1 + ISO-3166 pair, as "PT-BR". What determines the assumed pair is the number of speakers, so the most common variety is PT-BR.

After the presentation and Q/A, three polls were submitted to the participants to investigate their experiences as users of language technologies.







A few highlights from the participants' answers: the workshop attracted a significant number of participants that are equally interested in machine translation and in information retrieval systems, a few of them acknowledging their interest in speech recognition systems (on mobile). Also, they use LTs in Portuguese mainly and considered that the quality of these systems in Portuguese is "fair enough".

3.4 Digital Europe Automatic Translation Platform

The Digital Europe Automatic Translation Platform was presented by François Thunus, Directorate-General for Translation (DGT) – Machine Translation. He presented Connecting Europe Facility (CEF) eTranslation, which is for CEF digital service infrastructures, pan-European digital public services, public administrations in the Member States, Iceland, and Norway, and SMEs. He also presented DGT eTranslation, which is for EU translators and officials, and digital services of the EU institutions.

The eTranslation platform can be accessed in two ways:

- A web user interface for humans to translate texts automatically to be further reviewed by human translators, or just a quick translation to get a gist of the text.
- A machine-to-machine service (API) to integrate machine translation in workflows, websites, digital services, etc.

eTranslation supports all official 24 EU languages, plus Icelandic and Norwegian and provides not only a general language engine, but also domain-specific engines, such as the EU formal language engine. François Thunus commented on the translation output quality, underlining that, because eTranslation has been trained on a huge database of translated official EU texts, its performances are very good for the formal EU language, since the supporting data are all translations done in the EC since the beginning, but may not be as good when it comes to non-standard or creative texts, words, and expression isolated and context dependent. However, the availability of the general language engine, which is trained on respective non-official texts, delivers high-quality output. François Thunus highlighted the need to select the appropriate domain-adapted engine according to the text type to be translated. There are ten domains that can be chosen:

- Formal EU language
- General text
- Case-law of the Court of Justice
- Cultural
- Deutsche Bundesbank (Germany)
- Intellectual Property Law
- Ministère des Finances (France)
- Public Health
- Technical Regulations Information System
- Valtioneuvoston Kanslia (Finland)

Regarding plans for the development of the CEF AT platform, François Thunus noted the need to increase the domain coverage (e.g., scientific text); support more languages (not only European languages with economic or social importance); develop more language tools, such as anonymisation tool and a basic Computer-Aided Translation (CAT) tool. Some tools (eTranslation, multilingual tweet, Named-entity recognition, among others) have been made publicly available at https://language-tools.ec.europa.eu/.

Finally, François Thunus gave an overview of the eligible individual users. They must belong to the following institutions:

- Public Administrations
- Universities
- Projects financed by CEF
- SMEs (after validation)

He then presented the steps to self-registration and use of the web service integration (API), along with the technical documentation and the service desk.

- Self-registration via https://webgate.ec.europa.eu/etranslation/public/welcome.html
- Web service (API) Technical documentation: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/How+to+submit+a+translation+request+via+the+CEF+eTranslation+webservice
- eTranslation Service Desk: help@cefat-tools-services.eu
- Access to eTranslation web user interface: https://webgate.ec.europa.eu/ETRANSLATION

There were two questions asked through the chat.

Q: I have tried eTranslation several times and am not convinced by the results. I wonder if it has to do with the language combination (Dutch - Portuguese - Dutch).

A: Yes, unfortunately, it does, in an indirect way. The official engine is PT-PT (European Portuguese). The general text engine can be influenced by PT-BR (Brazilian-Portuguese). The data comes from the web and even if we try to "clean"/approximate the spelling to PT-PT, it is not possible, given the mass of data. Here the problem is the number of combinations. There are approximately 30 languages available. In the case of Dutch-Portuguese-Dutch translation, the process must use English as a pivot language and information gets lost along the way.

A: Automatic translation often ends up being done using Brazilian Portuguese because there is a larger amount of data in the computational universe.

Q: Can freelance translators also register for eTranslation?

A: Yes. The system varies from country to country, but you can contact the helpdesk.

A: Yes, but in principle, they are entitled, although sometimes the mechanics are a bit complicated and it is necessary to contact the service directly, it is not an automatic enrolment, but they are entitled.

