



# Chapter 27

## Language Report Maltese

Michael Rosner and Claudia Borg

**Abstract** This chapter is a highly abbreviated version of an update (Rosner and C. Borg 2022) to the META-NET White Paper on Maltese (Rosner and Joachimsen 2012). Like its predecessor, the update forms part of a series for all European Languages. Section 1 provides a brief description of the language, its national status, its general typology as a language, and its current usage in the digital sphere. Section 2 gives an overview of technologies and resources that are currently available. Finally, Section 3 frames the main shortcomings of Maltese language technology in terms of fragmentation, and offers some recommendations on how that might be reduced.

### 1 The Maltese Language

Maltese (il-Malti) is an official EU language and the national language of the Maltese archipelago. 97% of the Maltese population (ca. 400,000 people) consider it their mother tongue. It is also spoken by communities in Australia, Canada, the USA and the UK. Maltese is derived from late medieval Sicilian Arabic with Romance superstrata, and is often referred to as a mixed language due to the large number of loan words from Italian, English and French. It shares characteristics with other Semitic languages, making use of root-and-template morphology whereby various forms of the same lexeme are formed by interdigitating vowels between a fixed sequence of root consonants. The main distinguishing characteristics of Maltese are free word order, mixed morphology, aspect-based temporal system, and lack of a morphological infinitive. Unlike other Semitic languages, the Maltese alphabet is based on the Latin one with the addition of some letters with diacritic marks and digraphs (ċ, ġħ, ż, ġ, ħ). It contains 24 consonants and 6 vowels. According to Fabri (2011), the writing systems used for Maltese were somewhat ad hoc before 1920, but a degree of consistency among writers and in publications became a reality in the 1950s.

Within the digital sphere, there have always been several Maltese language newspapers. The broadcast media (radio and TV) are almost exclusively in Maltese. Since

---

Michael Rosner · Claudia Borg  
University of Malta, Malta, [mike.rosner@um.edu.mt](mailto:mike.rosner@um.edu.mt), [claudia.borg@um.edu.mt](mailto:claudia.borg@um.edu.mt)

the previous report, there has been a general decline in hard-copy newspaper readership, as all the media are now available online and the majority of readers prefer the online version. Various online-only news websites have appeared, one of which (Newsbook) operates bilingually. The full Maltese character set is now universally used. Social media are extremely popular (97% of the population according to a 2021 survey). Facebook remains the most accessed, but there is a trend of increased usage of Instagram and YouTube. Unlike other EU countries, Twitter usage in Malta is remarkably low. The Maltese Wikipedia currently ranks at 204/325 (for comparison, English, Portuguese, Irish, Icelandic, Romansch rank at 1, 18, 93, 95, and 213, respectively). It contains nearly 4 million words distributed over 4,400 content pages (cf. 6.5 million for English). This compares to about 3,000 pages in 2011; there are ca. 19,000 registered users with only about 40 active users (making changes every 30 days or less). YouTube gives rise to localised content in many other countries but the local website still operates predominantly in English. In general, there tends to be a gap between social media content creators and non-creators. However, a renowned online page which has successfully bucked this trend is Kelma Kelma which started in 2013 as a Facebook page and gathers many interesting original contributions by locals about the Maltese language. The top-level country domain for Malta, .mt, is administered by the Malta Internet Foundation, has currently ca. 17,000 domain names and subdomains, more than three times the figure in 2010.

## 2 Technologies and Resources for Maltese

Rosner and Joachimsen (2012) describe the main enablers and contributions to Maltese Language Technology up to ca. 2011. 2012 marked the public release of the MSE speech synthesiser (M. Borg et al. 2014), whilst Gatt and colleagues began re-vamping the University's MLRS resource server (Rosner 2008; Gatt and Čéplö 2013) to include semi-automated data-collection, a tagger, Korpus Malti v3.0 (2016), containing ca. 250 million annotated tokens, pattern-based search facilities, CLEM, a 1 million token Corpus of Learner English in Malta, Ġabra, an Open Lexicon for Maltese, and a Dictionary of Maltese Sign Language.

Most available corpora are monolingual written text. A few are spoken, and fewer still are multimodal such as MAMCO (Paggio et al. 2018). Many monolingual corpora form part of unannotated *multilingual* collections. Others are by-products of projects and annotated for MWE identification (PARSEME) or POS Tagging (MLRS), anonymisation (MAPA), morphological analysis (UniMorph), NER (WikiAnn) etc. Bilingual/multilingual resources include the Laws of Malta, the Government Gazette, and the Acquis Communautaire.

Regarding tools and services, besides low-level text preprocessing for tokenisation, sentence and paragraph splitting and POS-tagging, the Ġabra dictionary has evolved into the online Dizzjunarju tal-Malti app. Machine translation for Maltese has improved not only through the availability of free tools like Google Translate, but also as a result of DGT's eTranslation platform whose increased takeup by pub-

lic administration officials followed a series of workshops organised through ELRC. Much recent effort has been focused on dependency parsing and ASR. There is now a 2000-sentence Universal Dependency Treebank for Maltese which has supported experiments (Zammit et al. 2019) aimed at delivering a prototype dependency parser in 2022. Similarly, for speech technology, the locally funded MASRI project has delivered a fully annotated speech corpus (Hernandez Mena et al. 2020). Most resources mentioned above are freely available through MLRS and also EU platforms.

