

Malta National Language Technology Platform: A vision for enhancing Malta's official languages using Machine Translation

Keith Cortis

Malta Information
Technology Agency
Gattard House,
National Road,
Blata l-Bajda, Malta
keith.cortis@gov.mt

Judie Attard

Malta Information
Technology Agency
Gattard House,
National Road,
Blata l-Bajda, Malta
judie.attard@gov.mt

Donatienne Spiteri

Office of the State Advocate
Casa Scaglia,
16, Mikiel Anton Vassalli Street,
Valletta, Malta
donatienne.spiteri
@stateadvocate.mt

Abstract

In this paper we introduce a vision towards establishing the Malta National Language Technology Platform; an ongoing effort that aims to provide a basis for enhancing Malta's official languages, namely Maltese and English, using Machine Translation. This will contribute towards the current niche of Language Technology support for the Maltese low-resource language, across multiple computational linguistics fields, such as speech processing, machine translation, text analysis, and multimodal resources. The end goals are to remove language barriers, increase accessibility, foster cross-border services, and most importantly to facilitate the preservation of the Maltese language.

1 Introduction

Work in Computational Linguistics and Natural Language Processing (NLP) has been gaining a lot of traction in recent years. Particularly with the advance of the World Wide Web, this field now has access to thousands of languages. Whilst initially the field focused on a smaller number of these languages, namely high-resourced languages such as English, Chinese, and German, it is becoming increasingly evident that substantial effort needs to be invested in low-resource languages, particularly those at risk of becoming extinct. Here, the Web could act either as a catalyst to the disappearance of such languages, especially if the language is spoken by minorities, or it could help preserve such languages if the relevant resources, including NLP tools, are made available.

In the research conducted by Rosner and Joachimsen, the authors conclude that the level of technological support for Maltese, compared to that for other European languages, is extremely low, giving rise to the threat of "digital extinction" for the language (Rosner and Joachimsen, 2012).

In this context, "digital extinction" refers to the possibility that Maltese, while possibly maintaining its status as a national tongue with institutional support and recognition, is nevertheless virtually absent from digital media.

The current state of language technology (LT) support for the Maltese language is weak or provides no support with respect to **speech processing, machine translation (MT), text analysis and speech and text resources**. Considering this current state of Maltese LT support, the English language has become the default language of choice across most technological devices in Malta, simply because there are more language resources and tools available and they are easier to use. Taking this in consideration, prominence should also be directed towards Maltese language resources and tools, such as lexicons, corpora, lexical resources, language models and natural language processing tools, that can be used by any Artificial Intelligence (AI) technology initiative.

Investment in language technologies, tools and resources, especially those focused on low-resourced languages such as Maltese, provides the opportunity to address the complex issue of multilingualism in Europe and thus caters for citizens' needs to communicate across the language borders of the European Common Market. The developed Maltese language tools and resources can also feed into the European Language Grid (Rehm et al., 2021); the primary cloud-based platform for LT in Europe. The aim of this platform is to boost the Multilingual Digital Single Market towards a prosperous European LT community, which shall create jobs, new markets and new opportunities.

In a concrete effort to increase resources in the Maltese language and facilitating the preservation of the Maltese language, this paper presents a vision towards the development of a National Language Technology Platform (NLTP) for Malta.

This work is part of a CEF Telecom – Automated Translation (EU) project which has started in April 2021, whereby Latvia, Estonia, Croatia, Iceland, and Malta shall be developing a NLTP for each country for language tools and resources¹. With respect to Malta, the Maltese adaptation of the NLTP will be implemented through a collaboration between the Malta Information Technology Agency (MITA); the technology partner responsible for Malta’s NLTP development, the Office of the State Advocate (OSA); a pilot partner with a specific focus on the legal domain, and the University of Malta (UM); a research partner who shall be working on neural approaches towards multimodal machine translation (MMMT). Moreover, the Malta NLTP aligns with Malta’s Strategy and Vision for Artificial Intelligence (Schembri, 2019) where the country is investing in the development of Maltese language tools and resources.

