NexusLinguarum:
Where we are, where we go

Jorge Gracia

The 'European network for Web-centred linguistic data science' (*NexusLinguarum* in its short name) is a COST Action that was launched on 28 October 2019, for a four-year duration. At that time, the initial network was constituted by representatives of 33 different countries from different areas, like computer science, semantic web, artificial intelligence, linguistics, humanities, etc. The main aim of the network is to support the construction of an *ecosystem of multilingual and semantically interoperable linguistic data at the scale of the Web*. To this end, methods and techniques of the Semantic Web, Natural Language Processing and Language Resources are studied and combined. Such an ecosystem could reduce language barriers in Europe (and eventually beyond) and favour both electronic commerce and cultural exchange between countries with different languages.

In a previous article last year, we described the main aims of NexusLinguarum as well as their organisation in working groups and its initial developments. In the following paragraphs, we give a brief update on its current status and next steps.

Since its constitution, the Action has accepted new representatives from more countries, counting today researchers from 42 different countries. The number of members that actively participate in the Working Groups (WGs) are steadily growing, with 170 registered participants in one or several WGs. The Action is still open to welcome new members.

Unfortunately, as in many other aspects in our professional and personal lives, the COVID-19 crisis had an impact on NexusLinguarum activities. A number of Short Term Scientific Missions (STSMs), that is, research visits hosted by members of the network, had to be postponed or cancelled due to travelling restrictions. Similarly, some of the Action's events were postponed and/or changed to an online or hybrid format. For instance, the second NexusLinguarum plenary meeting took place in Lisbon, at the Universidade Nova de Lisboa, on 26-27 October 2020 in a hybrid setting, with most participants participating online.

In addition, the pandemic obliged us to postpone our first training school on Introduction to Linked Data for Linguistics, initially

Jorge Gracia is Chair of NexusLinguarum 'European network for Web-centred linguistic data science' COST Action. He works as senior research fellow at the Department of Computer Science and Systems Engineering (University of Zaragoza, Spain) as a member of the Aragon Institute of Engineering Research (I3A) and of the Distributed Information Systems research group. His main research interests are Semantic Web, Ontology Matching, Multilingual Web of Data, Query Interpretation, and Web Intelligence, and his recent work focuses on linked data-based lexicography as well as on methods and techniques for crosslingual linking and crosslingual information access. http://jogracia.url.ph/web/

planned as a face-to-face event, which finally took place in a purely online mode in February 2021, hosted by the Romanian Academy and University of Iași in Romania. Twelve lecturers and over 80 participants followed the school, which combined theoretical presentations with hands-on sessions. The slides and the training materials are available online. This initial experience will be continued with more training schools in the following years.

Recently, the Action published a policy brief Towards an open ecosystem of multilingual interoperable linguistic data, to help raise awareness among policy makers, stakeholders, and the general public about the social and technological interest of Linguistic Data Science, as a means to overcome language barriers in Europe and worldwide.

It is also worth noting that NexusLinguarum joined the interdisciplinary network COST Actions against COVID-19 , contributing from the field of linguistics. This is an initiative of several Actions wishing to connect and collaborate in treating COVID-19 issues from different angles, offering substantial potential in mobilising experts and tackling challenges as they arise, for this and future pandemics.

Another important outcome of this initial period has been the release of a document on Use Case Description and Requirements Elicitation, where a number of use cases and applications were described, in which the Action's methodologies and technologies can be tested and validated. An extensive summary of this document is available in this issue. The list of current use cases can serve to illustrate the broad range of application domains for the Action's topics:

- Media and Social Media

- Language Acquisition

- Humanities

- Social Sciences

- Cybersecurity

- Fintech

- Public Health

- Pharmacy

As for next steps in Nexuslinguarum, there is considerable effort and progress in the use cases as well as an increased interaction across WGs to further develop them, which generates also new research opportunities and collaborations among network members that have not been originally anticipated.

Several joint scientific publications have been published so far in the context of the different WGs, and a yet higher number of papers is currently under review or in preparation. Further, joint funding opportunities will be explored at different levels, with a transnational project proposal already submitted on the topic of humanities and social sciences.

Another policy brief in preparation concerns the "inclusion of data from under-resourced languages", to be released later this year. In fact, NexusLinguarum is especially sensitive to the situation of under-resourced languages and their deficit of language technologies, and is promoting the use of linked data technologies to improve this status.

The interaction with standardisation groups, such as the Ontolex and the Linked Data for Language Technologies W3C community groups, has been strengthened and is expected to continue in the future, leading to the publication of guidelines and best practises regarding linguistic data science topics.

On an educational dimension, we are currently preparing a common curriculum for a Europe-wide master degree that the participating institutions could adopt to train a new generation of researchers in the area, thus introducing linguistic data science in a cross-discipline academic infrastructure.

Finally, the organization of the Language, Data, and Knowledge conference (LDK 2021), the flagship conference of NexusLinguarum, will be held on September 1-4 in Zaragoza, Spain. This will be an excellent opportunity for sharing new ideas and raising awareness of recent advancements in the community. A number of workshops and tutorials co-located at LDK are also organised by NexusLinguarum participants.

In summary, despite the difficulties imposed by these pandemic times, we are moving ahead and steadily progressing towards an ecosystem of interoperable multilingual linguistic data. And we are doing so by connecting different research lines from different researchers, coming from different fields, through a number of networking tools such as STSMs, conferences, scientific meetings and, more importantly, the daily work at the level of the different tasks and WGs. That is, through connecting excellent researchers to achieve excellent research.