A: We are not yet fully open to the public, for physical reasons. The infrastructure is quite large and costs a lot of money and we don't have the financial resources of big tech companies such as Microsoft. For security reasons, the infrastructure must stay in Europe, in the same premises as the European Commission, and this also includes the IT Centre. The extension possibilities are limited. After opening to SMEs, the success has been huge. We have approximately 600 translators for all the languages, who translate millions of pages every year. Last week the system exceeded one hundred million pages. To reach this level with human translators, the EC would have to hire 30,000 more translators. Given that the EC staff amounts to 32-35 thousand, it would mean doubling the number of employees and it is not possible.

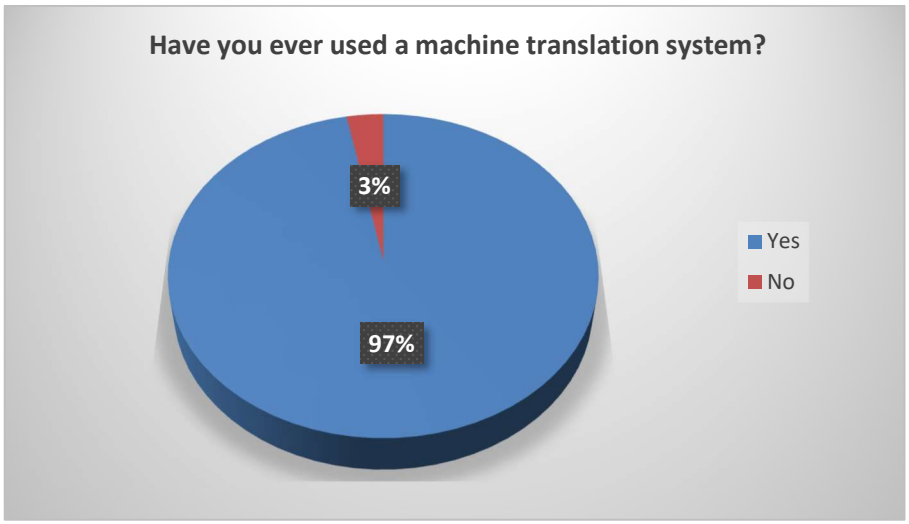
A: In any case, if freelancers can have access to eTranslation it will be very beneficial to us and to the European institutions themselves.

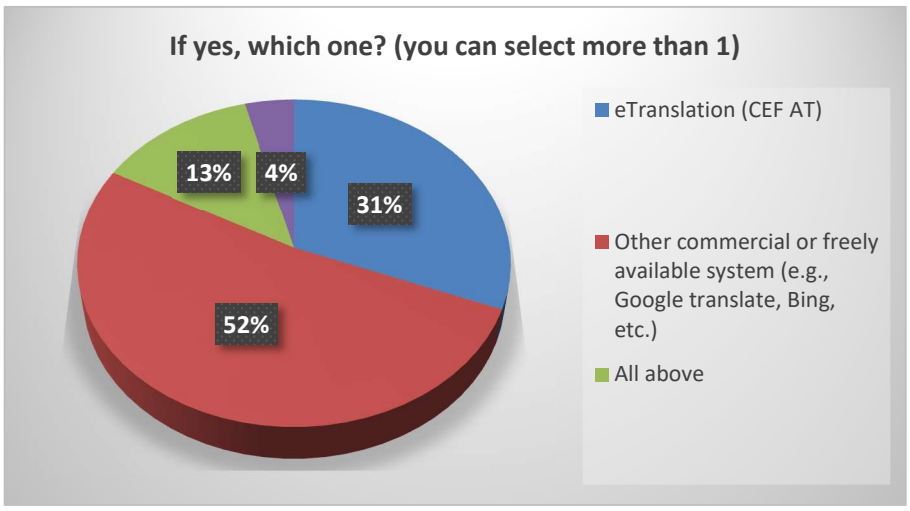
There was another question asked through the chat, but it will be answered after the workshop.