2 Multilingualism Challenges

Multilingualism is considered to be a cornerstone value within Europe, where the language diversity is embraced as a basis to good administration (García, 2002). This linguistic diversity however creates many challenges stemming from language insularity, and impacts a number of domains, including public administrations, public e-services, and the cohesion of the European Union Digital Single Market (DSM).

In Malta, both Maltese and English are official languages, with Maltese –a language of Semitic origin written in the Latin script– being the national language. Maltese, an official language of the European Union, is spoken by 97.25% of the population (around 500,000 people)². The public administration may use any of these languages for official purposes. Therefore, more prominence and importance need be given to the digital aspect of Maltese. Given that Malta is highly multilingual, the following are the main language challenges being faced, especially when the Maltese language is used for communication purposes:

- In the past decade or so, the number of foreign nationals living and working in Malta has increased. Therefore, the public administration deals with foreign nationals on a daily basis.

¹<https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/2020-eu-ia-0082>

²<https://languageknowledge.eu/countries/malta>

The national “Novel Coronavirus” web page³ shows an example of the country’s multilingualism needs and challenges, where all information about COVID-19 has been translated to English, whereas other general information has been translated in over 20 languages. The following are some statistics supporting the need for such translations: Top 10 EU Nationals employed between 2002-2018 (Jobsplus, 2019a) and Top 10 Third Country Nationals employed between 2002-2018 (Jobsplus, 2019b).

- National media has a challenge to reach all foreign nationals in Malta especially in the current COVID-19 pandemic, where the main Government communication (mostly press conferences) was generally in Maltese. Therefore, such language barriers (MaltaToday, 2020) should be tackled for inclusion of foreign nationals living in Malta.
- Not all citizen-facing Government websites are available in Maltese and English. An analysis of the eighteen Government Ministries’ (Gov.mt, 2021) websites identified only five of them which cater for both Maltese and English, with the other thirteen being localised in English only.
- Each entity within the public administration carries out their own language translation needs either in-house or through outsourcing to third parties. Most of the requested translations are from Maltese to English and English to Maltese. However, certain Ministries also carry out translations between other languages, such as between English and French, Italian, and German (Berzins et al., 2019).
- When translations are carried out in-house, no Computer Aided Translation (CAT) tools are used, although commercial tools are sometimes used (Berzins et al., 2019).
- No Government policy is available for sharing or pooling of translations and the respective translation memories. This is due to the lack of knowledge on the benefits of data contribution, both at a local and EU level (e.g., for eTranslation⁴) (Berzins et al., 2019).
- Lack of awareness on the importance of language resources and tools (Berzins et al., 2019).

³covid19health.gov.mt

⁴<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>

- Rare re-use of translated data (Berzins et al., 2019) both within the same entity/sector, at a cross-entity/sectoral and/or national level.
- Development of national digital services that are bilingual (Maltese and English) or multilingual are hindered due to the LT gap that exists between high-resourced languages, such as English, and other low-resourced languages, such as Maltese.
- Although it is possible for public entities to access the eTranslation system provided, this requires a specific login with the European portal. This operation might hinder the uptake of such a service.
- Lack of a platform that acts as a one-stop-shop for Maltese language tools and resources for all potential stakeholders, such as public administration, small and medium-sized enterprises (SMEs), researchers, and citizens.

The above are all different language challenges that motivate us further towards building a NLTP for sharing of language resources and tools with citizens, businesses, and the public administration, in particular ones that can cater for and support both the languages of Malta and other languages of foreign nationals living in Malta.

3 National Language Technology Platform

3.1 National Language Technology Platform Overview

The aim of the NLTP is to aggregate state-of-the-art LT tools and solutions in a single entry point. The NLTP will be an AI-driven solution that provides stakeholders with access to mature, tightly-integrated LT services, enabling the efficient and effective multilingual access to information, online services, websites, documents, etc. Stakeholders can include public administrations, businesses, and the general public. The end goals are to remove language barriers, increase accessibility, and foster cross-border services. The latter goals in fact are also priorities for the European Union (EU) and Member States (MS), as made evident by their increasing investment in various programmes intended to advance LT. This promotes democracy and civil society, cross-country cooperation, trade, investment, and European integrity.

