

# Joint Masters Degree in Digital Linguistics

Programme duration:

2 years

ECTS:

120

## Programme Description

### 1 Background

Digital Linguistics is a new interdisciplinary field of study at the crossroads between linguistics, translation studies, information technologies and social sciences. The digital age has brought profound changes to the ways texts and other types of content are produced, processed (by humans or machines), distributed and reused. Traditional graduate profiles, such as those provided by studies in Linguistics, Translation, Information Technologies or Communication Studies, fail to equip students with the skills and competencies required for the provision of language- and communication-related services in the digital world.

Digital Linguistics is not synonymous to Computational Linguistics (CL), although certain skills and research methods may overlap between these two disciplines. While CL is primarily concerned with modelling and processing language in computer applications, mainly by means of fixed rules or coarse-grained statistics, in order to provide language-aware tools and systems, the emerging field of Digital Linguistics is broader in the sense that it should provide the complete set of scientific, methodological and practical foundations pertaining to communication in the digital age and that it should enable students with backgrounds from various linguistic and translational subdisciplines to apply their knowledge and skills within a digital context.

This subsumes linguistic knowledge, such as native and foreign language competence, translation-related competencies and an understanding of language analysis procedures at all levels, and it also entails Natural Language Processing skills, particularly at the level of an in-depth understanding of state-of-the-art NLP techniques and basic programming skills. But on top of this “traditional” interdisciplinary blend between linguistics and IT, Digital Linguistics has several additional foci which justify its claim for a field of study in its own right.

The first is digital content authoring, which could at first glance be understood as one of the linguistic competencies formerly known as text production skills developed through text and discourse studies. Digital news media have revolutionized journalism and brought new paradigms into the concepts of journalistic research, credibility, authenticity, authorship and accessibility. The personalization of digital services means that content, including web sites, ads, user manuals and posts on social platforms, is produced in a targeted and user-centred fashion, whereby the cyber-identity of the target user is not to be confused with their real-world identity.

The second is that of super-fast multilingual digital communication. Machine translation and machine interpreting services have blossomed, yet their quality has been more than questionable. Still, they are being used around the globe and not rarely, with success. While the exact nature of the interplay between translation quality, communication risks and communicative flexibility has not yet been fully understood, it is clear that students need a deeper understanding of which tools can be used in which communicative contexts, which goals can be met by means of using which technology and how creators of such technologies need to be informed in order to improve the underlying technology.

Issues such as those named above can only be adequately addressed by bringing in the sociological, psychological and cognitive perspectives, and by putting communicative behaviour in digital media into the centre of study. Another aspect of content authoring is related to multilingual contents and activities such as translation, localization, subtitling and interpreting. While traditionally the providers of multilingual services were the ones generating content, contemporary translators compose texts by selecting from available hits offered by translation memories, machine translation engines and other multilingual resources. From the cognitive point of view, the process of [content] generation has been transformed into the process of selection, where the issues of critical assessment and trust have become paramount.

The issues of trust, identity, authorship and reuse inevitably lead to questions concerning intellectual property rights and data protection, but also ethical aspects of communication in digital media. As future managers of language resources Digital Linguists must be familiar with the national and international legal framework governing for example the compilation, protection and distribution of corpora, training data for Machine Translation, translation memories and similar.

Therefore, Digital Linguistics as a field of study combines insights and perspectives from different disciplines and does not overlap with Computational Linguistics, Information Science, Digital Humanities, Translation Studies, Sociolinguistics or Corpus Linguistics, though it may inherit methods and tools from all of the above.

## 2 Labour market needs

In Europe and worldwide, the language industry is growing rapidly as the world becomes increasingly globalized and multilingual, however the position of traditional linguistic profiles, such as that of translator, interpreter or linguist, is losing ground in the job market. Instead we encounter new profiles such as that of multilingual personalization specialist, junior language technician, global content provider, cross-language data consultant, expert in multilingual social network mining, localization project manager, speech analyst etc.

In order to assess the needs of the local and global economy for digital linguists and to adequately identify the skills and competencies needed by future graduates, two surveys were conducted.

The first was a small online survey among language service providers in Slovenia carried out in 2015, collecting responses from 20 largest companies providing translation, interpreting, software localization and other language services (results under <https://www.surveymonkey.com/results/SM-39FQZNTW/>). 90% of respondents believe that graduates of currently running HE programmes lack technological skills and that the need for digitally better educated young adults will continue to grow. 75% have to invest their own company resources to provide those skills and competencies, either through inhouse or external training.

