

Tamás Váradi and META-NET

Sabine Kirchmeier¹

¹ European Federation of National Institutions for Language
sabine.kirchmeier@gmail.com

1. Introduction

In his longstanding career, Tamás Váradi has participated in numerous projects on the European scene, but probably the one with the greatest impact on the development of language technology in Europe is META-NET. META is an acronym for Multilingual Europe Technology Alliance, and the network aims at bringing together researchers, commercial technology providers, private and corporate language technology users, language professionals and other information society stakeholders. META-NET is a joint effort towards furthering language technologies in Europe.

2. The key components of META-NET

META-NET started as a network of excellence in 2010 pursuing the following goals:

1. fostering a dynamic and influential community around a shared vision and strategic research agenda (META-VISION),
2. creating an open distributed facility for the sharing and exchange of resources (META-SHARE),
3. building bridges to relevant neighbouring technology fields (META-RESEARCH)¹

The network consisted of 60 members in 34 European countries, one of these the Research Institute for Linguistics at the Hungarian Academy of Sciences represented by Tamás Váradi.

2.1. META-VISION

META-VISION comprised two main activities: 1. addressing European decision makers through the development and promotion of a strategic

¹ <http://www.meta-net.eu/>

research agenda (Rehm and Uszkoreit, 2013), and 2. conducting a large and comprehensive study on 30 European languages and the level of support they receive through language technologies, the META-NET White Papers (Rehm and Uszkoreit, 2012). The two initiatives together turned out to provide an excellent basis for creating an understanding and gaining support among decision makers for the necessity of developing better language technology for all European languages.

2.1.1. The strategic research agenda

The strategic research agenda, published in 2013, set out to describe what kind of technological innovations could be expected by 2020 and what role language technology could play as part of these innovations. Robotics and AI were quite optimistically pointed out as fields expected to profit immensely from language technology:

“Within this decade, specialised mobile robots will be deployed for personal services, rescue missions, household chores, and tasks of guarding and surveillance. Natural language is by far the best communication medium for natural human-robot interaction. By 2020 we will have robots around us that can communicate with us in human language, but their user friendliness and acceptance will largely depend on progress in LT research in the coming years” (Rehm and Uszkoreit, 2013: 35).

2.1.2 The META-NET White Paper Series

The META-NET White Papers (Rehm and Uszkoreit, 2012) describe the technological status of 30 European languages and the level of language technology support available for each of them. 200 experts from all over Europe, the META-NET Network of Excellence, contributed to this comprehensive study that discusses the threats and opportunities for the languages in question.

The key results and the cross-language comparison of the collected data sent a clear message that for most languages, except for English, it was extremely urgent that efforts were made to bring them up to a level where they could be preserved from digital extinction. The Hungarian language was among the languages in the danger zone with only fragmentary support for machine translation, speech processing and text analysis. For text and speech resources, Hungarian, together with Czech,

Dutch, French, German, Italian, Polish, Spanish and Swedish, was reported to have moderate support.

The analyses of the different languages served as an excellent starting point for public initiatives to create better language technology support for many languages.

2.2. META-SHARE

The collection and sharing of language resources is still today at the core of developing language technology. The demand for more and better text and speech resources has grown drastically, especially due to the use of AI-techniques during the last decade. The second META-NET initiative, META-SHARE, was established to address this problem by facilitating the sharing of resources, whereas the collection of resources was mainly the responsibility of national programmes.

2.3. META-RESEARCH

Finally, META-RESEARCH established several working groups, research workshops and a collection of online tutorials, mainly on machine learning for MT, but also other topics to support the development of language technology expertise for the languages involved.

3. META-NET and CESAR

Soon after its formation, META-NET succeeded in interlinking four EC-funded projects, resulting in a Europe-wide cooperation of computational linguistic and NLP communities. One of the projects involved was CESAR (Central and South-East European resources) (Váradi, 2013). It was funded by EC and national funding sources. The project started on 1st February 2011, its duration was 24 months, and the coordinator was Tamás Váradi.

The central objective of the project was to produce and make available a comprehensive set of language resources and tools covering Bulgarian, Croatian, Hungarian, Polish, Serbian and Slovak. CESAR was not only about the creation of new resources but also about enhancing existing resources and tools, for instance their size, coverage, accuracy, compliance with current standards for interoperability, and regarding licencing and IPR issues.

A huge effort was made to make the linguistic development environment NOOJ freely available on all platforms allowing linguists to formalize several levels of linguistic phenomena: typography and spelling; lexicons of simple words, multiword units and discontinuous expressions; inflectional, derivational and productive morphology; local and structural syntax, transformational and semantic analysis and generation.²

By the end of the project, the resources and tools developed by CESAR were made available through META-SHARE, thus contributing to the extension of the linguistic coverage of the platform and ensuring the availability of key resources for the development of improved language technology and AI applications for Central and South East European languages.

4. The impact of META-NET on European language technology

META-NET and its associated projects came to play an extremely important role for the development of language technology for European languages for several reasons. First, META-NET provided the basis for a strong European language technology community that worked together instead of competing with one another. Second, META-SHARE stimulated the collection and exchange of language resources for commercial use, in contrast to collections in other networks, such as CLARIN, which were more focussed on resources for research purposes. Finally, the META-NET White Paper Series gave the participating language communities a well-documented offset for the discussion about the future of languages and language technology all over Europe (Rehm et al., 2014).

4.1. A European LT community

The META-network extended constantly, not least through a long series of events and conferences (META-FORUM etc.)³ and through the inclusion of new stakeholders, individual researchers, companies, and organisations such as the European Federation of National Institutions for Language (EFNIL) and the Network to Promote Linguistic Diversity (NPLD). In this way, META-NET was able to engage and include stakeholders in all European countries and to present itself to the European Commission as a strong language technology community with an

² <http://www.nooj-association.org/>

³ <http://www.meta-net.eu/events>

impressive network of supporters consisting of private vendors, public institutions, and researchers all over Europe.

