Slovene
terminologies

Dr Nataša Logar Berginc from the University of Ljubljana leads a multi-collaborative project that is using a corpus approach to develop a model for the systematisation of Slovene terminologies

What would you say has been the driving force behind this project?

In 1998, leading Slovene expert on public relations, and a colleague on the project, Dejan Verčič wrote: 'Language is the basic tool of any profession. It is impossible to think without it, or even communicate. Until it is possible to think and communicate about public relations in Slovene language, it will be difficult to speak about Slovene public relations. As long as we do not have our own language to talk about a discipline, then in a sense, we still do not have our own discipline'.

This opinion summarises the key task of any discipline: that by gaining expert knowledge, the terminology of the discipline in one's own language develops as well.

Could you outline the corpus approach and explain how it feeds into your research?

More than two decades ago, foreign, particularly English lexicography proved that modern dictionaries can no longer be produced qualitatively without corpora – collections of authentic texts collected according to predetermined criteria and with a certain goal. These are available in electronic form and equipped with tools that enable multi-layer searching and statistical processing of linguistic data. They are also a source of evidence for linguistic descriptions and argumentation. The data from texts such as linguistic usage enhances the validity and reliability of generalisations, which are provided in the form of a linguistic description especially if compared to a description, which is produced on the basis of the introspective assessment of only a few language speakers.

By satisfying a predetermined network of criteria for the selection of texts, our project is an honest attempt to represent the entire discipline of public relations, and it is the first corpus of this kind in Slovenia.

How has the current financial climate of Europe impacted on your research?

Unfortunately, the economic crisis has an important impact on the greater fragmentation in R&D. A strong economy is a condition for a strong share of R&D, and the countries which are most affected by the economic crisis will certainly be deprived in this area. The anticipated austerity measures in Slovenia will severely affect the budget for R&D, which will further reduce our ability to participate in international research projects, and result in reduced competitiveness in comparison with more developed countries. Countries which maintain stable economies or overcome the economic crisis sooner will primarily support their own R&D projects. Unfortunately, competition rather than cooperation between countries in this area is expected – at least until the return of economic growth.

Influencing language can prove difficult, particularly when ideologies and politics are concerned. Are you looking to develop any new methods or strategies to address these issues?

Normative handbooks, particularly dictionaries, function in a society as authoritative books and, as such, the solutions they provide may even be used as arguments for the promotion and dissemination of intolerance. While working on the project, we are aware that at this point we have to be socially responsible and sensitive.

In our project, these problems are partly neutralised by the corpus data, which display the development of terms through time, frequency and distribution according to texts, authors and typical co-texts. In short, the data on actual linguistic use directs the editors of the dictionary on public relations away from subjective and quickly formed opinions. Furthermore, when marking the terms (such as preferred term, deprecated term, proposed term), we will comply with the relevant international standards.

What applications do you hope your research will have?

We hope that the forthcoming web dictionary of public relations will become recognised as a model worth following, and that the technological infrastructure that is developed will have many users who together will build a large base of Slovene terminology. We would also like companies and organisations that are willing to at least symbolically, as well as financially, support this project to become part of this development story. We plan to share and upgrade our knowledge in the international sphere. Although the project is currently based locally due to the language, its results can certainly be transferred across language borders.

Preserving Slovene terminology in a globalising world

Globalisation is challenging the full functionality of Slovene language, more specifically terminology. However, a new research model using an exemplary discipline may hold the key to standardising terminology in future fields

TERMINOLOGY, THE STUDY of terms, is an important aspect of all modern national standard languages. Unlike general vocabulary, which is characterised by general development, social and aesthetic factors, terminology is defined by those that regulate the meanings given to a system of words in a particular discipline.

The development of terminologies, particularly the standardisation of terms through the creation of dictionaries, is important for preserving a national language. Additionally, modern linguistic theories see terminology as an important aspect of linguistic policy and planning, which concerns linguists and experts from individual fields, as well as society and all the users of a language.

However, the popularity and spread of the English language, brought on by the globalisation of politics, economy and culture, is threatening to create lexical among other holes in different languages. One such country facing this problem is Slovenia. Many Slovene experts in various disciplines, ie. translators, journalists, etc., face the task of developing Slovene (especially when the naming of things is concerned) through the English – Slovene encounter on a daily basis.

Printed evidence of Slovene terminologies was first found in the texts of Protestant writers during the 16th Century before development became more extensive during the second half of the 18th Century. Fuelled by the competitive threat of the German language, Slovene intellectuals published the first terminological dictionary in 1880, with many more created after.