The second was a large online survey carried out within the framework of the EU-funded Erasmus+ project DigiLing ([www.digiling.eu](http://www.digiling.eu)), which serves as the ground-preparing platform of our joint programme. The survey was aimed at companies outside of the language industry and inquired about their practices and needs in the field of processing, analysing, translating language data, as well as in the field of web content creation and multimodal interfaces. We collected 90 responses from 11 countries and the results indicate predominantly needs for automatic sorting, analysis and categorization of e-mail communication and internal documents, while the need to automatically analyze social media or competitor's websites is currently less pronounced. Another discernible trend is increasing multilinguality in the enterprise world. In every relevant category (content, categorization, attitude analysis, terminology, automated communication and speech, digital agents), the need for automatic creation or processing of multilingual content was expressed by more than half the respondents, with terminology consistency checks figuring the highest at 75 percent. The results of both surveys also indicate that respondents are either in the process of or planning expansion of their activities in the area of digital language processing.

These results are in line with EU-wide trends. Within the EU, the demand for digitally skilled employees is growing by around 4% a year. Shortages of ICT professionals in the EU could reach 825,000 unfilled vacancies by 2020 if no decisive action is taken (DSMS 2015). At the same time, the European language industry was estimated at 16.5 billion Euros in 2015 and is expected to continue growing by 10% per year (META-NET SRA 2012, TAUS 2015). The global speech technology market grows even faster and will reach 31.3 billion US-Dollars by 2017, yet most commercial services such as Nuance's speech tools or Apple's Siri are provided by US-based enterprises. Multilinguality remains one of the key challenges in Europe. Our continent has half a billion citizens who speak one of over 60 European and many non-European languages as their mother tongue, yet 21 of these languages are facing digital extinction. With migrant inflows from crisis areas, the need for language services to facilitate communication, inclusion and integration is greater than ever before. Human translators are not in the capacity to satisfy current and future needs - already today, Google Translate translates the same volume per day that all human translators on the planet translate in one year. Yet according to recent surveys, for 23 out of 30 studied languages Machine Translation was evaluated as having very limited quality and performance. In addition to these facts, today's digital world is increasingly shaped by social media, where language is used in entirely different ways, which have scarcely been explored and are largely unsupported by current technologies.

### 3 Target skills and competencies

A Masters graduate in Digital Linguistics has the following skills and competencies:

- Language competence in at least two languages,
- An understanding of the way written and spoken language works at all levels of linguistic analysis,
- An understanding of the principles of multilingual communication, including skills in intercultural mediation, translation/interpreting/localization and multilingual content authoring,
- Skills in the compilation of digital language resources, such as corpora, lexica, acoustic databases and similar, including competencies in methodological design and technical implementation of LR compilation,
- Skills in analysing and processing natural language, including the ability to design and develop own tools as well as basic competencies in implementing existing ones in order to analyse or process language data,
- Basic understanding of digital media from the sociological, psychological and legal perspective,
- Ability to perform independent research and acquire new skills,
- Ability to work in interdisciplinary/multilingual teams.

The diagram below illustrates the network of target skills and competencies.



### 3 Curriculum structure

The Digital Linguistics curriculum is structured as follows:

The entire curriculum is composed along three complementary pillars: Foundations, providing theoretical principles in each of the contributing fields: Linguistics, Multilingual Communication, Programming & IT and Digital Media. Since students will enter the programme with different Bachelors degrees and hence different backgrounds, the main purpose of this module is to equalize the differences and provide a solid interdisciplinary basis for all in accordance with their background. Modules can be adjusted such that e.g. students with a Bachelor's degree in Computer Science will not need to attend classes on basics of programming etc.

The second pillar consists of Methods and Tools where students will acquire applied skills in selected areas of Digital Linguistics, including text analysis, digital content authoring, statistics, ethics & law and related. Some parts of this module will be obligatory, other elective to broaden the scope of study. The main purpose of this module is to equip students with an inventory of methods and tools which will allow them to engage in research and applied projects developing own solutions to specific language-related problems.

The third pillar consists of applying the knowledge and skills to concrete problems, either through internship at partner companies/institutions or through projects developing the students' problem-solving and teamwork skills.