Whilst the benefits of language technologies, tools, and services are already apparent, the use of many solutions is unfortunately still limited, po-

tentially due to low discoverability and the specific applicability of such solutions. Therefore, through the matching and integration of existing solutions, including solutions developed in the CEF Telecom - Automated Translation programme, as well as other European and national programmes, the NLTP will provide translation and speech processing services that are easily customised to the specific requirements of each EU MS. More specifically, the adaptation will be carried out around the eTranslation building block supplemented with customised Neural Machine Translation (NMT) systems and speech processing tools, accessible through a unified user interface. The latter high-quality tools will be tailored for particular key domains within the public administrations and the relevant specific language, terminology, and communication style. The following are some of the LT tools and modules that will be available through the NLTP:

- eTranslation Digital Service Infrastructure (DSI);
- custom MT systems for specific domains and languages;
- terminology management services;
- translation memory (TM) accumulation and management functionality;
- widget for website translation;
- CAT environment;
- data repository; and
- speech-to-text and text-to-speech services.

In particular, the data repository shall be used for collecting source data from the stakeholders for further use and processing by the NLTP tools and modules. Furthermore, apart from the above tools and modules, integrated speech technologies will give access to information to people with disabilities such as visual or hearing impairments. The NLTP will therefore be an indispensable platform that enables stakeholders to access the necessary LT tools, thus enabling public administrations to simplify and automate their interactions with their citizens, businesses to target new markets, and citizens to access information and communicate also across borders.

3.2 Maltese Adaptation for the National Language Technology Platform

Malta's NLTP shall act as a central repository for interested stakeholders, including public entities, SMEs, researchers, citizens, and foreign nation-

als living and working in the country. The NLTP will continuously support the curation of Maltese language tools and resources, thus ensuring the presence of Maltese in digital environments. Moreover, the NLTP shall be adapted to the country's specific requirements, with the following being the most relevant ones:

- **Assessment and aggregation of linguistic tools and resources:** The necessary language resources (parallel and monolingual corpora) shall be collected and processed for general domain and domain-specific areas (including legal, culture, social security, health, and taxation) for the Maltese-English-Maltese MT system development. These resources consist of significant amounts of tokens per language –approximately 6.5 million per language for the parallel datasets within the legal domain⁵, and approximately 250 million tokens for the Korpus Malti (Gatt and Čéplö, 2013) Maltese monolingual dataset (containing genres such as the press, literature, laws, parliament debates, and Wikipedia) and 1,000,000 words for an English monolingual dataset.
- **Development of custom MT systems:** A set of Maltese-English-Maltese custom NMT systems shall be developed using the linguistic resources collected and processed as listed above. The plan is to train a general domain NMT system and domain-specific systems based on the type of parallel data collected, language-specific terminology, and communication style used.
- **Multimodal data sources:** Certain Government data collected for the development of NMT systems contains data in multiple modalities, such as text and images, therefore, requiring multimodal data processing. Some data source examples are the Maltese laws (primary and subsidiary legislation)⁶, and the Government services and information (e.g., the Government Gazette and Press Releases) provided by the Department of Information⁷. In the case of the latter, for example, the machine readable version of specific Government Gazettes represent certain data (e.g., tabular) as images.

⁵Quoted figure based on Chapters of the Laws of Malta from <https://legislation.mt/>, such as the Criminal Code, Civil Code, and Companies Act

⁶<https://legislation.mt/>

⁷<https://doi.gov.mt/>

- **Deployment:** MITA's Hybrid Cloud Enabling Infrastructure⁸ is the platform of choice for the digital services of the Government of Malta.
- **National policies and directives:** Any national requirements, in terms of directives, such as Web Accessibility Directive (FITA, 2018) in accordance with the European Standard EN301549v1.1.2 (ETSI, 2016), and policies, such as the Government of Malta ICT Website Policy (MITA, 2019), shall be adhered to.