During the last couple of decades Slovene experts of various disciplines have been unable to meet the demands for terminological dictionaries. As – combined with the proliferation of English scientific publications and English speaking lecturers at the universities – the naming of concepts usually takes some time. Since it is necessary for the formation of terms to take place simultaneously, it has been realised that in order to have a fully developed language, Slovenia must focus on creating and developing its terminological dictionaries and databases rapidly, at low-cost and with modern language technology tools. This movement is backed by the national government who has already co-financed several language corpus research projects.

LEADING THE WAY

Dr Nataša Logar Berginc is leading one such project, intent on continuing and adding to those projects that have gone before. Financed by the Slovenian Research Agency, Pristop and the Chamber of Commerce and Industry of Slovenia (CCIS), the project, 'Terminology data banks as the bodies of knowledge', intends to transfer the knowledge of reference corpora to the corpora of professional texts, done so through the provision of free tools. Using these tools, the project aims for experts in various disciplines, alongside translators, journalists and those who determine Slovene terminology, to create a free online dictionary at a fairly rapid pace.

With many initiatives and individuals in Slovenia already attempting to create terminological dictionaries and databases, Logar's efforts have been collaborative with other contributions from a number of experts at the University of Ljubljana's Faculty of Social Sciences, as well as colleagues from the University of Ljubljana's Faculty of Arts, consultancy and communications company Pristop, Amebis, a company working in the field of language technologies, the Institute for Applied Slovene Studies, Trojina, and the Fran Ramovš Institute of the Slovenian Language. In total, six different institutions and at least four different scientific fields have collaborated on the project.

A CORPUS APPROACH

Logar chose to take the corpus approach, believing it to produce technologically and methodologically more advanced and user-friendly results. Using the public relations field as an example of a particular discipline, the ultimate goal is to produce a free, online terminological database as a model for future fields and disciplines. Comprised of 2,000 entries, the database will also aim to provide information regarding the form, volume and linguistic quality of dictionaries using this methodology and scale.

Alongside the corpus approach, the project also intends to develop a technological offering that allows linguistic annotation, provides a user-friendly interface and creates programmes for the automatic extraction of terminology. The technology will also incorporate lists of term candidates in a web program to quickly and easily allow non-specialists to contribute to the database.

Additionally, the technology interface will provide automatic links between parts of the dictionary and a specialised corpus to allow both the general public and specialised experts access to additional information that does not belong to the dictionary entry, but could be deduced from the use of a term in context.

The database will be available to everyone that possesses an interest in sharing Slovene terminological or general language knowledge. The portal will also offer translations in a bilingual or multilingual capacity whilst general policy for the use of the database online and offline will be overseen by the Creative Commons Attribution Non-Commercial Share Alike license.

To accommodate the natural development of terminology, Logar explains that the Termania portal on which the project is based will allow ongoing updates with relatively minimal effort and minimal or no cost.

With the creation of Slovene terminological dictionaries not deemed significant enough to receive any considerable funding, one of the biggest challenges facing the project's formation was financial-backing. As an applicative project, the Slovenian government demands a quarter of the funding to be given by private investors, which considering the current economic crisis is regarded as a daunting amount to procure. Logar explains that companies are reluctant to co-finance such projects as the market is small and they yield only a minimal return on investment.

Fortunately, Pristop, the CCIS and the Slovenian Research Agency stepped up to fund the database. Logar believes the success of this first model will determine future cooperation between private financers and the extension of Slovene databases in other fields.

EARLY SUCCESS

The two-year project, now underway for almost a year, is already proving its worth. Following a number of appearances at conferences and meetings to outline her goals, Logar began receiving emails and phone calls from particular disciplines wanting to be included or wanting to know when to expect the completion date. Additionally, interest began pouring in from abroad, especially from the countries of former Yugoslavia, which confirmed to Logar and her team that the project was on the right track.

Logar comments that: "The experiences and knowledge which we acquire together with our colleagues who are involved in the preparation of lexical database for Slovene on the 'Communication in Slovene' project are exceptionally valuable. The automatic extraction of terminological candidates, which in our project was managed by Špela Vintar from the Faculty of Arts in Ljubljana, was – among others – the result of the project Slovene Terminology Web Portal". Logar is also hopeful that the project will continue to thrive as the years progress: "It is evident that the project 'Terminology data banks as the bodies of knowledge' is continuing to build on related past and current projects in Slovenia and combines the latest national and international knowledge in this field," she concludes.

University *of Ljubljana*
Faculty *of Social Sciences*

Gospodarska zbornica Slovenije

PRISTOP

Mercator

Zanesljivo vsepovsod
POŠTA SLOVENIJE

ZM ZAVAROVALNICA MARIBOR

ELEKTRO LJUBLJANA