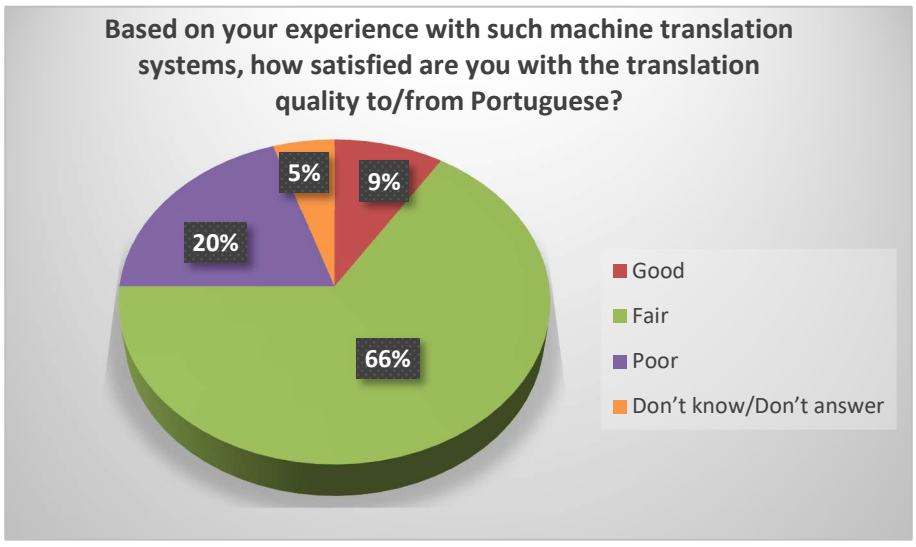
Q: In e-translation, when we correct the result in Portuguese and evaluate it with the star system, is this contribution considered and entered into the automatic translation system?

For the records, the brief answer is "yes". As there was no more time left, the participant was answered by the panellist with a more complete answer after the workshop.

At the end of this presentation, three polls were submitted to the participants to investigate their experiences with machine translation systems.







3.5 eTranslation – Demonstration of the tool

Nelson Loureiro, Deputy Head of Unit and Translator at the European Commission made a demonstration of to how use eTranslation. The main steps are as follows:

- File upload in one of the supported formats (multiple files can be uploaded simultaneously)
- Domain selection (one of the domains mentioned above)
- Language's selection: original language and output language
- Output format: same as original
- Output reception: receive the file translated by email

He explained that if we want to translate a specific sentence from a language that we don't know, the best option is to translate it into English. That way, we can get an idea of the meaning in the original sentence.

Finally, Nelson Loureiro concluded his presentation by stating that translations done by professional translators feed machine translation engines. If there is an error in a human translation, it will be reflected in the MT engine. To add value to machine translation, translators must ensure the quality of their translations which will also improve machine translation outputs, therefore reducing the time spent on these corrections and increasing their productivity. In "older" domains, there are fewer errors to correct than in new domains, with new terminology, so professional translators will spend more time on corrections, but will improve machine translation results in these cases.

There were many questions asked through the chat.
Q: If our text contains personal data, for example, about of customers, are we breaking rules? Do we have to anonymise our texts?

A: If it is possible to anonymise, it is better. We do our best to have a maximum level of security, but the ideal process is to anonymise. However, if it is a large text, it becomes more difficult to anonymise. As said, there will be a tool for anonymisation, which will make the process easier. As the tool is for the EC, the risks are limited.

A: Unlike other services, eTranslation does not keep the original texts nor their translated version. All files are deleted. It is a safe tool, only saving the data for 24h or until it is downloaded.
A: As soon as the text is no longer needed by the system, it is immediately deleted. The intermediate files are created on the spot and no relation between source and intermediate or temporary files can be made.

Q: Can or should we, the users, also correct the results? – This question was reworded as: how can external users contribute to eTranslation? What are the ways to contribute texts (monolingual or bilingual) to improve the tool?

A: There is a page to contribute. The link is provided in the chat https://elrc-share.eu/static/metashare/Walkthrough_for_Contributors.pdf.

Q: From a translator's point of view, what is your opinion on the results of automatic translation processed on eTranslation? Do you consider them to be satisfactory, or do they need any corrections? Which language pair gives the best results?

A: In 4.5 years, Nelson Loureiro has worked with 3 French texts and the rest in English, so he doesn't have much experience with other language pairs in everyday life. English is very good in domains where the terminology and phraseology are already very settled. For new domains, such as AI, there is still a lot of new stuff, and there you lose some of the value of machine translation as a resource. Machine translation, to translate in itself, is not great, but with continuity, it will improve. More objectively, a 15-page document for an old domain I can do in one day; for a new domain, I will need

2 or 3 days, as I will have to make many more interventions at the machine's suggestion. Besides the issue of the translator's translation style.