MITA shall provide the necessary assistance to any Maltese entity (ministries, departments, agencies, etc.) within the public administration who shows interest in integrating certain NLTP modules for further use within their domain, such as websites or internal systems. The “Webpage translation widget” is one of the modules that shall be integrated within citizen-facing Government websites to ensure that they are available in both Maltese and English.

Based on the requirements above, the following are the **main goals** that will define the success of the Malta NLTP:

- Increasing of language tools and resources for the Maltese low-resourced language, in particular as required for MT;
- Building and validating custom NMT systems targeting multiple domains (through pilots);
- Providing MT support and development of custom MMT systems;
- Integration of NLTP with multiple eGovernment services and public web services;
- Enabling and facilitating the addition of other languages through the use of English or Maltese as an intermediary language, in order to provide further translations as required by the Maltese context; and
- Increasing efficiency in the public administration and reducing the expenditure on manual translations through the provision of automated MT services.

3.3 Legal Domain Pilot

The legal domain will be taken as an initial pilot for Malta's NLTP. The particular Maltese context, where both Maltese and English are official languages, provides an interesting basis for this pilot due to the specific requirements of the domain. In

⁸<https://cloud.gov.mt/>

fact, as defined by the Maltese Constitution, Maltese laws need to be enacted in both languages, however in case of any discrepancies between the two languages, the Maltese law prevails⁹.

The drafting and the translation of Maltese legislation generally take place at Ministerial level or at more granular levels, such as specialised authorities, government agencies, and other public entities. When the latter are not in a position to provide a translation, the Legislation Unit within the Office of the State Advocate will carry out the translation itself, as part of its final review process of legislation. Although the Maltese version of the laws prevails over the English version in case of conflict, the vast majority of laws are first drafted in English and then translated into Maltese. This places a lot of responsibility on the translator who is usually further pressured by tight deadlines. For this reason, it is imminent that drafters and/or lawyer-linguists within the Maltese public administration are provided with appropriate language tools that facilitate the translation process and ensure consistency of terms and terminology across ministries and a multitude of domains.

The Laws of Malta comprise language and legal data going back almost 200 years. The data is bilingual, readily available and provides an excellent start for Maltese-English/English-Maltese AI solutions, such as MT in the legal domain due to its vastness and variety of areas it covers. The implementation of this pilot will therefore enable the relevant stakeholders to more easily and efficiently translate drafted legislation, facilitating the legislation enactment process and aiding in the reduction of any discrepancies in the respective texts.

4 Related Work

As a low-resource language, language technologies for Maltese are still lacking when compared to other languages, although there have indeed been developments in the domain in recent years. MT technology, especially MMT technology, is still very limited.

4.1 Current Initiatives: Maltese Language Tools and Resources

This section provides a discussion on current landscape with respect to Maltese language tools and resources. These cater for multiple NLP tasks, however are limited to the textual modality, therefore

⁹<https://legislation.mt/eli/const/eng>

highlighting the need for more research and development focused on other modalities, such as images, videos and spoken language.

The Maltese Language Resource Server (MLRS) (Rosner et al., 2006) currently serves as a central repository for Maltese language resources and tools built by the UM, such as a *tokeniser*, a *part-of-speech* (POS) tagger and *corpora*, such as the Korpus Malti. Ġabra¹⁰ is an open *lexicon* for Maltese which contains several different lexical resources into one common database (Camilleri, 2013).

Borg presented a comprehensive and systematic treatment of Maltese *morphology* using machine learning techniques (Claudia Borg, 2015). The research resulted in a morphological classification system for verbs, nouns and adjectives, which although not sufficiently accurate, it can act as a basis for a more complete morphological analyser with broader coverage (Borg and Gatt, 2014). Ravishankar et al. developed a morphological analyser of Maltese, generated from a combination of affixation and form-generation rules. Evaluated over two corpora, the system fared reasonably well. (Ravishankar et al., 2017).

Tiedemann and van der Plas conducted a number of experiments on *dependency parsing*, in order to evaluate the practicality of cross-lingual learning for Maltese, as a low-resource language (Tiedemann and van der Plas, 2017). Despite encouraging results with model and data transfer, the outcome of the experiments indicate that cross-lingual parsing still lags far behind fully supervised models. Moreover, a new project focused on the development of a Maltese *spelling* and *grammar checker* solution (spellchecker) and owned by the Ministry for the National Heritage the Arts and Local Government, is currently in its initial stages.