Over the years, the network continued to grow as its members participated in follow-up projects such as the EU-project CRACKER (2015–2017).⁴ CRACKER’s objectives were, among others, preparing and publishing research and innovation agendas (Rehm, 2015). It managed to establish the Cracking the Language Barrier federation,⁵ a kind of umbrella initiative for European language technology projects and organisations. The cooperation also formed the basis of the European Language Grid.

Many of the META-NET members also participate in EU’s European Language Resource Coordination project (ELRC) aiming at collecting language resources and providing a platform for sustainable language data sharing to support language equality in multilingual Europe, especially the Digital Single Market.⁶

The network created through META-NET also made it possible to conduct a survey (Rehm and Hegele, 2018), which covered more than 600 respondents from more than 50 countries working on language technology, emphasising the need of a programme specifically designed to develop the technology that could meet the linguistic challenges in Europe.

Many members of the network are involved in the latest European language technology project, European Language Equality (ELE), with the goal to establish a roadmap for the development of sustainable language technology for all European languages by 2030.⁷

4.2. Resources in META-SHARE

In 2012, there were 1248 – in 2020, there were 2888 language resources, tools, or services accessible through META-SHARE distributed over 100+ languages, four main resource types (corpus, lexical/conceptual model, tool/service, language description) and four main media types (text, audio, image, video). Currently, the most frequently viewed and downloaded datasets are those containing semantic annotations and gold standards. This clearly indicates the direction that language technology is taking towards becoming an integrated part of the development of artificial intelligence applications.

⁴ <http://www.cracker-project.eu>

⁵ <http://www.cracking-the-language-barrier.eu>

⁶ <https://lr-coordination.eu>

⁷ EU-CALL: Developing a strategic research, innovation and implementation agenda and a roadmap for achieving full digital language equality in Europe by 2030 (PPPA-LANGEQ-2020)

4.3. Political attention

The cross-comparison of the digital fitness of the participating languages regarding text analysis, speech processing, MT and language resources in the META-NET white papers clearly made an impression on the government in many countries and has led to the development of strategies and plans for the advancement of high-quality language technology, for instance in Denmark, Finland, Iceland, Latvia, Norway and Sweden.

In the EU, it led to a hearing initiated by the Scientific Foresight Unit (STOA) of the European Parliament in 2017 (STOA, 2017), who also commissioned the study *Language Equality in the digital age – Towards a Human Language Project*, published in March 2017 and presented to the European Parliament on 11 Sept. 2018.⁸ The EP adopted the report, with an overwhelming majority of 592 votes in favour, 45 against, and 44 abstentions.

The newly initiated ELE Project (2021–2022) is the first step towards implementing the vision of language equality laid out in the STOA-report. It is envisaged to be a multidisciplinary initiative including stakeholders from research institutions, industry, the public sector and civil society, collaborating on European, national and regional level. The primary goal is the preparation of the European Language Equality Programme, specified in the form of a strategic research, innovation and implementation agenda and a roadmap for achieving full digital language equality in Europe by 2030.

With his long-standing experience from not least META-NET and CESAR and his strong and untiring dedication to language technology, Tamás Váradi is of course also involved in the ELE project laying out the path for European language technology in the future. In fact, he is involved both in his capacity as head of his institute and as general secretary of EFNIL as well. There is no doubt that his dedicated engagement in language technology through the years is of immense importance for the digital future of the Hungarian language.

References

⁸ Language equality in the digital age (A8-0228/2018, P8_TA-PROV(2018)0332)

- Rehm, G., Uszkoreit, H. (eds.) META-NET White Paper Series: Europe's Languages in the Digital Age. Springer, Heidelberg/New York/Dordrecht/London (2012) www.meta-net.eu/whitepapers. [31 volumes on 30 European languages.]
- Rehm, G., Uszkoreit, H. (eds.) META-NET Strategic Research Agenda for Multilingual Europe 2020. Presented by the META Technology Council. Springer, Heidelberg, New York etc. (2013) <http://www.meta-net.eu/sra>
- Rehm, G., Uszkoreit, H., Dagan, I., Goetcherian, V., Dogan, M. U., Mermer, C., Váradi, T., Kirchmeier-Andersen, S., Stickel, G., Jones, M. P., Oeter, S., and Gramstad, S.: An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age". In Laurette Pretorius, et al. (eds.) Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era (CCURL 2014), pp. 30–37. Reykjavik, Iceland (2014)
- Rehm, G.: Cracking the Language Barrier for a Multilingual Europe. In: Nuolijärvi, P., Stickel, G. (eds.) Language Use in Public Administration – Theory and practice in the European states. Contributions to the Annual Conference 2015 of EFNIL in Helsinki. pp. 41–58. European Federation of National Institutions for Language, Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary (2015)
- Rehm, G., Hegele, S.: Language Technology for Multilingual Europe: An Analysis of a Large-Scale Survey regarding Challenges, Demands, Gaps and Needs. In: Calzolari, N. et al. (eds) Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018), Miyazaki, Japan. pp. 3282–3289 (2018)
- Váradi, T.: Veni, Vidi, Vici: The Language Technology Infrastructure Landscape after CESAR. In: Gajdošová, K., Žáková, A. (eds.) Natural Language Processing, Corpus Linguistics, E-learning. Seventh International Conference Bratislava, Slovakia, 13–15 November 2013 Proceedings, pp. 261–279 RAM-Verlag, Lüdenscheid (2013)