A: The system is not static, meaning we are making new engines every 3 or 4 months. Language changes and things are not said in the same way today as in the 1950s, for example. Retraining the neuronal network takes 10 to 12 weeks, day, and night, on very powerful machines. It is always important to have a human review. The problem with neural networks is that they are too good. In the old days, it was obvious that it was a machine translation. Now the machine is much more subtle, a sentence may be missing - because the machine made a mistake in segmentation. It's obvious that it must be revised.

Q: Which language pair has the best machine translation results? What about the DE-PT language pair? Do you have any information from your colleagues about this language pair?

A: Working with DE-PT pair (too) is rare, so I don't have much information about the quality of eTranslation in that case. However, I think it will also use English as a pivot language. So, as already mentioned, it is natural that the quality suffers a bit due to the successive treatment.

3.6 eTraducao.gov.pt – Language technologies by and for the public sector

Paulo Vale, PT-NAP for ELRC, Digital Transformation, AMA IP, presented the eTraducao.gov.pt repository, reinforcing what was said earlier about the importance of data for the development of LT.

The quality of translation is linked to the volume of data shared with the machine. In public administrations, a lot of multilingual data are created daily, however, much of this information is not reused and their full potential is not used. Paulo Vale reinforced the importance of sharing data among public and private institutions to improve machine translation. The greater the sharing, the better the quality of the machine translation.

This is one of the reasons why the European Commission created ELRC, aiming at improving the quality and coverage of CEF eTranslation and especially looking at its use in online digital public services, which will benefit all European citizens.

Paulo Vale showed the ELRC repository (https://elrc-share.eu/), the repository of resources collected from Public Administrations that aims to improve the performance of the MT engine by increasing language and domain coverage. Only 147 resources are in Portuguese. Some of the reasons for the low level of sharing are listed hereafter:

- Distrust of a service outside Portugal
- Unawareness of what can be shared
- Lack of mandate to authorise sharing
- No obvious benefit from sharing
- Resistance to adopting new working methods
- Low quality of machine translation outputs

After that, Paulo Vale presented eTraduçao, the national repository of translation resources. This national infrastructure enables the collection, processing and sharing of language resources, namely translations to and from the Portuguese language that can be used to improve machine translation services. eTraduçao complements ELRC-share repository.

eTraduçao is an online platform with restricted access (registration required). Below are some of its features:

- Web-based platform
- File upload restricted to authenticated users
- Downloading resources according to authorisation level
- Resource search including filtering criteria
- Accepted formats: doc(x), odt, tmx, sdltm, xml, tbx, xls(x), txt
- Access to the upload history.

The steps for resource sharing process are:

- Uploading of resources (language pair can be uploaded simultaneously)
- Sharing conditions:
    - Only eat the institution itself + AMA (level 1)
    - With other national institutions (level 2)
    - With European institutions (level 3)
    - Public (level 4)

The resources are automatically processed for alignment, formatting, language, cleaning, and conversion to TMX format. The resources are available for download once they have been processed. Then the resource can be used with any machine translation tool (Trados, MemoQ, eTranslation).

The eTraduçao repository is a safe place for storing translated resources. It can be used in machine translation of digital services – making some digital public services multilingual and contributing to the development of Information and Communication Technology tools. With this repository, security, confidentiality, and compliance with intellectual property right are ensured. Finally, it is a great tool to promote the Portuguese language.

Paulo Vale ended his presentation reinforcing that every document shared will help to support the presence of the Portuguese language on the European scene, preventing its digital extinction, and it will improve the quality of machine translation services for everyone working in the Portuguese public administration.

One question was posed through the chat.

Q: How do you do quality management? Can you ensure the quality of the translations shared?
A: At the repository level, we do not interfere or even look at the resources. However, this is a restricted access repository, entities (Public Sector bodies/services/depts) end up identifying themselves, which means that, in the case of translations, a certain quality level is guaranteed. We do not validate whether one translator is better than another or whether one translation is better than another. We are interested in having volume, which is the key word that answers that question; bad translations end up being swallowed up by good translations; volume will give quality, it is the statistics applied to the resources that are collected that give quality. We could go even further, have even more shared resources, contacting the entities directly, but it takes time for the resources to be shared.