The Multilingual Anonymisation toolkit for Public Administrations (MAPA) project is developing a multilingual anonymisation kit based on *Named Entity Recognition and Classification* (NERC) that is applicable for all EU languages, including Maltese (Ajausks et al., 2020). In (Cortis and Davis, 2021), the authors published a dataset that can be used for developing *opinion mining* tools, such as for *sentiment analysis* and language technologies for the Maltese official languages. It consists of user-generated content in English (high-resourced), Maltese-English (code-switched) and Maltese (low-

¹⁰<https://mlrs.research.um.edu.mt/resources/gabra/>

resourced) and has been annotated for multiple opinion dimensions, namely *subjectivity*, *sentiment polarity*, *emotion*, *sarcasm*, and *irony*.

With respect to *speech recognition*, the MASRI (“Maltese Automatic Speech Recognition”) project¹¹ is focused on developing the first Automatic Speech Recognition (ASR) technologies for the Maltese language. To this end, the MASRI-HEADSET (Hernandez Mena et al., 2020), is the first spoken Maltese corpus designed for ASR. Recently, a system for Maltese ASR –bilmalti.com¹²– was launched online with the aim of advancing the speech-to-text technologies. As for *speech synthesis*, a Maltese speech engine (UM, 2020) shall be developed as further development of the Maltese text-to-speech Engine (FITA, 2012).

4.2 Machine Translation for Maltese

Rosner and Bajada tackle an issue in Statistical Machine Translation (SMT) where good quality translation is not strictly word-to-word (Rosner and Bajada, 2007). The authors propose a phrase extraction algorithm and carry out a preliminary evaluation on a set of bilingual texts. The latter indicated that parameterisation of the algorithm would improve the results.

Azzopardi carried out a large-scale evaluation of translation technologies for Maltese translators, including electronic dictionaries, termbases, corpora, spell-checking software, translation memory software, terminology memory software, terminology management software, and machine translation (Brandon Sam Azzopardi, 2015). For the latter, Azzopardi evaluated Google and Bing (the top two MT systems for Maltese) in a number of fields.

Habash et al. present Arab-Acquis; a dataset for evaluating MT between 22 European languages and Arabic (Habash et al., 2017). In the resulting benchmarking results on the 22 languages, the direct translation from English to Maltese scored the highest BLEU (Papineni et al., 2002) score, whilst the translation from Maltese to English scored around the average. For pivoting translation, the score for Arabic-English-Maltese translation is slightly below average, whilst the Maltese-English-Arabic translation scored the highest.

Aharoni et al. perform a large number of experiments to train massively multilingual NMT, involving up to 103 distinct languages and 204 translation

directions simultaneously (Aharoni et al., 2019). Based on their results, the authors report that massively multilingual many-to-many models are effective in low-resource settings, and outperform previous state-of-the-art. The Maltese language was one of 102 languages paired with English, used for these experiments.

5 Envisaged Benefits

The following are the main benefits of Malta’s NLTP:

- Language resources and tools are enablers to key technologies at a multimodal level, in terms of text, visual (video and image), audio, and speech, and to automate certain processes;
- Provide the Government with the required language resources and tools to build bilingual/multilingual digital services and AI tools and technologies;
- Enable more people and businesses to collaborate on LT initiatives, especially in MMTT;
- Ability to integrate and/or use the language resources and tools within key software environments;
- Improved citizen service support at a cross-sectoral level;
- Reachability in terms of different public entities within the public administration; and
- Reduction on costs for translation services.

6 Conclusions and Future Work

Malta’s NLTP shall act as a central repository for Maltese language tools and resources, primarily in the context of automated translation, and act as a one-stop-shop for its stakeholders. Moreover, it shall continuously support the curation of any Maltese language tools and resources for multiple domains, thus ensuring the presence of the currently low-resourced Maltese language in digital environments and countering the threat of “*digital extinction*” of the language. This project is anticipated to leave both a social and economic impact to the Maltese society and economy which benefits the citizens, the public administration, private entities, and academic institutions.