Currently, the main drivers for the evolution of future Maltese LT are targeted national initiatives, against a mixed background of projects at EU level. At the national level, the National AI Strategy (2019) focuses on the creation of an AI ecosystem infrastructure including tools to enable Maltese Language AI solutions, with funds earmarked for Maltese LT resources. The Malta Digital Innovation Authority (MDIA) is committed to supporting Maltese LT tools which will focus on morphological analysis, dependency parsing, named entity recognition and POS tagging. In 2019, the Government also committed funds to the development of a spell checker. However, there is no information with respect to the progress of this important initiative. Meanwhile at the EU level Maltese participation in a wide range of projects, actions and initiatives including ELE, ELG, ELRC, DARIAH, LCT, LT-Bridge, MAPA, Nexus Linguarum, and NLTP, has ensured a level of Maltese presence on the European scene and also produced some specialised resources and tools.

### 3 Recommendations and Next Steps

Maltese LT is indeed alive, but manifests an important weakness: it is highly fragmented, in different ways: 1. between national efforts (small-scale, Maltese-focused) and international ones (large-scale, language-independent); 2. across resources/tools which are not necessarily compatible with each other; and 3. between users and developers of LTs (reduces the perceived relevance of the technologies developed). To address these requires further investigation of techniques like transfer learning, as seen, for example, in the MAPA project where general language models were successfully used for Maltese NER. Issue 2. can be reduced by insisting that such resources inhabit a framework which includes the necessary protocols to ensure interoperability, as seen in European infrastructures like ELG and NLTP, funded under CEF, aiming to build a National Language Platform for Maltese integrating eTranslation services developed by the European Parliament with fine-tuned local translation memories, and providing a central point for collecting different LT services together. 3. is in part the result of insufficient involvement of the IT industry in LT. Despite the latter being a major component of the local economy, the number of technical LT providers is very low. LT has a crucial role to play as a natural bridge linking IT, AI, communication and multilinguality. More needs to be done to support that role by encouraging participation in ELG by local IT players, among others. In 2016, the IT subcommittee of the Council for the Maltese Language had recognised the need for the long-term curation of resources, recommending the creation of a central

repository, and efforts to involve more stakeholders concerning the availability and importance of resources. Some progress towards the realisation of these recommendations has been made but the effort needs a substantial and sustained coordinated investment across the different sectors involved.

## References

- Borg, Mark, Keith Bugeja, Colin Vella, Gordon Mangion, and Carmel Gafa (2014). “Preparation of a Free-Running Text Corpus for Maltese Concatenative Speech Synthesis”. In: *Perspectives on Maltese Linguistics, Studia Typologica 14*. Ed. by Albert Borg, Sandro Caruana, and Alexandra Vella, pp. 297–318.
- Fabri, Ray (2011). “Maltese”. In: *The Languages of the 25. Revue belge de Philologie et d’Histoire: RBPH*. Ed. by Christian Delcourt and Piet van Sterkenburg. Amsterdam, Philadelphia: John Benjamins, pp. 17–28.
- Gatt, Albert and Slavomír Čéplö (2013). “Digital corpora and other electronic resources for Maltese”. In: *Proceedings of Corpus Linguistics*. Ed. by Andrew Hardie and Robbie Love. University of Lancaster, UCREL.
- Hernandez Mena, Carlos Daniel, Albert Gatt, Andrea DeMarco, Claudia Borg, Lonneke van der Plas, Amanda Muscat, and Ian Padovani (2020). “MASRI-HEADSET: A Maltese Corpus for Speech Recognition”. In: *Proceedings of LREC 2020*. Marseille, France: ELRA, pp. 6381–6388.
- Paggio, Patrizia, Luke Galea, and Alexandra Vella (2018). *Prosodic and gestural marking of complement fronting in Maltese*. DOI: [10.5281/zenodo.1181805](https://doi.org/10.5281/zenodo.1181805).
- Rosner, Mike (2008). “Electronic Language Resources for Maltese”. In: *Proceedings of Bremen Workshop on Maltese Linguistics*. Springer.
- Rosner, Mike and Claudia Borg (2022). *Deliverable D1.25 Report on the Maltese Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. <https://european-language-equality.eu/reports/language-report-maltese.pdf>.
- Rosner, Mike and Jan Joachimsen (2012). *Il-Lingwa Maltija Fl-Era Digitali – The Maltese Language in the Digital Age*. META-NET White Paper Series: Europe’s Languages in the Digital Age. Heidelberg etc.: Springer. <http://www.meta-net.eu/whitepapers/volumes/maltese>.
- Zammit, Andrei, Slavomír Čéplö, Lonneke van der Plas, and Claudia Borg (2019). *A Dependency Parser for Maltese: Comparing the impact of transfer learning from Romance and Semitic Languages*.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