3.7 Panel discussion – Creation, management and sharing of the linguistic data: Existing practices and challenges

The panel discussion addressed the existing practices and challenges in the creation, management and sharing of linguistic data. It hosted three panellists, two from public services of translation (Directorate-General for Translation of the European Commission and AMA) and one from the industry. Ana Garrido, Representative of the European Commission in Portugal, and moderator of the panel discussion, initiated the session by presenting the theme and introducing the panellists:

- João Coelho Lopes (JCL) - Head of the Portuguese Language Department of the EC's Directorate-General for Translation
- Fátima Castanheira (FC) - director of Traducta and founder and former president of APET
- João Curado (JC) - AMA data science team, comprehensive overview of the situation in Portuguese public administration

Discussion points:

What is the situation in terms of language data creation and sharing in the EC? What developments have been achieved recently?

JCL – At the European Commission we encourage sharing (regardless of the quality of the language pairs) of everything, not only what is translated internally but also externally (by freelances). Machine translation improves with contributions - neural MT is performing very well, including in the English-Portuguese pair. By feeding the translations into the database, we are improving the MT results. Given the volume of work we have, MT is crucial for them to be able to respond. It is essential that within the EC data is shared. The same goes for the Portuguese public service, and Portuguese e-translation. The more data are contributed to the repository, the better the results will be, which will contribute to increase the confidence in MT with quality improvements in the output.

How have these developments in translation tools affected the work of translators in a European institution? Notably in terms of diversification of tasks and productivity.

JCL – MT had a significant impact on the department's translators (specifically Portuguese). The most significant contribution was the possibility to expand the range of knowledge that translators can do, becoming more productive. MT contributed to the diversification of translators' knowledge and evolution, considering the policy of contention and reduction of resources within the European Commission.

What is the vision of a translation company when it comes to sharing data? Do confidentiality or trade secret issues arise?

FC – Traducta has existed for over 35 years and has a wealth of archived information on much of the work done that can be shared, especially since the use of computer-assisted translation tools and translation memories. The Portuguese language must be given more volume of data.

Then, we recognise that post-editing is not well covered in the translation curricula. Young translators have very little knowledge of what post-editing is and are not prepared for it. There is a gap between the market demands and classes followed during translation studies to prepare translators to the market's needs. For many translators, reviewing content that has been translated by a MT engine is challenging. MT output can give the impression that the translation is correct (syntax, terminology) whereas the meaning is not carried properly which can mislead young and untrained translators.

People are under the impression that, nowadays, any machine can do a translation and you do not need a human translator, much less a specialized professional. How do you see the evolution of the translation market? Is it a market in evolution or decline? What are the implications for the profession?

FC – I don't consider all this technological evolution as a threat, and it has a huge impact on the way translation activity exists and will continue to exist in the future. But this requires openness on the part of translation professionals, which does not always happen. In part, university teachers can probably be blamed for not keeping up with the evolution and needs of the markets. As in all areas, new technologies have already had and will continue to have a major impact on the translation business. And only professionals who are open and willing to invest in continuing education and keep up with the evolution of our profession will play a decisive role in the future of our profession. If we close ourselves off, limit our knowledge, and do not keep up with what is being demanded of us, it will be very difficult for us to survive. But I count on the intelligence and cleverness of all translators to understand this and invest in their continuing education to take the best out of all these new technologies. There is no doubt that machine translation, especially when based on many years of work and study, allows us to enter areas that we previously dared not enter because we feel more confident with the result of machine translation. For all these reasons, I believe that we translators can only benefit and should not be afraid of all these challenges, because this is life, it is a continuous challenge. When the so-called computer-aided translation tools were introduced, it was very difficult to convince translators of the need to use these tools, which improved the productivity and the quality of their translations, by ensuring terminological consistency. Nowadays, we must also convince translators that we need to invest in post-editing, and machine translation is something that helps translation, and increases productivity. We should not close ourselves off, but rather be open and participate in training courses, be enlightened so that we can take advantage of this market development.

Can you explain us the data sharing policy in public administration, not only language data, but data in general? What are the challenges? How do data protection regulations apply?

JC – In Portugal, we have a GOV data portal. This portal exists since 2011 and has been revised three years ago to be better prepared for the challenges in terms of data, both in volume and typology, and in the management of the open data community and ecosystem in Portugal. It is recognised as the official portal for Portuguese public administration, as a repository of open data, and can gather information from other open data portals that already exist in Portugal, namely sectoral portals, such as justice, health and environment, and more local portals, such as municipal councils. Lisbon Municipal Council is a successful example of what has been done in this domain. The GOV data portal aims to gather information on open data in Portugal, not only at the Portuguese public administration, but it also has the ambition to become a portal that allows sharing data from the private sector as well.