Acknowledgments

This work is funded by the European Commission under Connecting Europe Facility (CEF) - Telecommunications Sector (Agreement No. INEA/CEF/ICT/A2020/2278398).

¹¹<https://www.um.edu.mt/projects/masri/>

¹²<https://www.bilmalti.com/>

References

- Roe Aharoni, Melvin Johnson, and Orhan Firat. 2019. *Massively multilingual neural machine translation*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3874–3884, Minneapolis, Minnesota. Association for Computational Linguistics.
- Eriks Ajausks, Victoria Arranz, Laurent Bié, Aleix Cerdà-i Cucó, Khalid Choukri, Montse Cuadros, Hans Degroote, Amando Estela, Thierry Etchegoyhen, Mercedes García-Martínez, et al. 2020. The multilingual anonymisation toolkit for public administrations (mapa) project. In *Annual Conference of the European Association for Machine Translation*, pages 471–472.
- Aivars Berzins, Khalid Choukri, Maria Giagkou, Andrea Lösch, Helene Mazo, Stelios Piperidis, Mickaël Rigault, Eileen Schnur, Lilli Small, Josef van Genabith, Andrejs Vasiljevs, Andero Adamson, Dimitra Anastasiou, Natassa Avraamides-Haratsi, Núria Bel, Zoltán Bódi, António Branco, Gerhard Budin, Virginijus Daturkevicius, Stijn de Smeytere, Hristina Dobreva, Rickard Domeij, Jane Dunne, Kristine Eide, Claudia Foti, Maria Gavriilidou, Thibault Grouas, Normund Gruzitis, Jan Hajic, Barbara Heinisch, Veronique Hoste, Arne Jönsson, Fryni Kakoyianni-Doa, Sabine Kirchmeier, Svetla Koeva, Lucia Konturová, Jürgen Kotzian, Simon Krek, Gauti Kristmannsson, Kaisamari Kuhmonen, Kristin Lindén, Teresa Lynn, Armands Magone, Hélène Mazo, Maite Melero, Laura Mihailescu, Simonetta Montemagni, Micheál Ó Conaire, Jan Odijk, Maciej Ogrodniczuk, Pavel Pecina, Jon Arild Olsen, Blette Sandford Pedersen, David Perez, Andras Repar, Ayla Rigouts Terryn, Eirikur Rögnvaldsson, Mike Rosner, Nancy Routzouni, Claudia Soria, Alexandra Soska, Donatienne Spiteri, Marko Tadic, Carole Tiberius, Dan Tufis, Andrius Utka, Paolo Vale, Piet van den Berg, Tamás Váradi, Kadri Vare, Andreas Witt, Francois Yvon, Janis Ziedins, and Miroslav Zumrik. 2019. *Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe - Why Language Data Matters: ELRC White Paper*, 2 edition. ELRC Consortium.
- Claudia Borg and Albert Gatt. 2014. *Crowd-sourcing evaluation of automatically acquired, morphologically related word groupings*. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3325–3332, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Brandon Sam Azzopardi. 2015. *An evaluation of the computer-aided translation tools and resources available for Maltese translators*. Ph.D. thesis, University of Malta.
- John J Camilleri. 2013. A computational grammar and lexicon for maltese. Master’s thesis.
- Claudia Borg. 2015. *Morphology in the Maltese language : a computational perspective*. Ph.D. thesis, University of Malta.
- Keith Cortis and Brian Davis. 2021. A dataset of multidimensional and multilingual social opinions for malta’s annual government budget. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 971–981.
- CCEN ETSI. 2016. Accessibility requirements suitable for public procurement of ict products and services in europe. https://www.etsi.org/deliver/etsi_en/301500_301599/301549/01.01.02_60/en_301549v010102p.pdf.
- FITA. 2012. Erdf 114 maltese text to speech synthesis. <https://fitamalta.eu/projects/maltese-speech-engine-synthesis-erdf-114/>. ERDF 114 Maltese Text to Speech Synthesis.
- FITA. 2018. Web accessibility directive. <https://fitamalta.eu/web-accessibility-directive-update/>. Web Accessibility Directive.
- R Alonso García. 2002. The general provisions of the charter of fundamental rights of the european union. *European Law Journal*, 8(4):492–514.
- Albert Gatt and Slavomír Čěplö. 2013. Digital corpora and other electronic resources for maltese. *Lancaster, UK*.
- Gov.mt. 2021. Government of malta ministries and entities. <https://www.gov.mt/en/Government/Government%20of%20Malta/Ministries%20and%20Entities/Pages/default.aspx>.
- Nizar Habash, Nasser Zalmout, Dima Taji, Hieu Hoang, and Maverick Alzate. 2017. *A parallel corpus for evaluating machine translation between Arabic and European languages*. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 235–241, Valencia, Spain. Association for Computational Linguistics.
- Carlos Daniel Hernandez Mena, Albert Gatt, Andrea DeMarco, Claudia Borg, Lonke van der Plas, Amanda Muscat, and Ian Padovani. 2020. *MASRI-HEADSET: A Maltese corpus for speech recognition*. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6381–6388, Marseille, France. European Language Resources Association.
- Jobsplus. 2019a. Top 10 nationalities of employed eu nationals (2009-2019). <https://jobsplus.gov.mt/resources/publication-statistics-mt-mt-en-gb/labour-market-information/foreigners-data/#title2.1>. Top 10 Nationalities of Employed EU Nationals (2009-2019).