Data protection is often seen as an obstacle to data sharing. It is not always true. For instance, some open data that do not contain personal data nor sensitive and confidential information, can be published under specific licences created for this purpose. In other words, they are protected in that way in legal terms.

The challenges we face are related to the volume and typology of data, increased real-time data sharing, and the need to change our mindset. As entities share data, they increasingly realize the benefits of such sharing because this data is used and combined, applications are made, has various uses, and knowledge is built, which also benefits the entities themselves, which often do not have technical or financial resources to process this data.

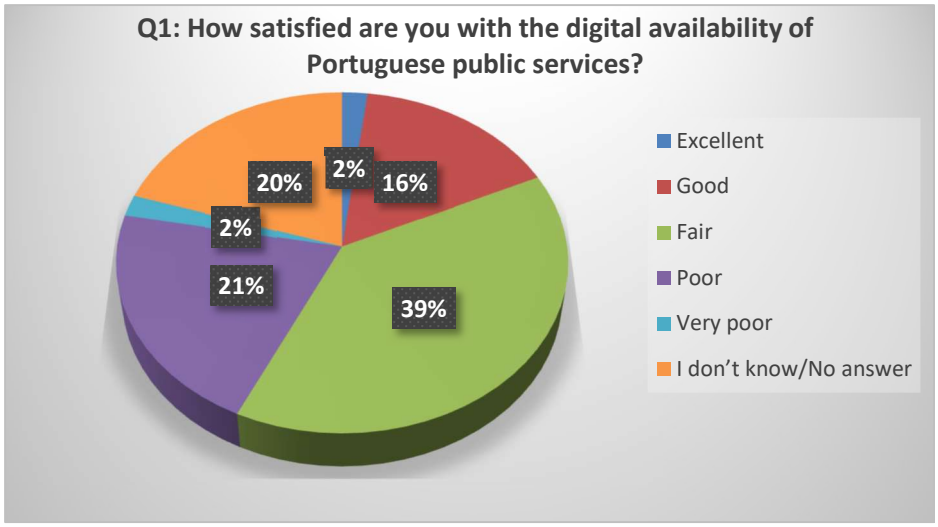
What are the prospects in terms of data use and sharing? What are the concrete benefits for individuals and what, if any, are the dangers?

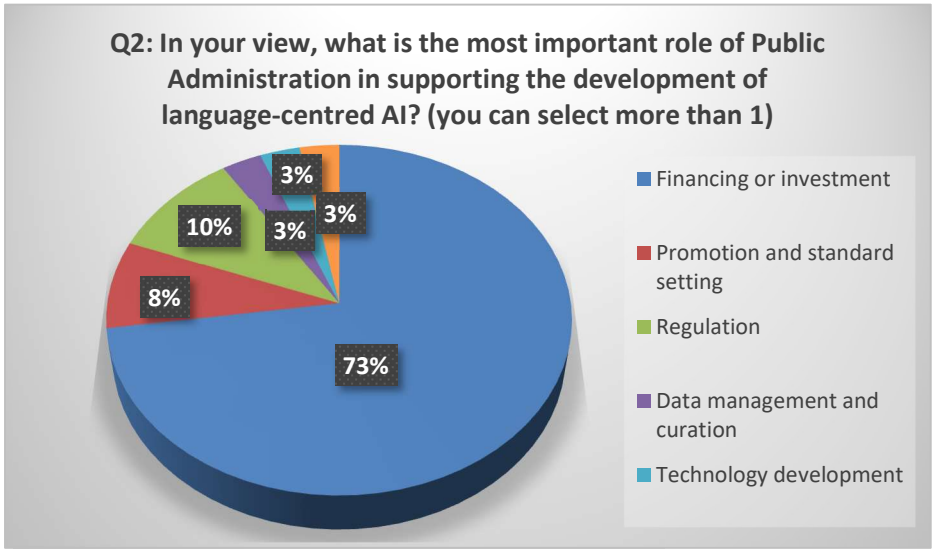
JC – Open data contributes to transparency and accountability and promotes democracy. They help to combat the phenomenon of fake news, with specialist journalists who research to validate the information to be disseminated. These advances allow easier search and access to data, with less bureaucratic formalities, less redundant and less costly. Using open data, we can develop applications and websites. All of this can bring greater credibility to political decisions by public and private entities, as they will be better grounded and adjusted to the needs of the communities. The sharing of open data allows easier access to information and knowledge acquisition for everyone, also contributing to the development of the scientific community. In addition, this sharing allows faster and more adjusted responses to crisis contexts, such as, for example, in the context of the pandemic we are going through. The EU estimated, by 2030, that the value of open data and its reuse could reach 194 billion euros. Allow the development of the business community. The progress of digital technologies stimulates digital innovation, especially concerning AI itself, which without quality data cannot evolve in the direction we want, in a correct perspective in ethical terms. The issue of opening data and combining data between the public and private sectors also allows for the construction of smarter cities, in terms of energy efficiency and mobility, for example.

Directive 1024, on open data and held information about the public sector, defends the need for data downloaded in bulk to be machine-readable and accessible through APIs. This motivates the Member States and their respective entities to open each time plus the data. We must contribute to data literacy and open data to eliminate possible cultural or other barriers so that we can more quickly reap the benefits that will flow from this.

There were questions/comments through the chat defending the teaching of post-editing in the translation courses of the various institutions.

At the end of this presentation, we asked the participants to answer two poll questions to investigate their satisfaction with digital Portuguese public services, and the role of Public Administration in supporting the development of language-centred AI.





3.8 Take-home message and conclusions

This evolution in machine translation and digital services is unstoppable, and we all must keep up with in order not to be left behind. Accepting the change and the benefits machine translation brings to human translation have a key role in changing mindset.

There is an urgent need for increased sharing of data that feeds machine translation.

The prospects are continuing with these conversations, in less ambitious initiatives, in small seminars, to talk about this and give evolution to the work.

4 Synthesis of Workshop Discussions

The existence of the eTranslation service and its enlarged availability for a wider range of types of users was visibly very much appreciated by the audience, in general. In that respect, the workshop achieved one of its major goals, namely the awareness raising about this service.

The workshop was also instrumental and highly invaluable in as much as it counted with the opening address by the Secretary of State for the Public Administration. In her speech, she highlighted the importance of multilingualism, the need to provided public services in a wider range of languages to serve a wider range of purposes and citizens, both national and foreigners, and the very positive role of machine translation in this respect, namely the more widespread adoption of machine translation by the Public Administration, under her political direction. This was very important in more than one count. On the one hand, the elaboration and delivery of this speech permitted that she grew and consolidated a positive attitude by herself towards this subject. On the other hand, this send a strong and authoritative signal to the audience about the need of pushing for the promotion of multilingualism and machine translation, which reached many officers, from top to low ranked ones, as well as many stakeholders in the field, that were present in the audience.

The workshop was also useful and important to further disseminate the existence and availability of the national data sharing platform eTradução, by means of which users may upload and share their language resources, including multilingual datasets, to be reused by other users and by the EC in the further development of the eTranslation service.

Interestingly, the issues that emerged representing roadblocks for an ever more widespread sharing of language resources that may support enhanced machine translation point towards the same key lessons learned in the previous editions of the workshop. While translators and other stakeholders are more aware of the invaluable impact of sharing language resources and which platforms to resort to undertake it, they still underline the unclear aspects related to the responsibility involved in that sharing. These aspects touch on the problems of privacy and anonymization of the data and, when it comes to the Public Administration, also to the legitimate source of authorization for the sharing of language resources. Sometimes some questions from the workshop participants, while apparently asking about some lateral questions, one understands that they boil down to the issue of the losing of competitive edge to competitors when releasing data that the latter can use to overcome the business advantage of the original owners of that data.

Despite the sanitary emergency, this challenge was sought to be addressed with the workshop being undertaken as a fully online event. When compared to the onsite modality and its full benefits, this was far from an optimal way of delivering the workshop. Nevertheless, given the very positive results described above, it was still very worthwhile organizing this workshop.

5 Country Profile: Language data creation, management and sharing

The current situation in Portugal with regards to the practices for multilingual data creation and sharing received a significant change with the availability of the platform eTradução, established and run by AMA the national Agency for Administrative Modernization. As described at length during the workshop in the talk by its coordinator, Paulo Vale, this platform allows its users to upload and share language resources aimed at supporting multilingualism and machine translation, with other users and also with the ELRC central repository. This has permitted that the number of resources upload and shared has been increasing steadily, as presented in detail in Paulo Vale's slides corresponding to his talk in the workshop.