- Jobsplus. 2019b. Top 10 nationalities of employed third country nationals (2009-2019). <https://jobsplus.gov.mt/resources/publication-statistics-mt-mt-en-gb/labour-market-information/foreigners-data/#title3.1>. Top 10 Nationalities of Employed Third Country Nationals (2009-2019).
- MaltaToday. 2020. Employees cannot refuse to work if outlets have re-opened, minister says. https://www.maltatoday.com.mt/news/national/102136/watch_employees_cannot_refuse_to_work_if_outlets_have_reopened_minister_says/#.YPAJRugza72.
- MITA. 2019. Gmict - website policy. https://mita.gov.mt/wp-content/uploads/2020/07/GMICT_P_0051_Website.pdf. GMICT - Website Policy.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Vinit Ravishankar, Francis M. Tyers, and Albert Gatt. 2017. A morphological analyser for maltese. *Procedia Computer Science*, 117:175–182. Arabic Computational Linguistics.
- Georg Rehm, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Victoria Arranz, Andrejs Vasiljevs, Gerhard Backfried, José Manuel Gómez-Pérez, Ulrich Germann, Rémi Calizzano, et al. 2021. European language grid: A joint platform for the european language technology community. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 221–230.
- Mike Rosner and Jo-Ann Bajada. 2007. Phrase extraction for machine translation.
- Mike Rosner, Ray Fabri, Duncan Attard, and Albert Gatt. 2006. Mlrs: a resource server for the maltese language. <https://mlrs.research.um.edu.mt/>.
- Mike Rosner and Jan Joachimsen. 2012. II-Lingwa Maltija FI-Era Digitali – The Maltese Language in the digital age. *META-NET White Paper Series*. Springer, Heidelberg etc. <http://www.meta-net.eu/whitepapers/volumes/maltese>.
- Silvio Schembri. 2019. Malta: The Ultimate AI Launchpad - A Strategy and Vision for Artificial Intelligence in Malta 2030. https://malta.ai/wp-content/uploads/2019/10/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf. Parliamentary Secretariat for Financial Services, Digital Economy and Innovation, Office of the Prime Minister.
- Jörg Tiedemann and Lonneke van der Plas. 2017. *Bootstrapping a Dependency Parser for Maltese - A Real-World Test Case*, number 32 in Tributes, pages 355–365. College publications, United Kingdom.
- UM. 2020. Um to further develop speech engine for the visually-impaired. <https://www.um.edu.mt/newspoint/news/features/2020/02/maltese-speech-engine>.