

Language Policy in Slovenia: Language Users' Needs with a Special Focus on Lexicography and Translation Tools

Mojca Šorli¹, Nina Ledinek²

¹ Faculty of Arts, University of Ljubljana, Aškerčeva 2

² Research Centre of the Slovenian Academy of Sciences and Arts, Novi trg 2

E-mail: mojca.sorli@guest.arnes.si, NLedinek@zrc-sazu.si

Abstract

In the following contribution we present the design and the sociolinguistic background of the government-funded **Slovenian Language Policy and User Needs CRP 2016** project conducted between October 2016 and September 2017 under the leadership of the Research Centre of the Slovenian Academy of Sciences and Arts. Broadly speaking, the survey, which constitutes the core of the project, focuses on the language needs of four main categories: speakers of Slovene as their mother tongue; Slovenian minorities living across the border in Italy, Austria, Hungary and Croatia, with their specific linguistic and cultural background (bilingualism); users/learners of foreign languages; and users with special needs. All of these are investigated from the perspective of the legal framework regulating language use in individual fields, communicative practices, empirical evaluation of users' habits and attitudes; and, of particular importance for the present contribution, the current state-of-the-art in language infrastructure, including language technologies and digitisation. "Language description and language infrastructure in Slovenia" is a topic covered by the ZRC SAZU CRP 2016 project that will be treated in this paper in more detail, with special attention given to the questions asked about the use of the existing monolingual and bilingual (multilingual) language resources, in particular, dictionaries and other lexical resources. An in-depth survey will cover different groups of language professionals who use Slovene/foreign languages on a regular basis in the production of written and spoken texts for public use, such as journalists, publicists, fiction writers, bloggers, researchers, copywriters, PR professionals, legal document compilers, business and public administrators, as well as proofreaders and language editors and, last but not least, translators and interpreters.

Keywords: language infrastructure; interlingual resources; translation tools; user needs; online survey

1. Introduction

The **Slovenian Language Policy and User Needs CRP 2016** project conducted by the Research Centre of the Slovenian Academy of Sciences and Arts (henceforth: ZRC SAZU)—specifically its subtopic "**Language description and language infrastructure in Slovenia**", which is treated in more detail in the present paper—includes a study of the use of monolingual as well as bilingual and multilingual

(henceforth: interlingual) language resources, in particular, **dictionaries, corpora and other lexical resources**. The term “infrastructure” is used to incorporate language resources, i.e., sources and tools, as well as language technologies. Due attention is given to the supporting government documents, notably the *Slovenian 2014–2018 Action Plan (for Interlingual Resources)* (henceforth: SAPIR) and the legal framework necessary for the implementation of a language policy. A significant lack of data about the needs and expectations outside the formal education system, especially those of expert user groups, and increasingly intertwined private and public interests in the development of language resources, calls for an in-depth, comprehensive and, as far as possible, unbiased study of the actual habits and attitudes of the various user groups. Such a study could form the basis for future action plans and other binding language planning documents at the national level. Given that, in view of a general lack of surveys, the role of language resources in formal education has been reasonably well investigated, in the part of the survey presented here we focus on the different groups of **language experts who use Slovenian and foreign languages on a regular basis** in the production of written and spoken texts for public use, such as journalists, publicists, fiction writers, bloggers, researchers, copywriters, PR professionals, legal document compilers, business and public administrators, as well as proofreaders and language editors and, last but not least, translators and interpreters. In so doing, we indirectly address the question of contexts of the use of dictionaries and other, primarily lexical, resources. Furthermore, the survey questions elicit information on how specific user needs/aims are related to the use of specific sources and translation tools. In view of the targeted user groups, it was mandatory to survey the use of both interlingual and monolingual resources for both the language of origin (Slovene) and the target language (various foreign, mainly European, languages).

2. (Socio)linguistic background

The Slovenian language community is somewhat specific due to the small number of speakers, resulting in an imbalance between the number of users of the language of origin and any target language. At the same time, it is universal and conditioned by numerous radical changes in attitudes towards the use and planning of language resources, as well as access to them.

In light of the (linguistic, social and sociolinguistic) issues raised in the following paragraphs, our aim here is to relate this (socio)linguistic reality of Slovenian speakers to their actual needs and expectations, thus providing language policy makers and language infrastructure developers with some hard evidence on what they should prioritise. As the final results of the related online study will be available by September 2017, in the present paper we have chosen to showcase the vast range of issues addressed, directly or indirectly, by the survey.

While monolingual dictionaries and other reference books have been published by the Fran Ramovš Institute of the Slovenian Language (henceforth: FRISL) of ZRC SAZU,

an institution that regulates the standardised use of written (and spoken) language, the production of bilingual resources has not been subject to any systematic planning and control at the national level. Since the collapse of most commercially driven bilingual dictionary publishing about a decade ago, the development of interlingual resources has simply been determined by the demands of the free market. However, considering the (small) number of Slovenian speakers and the importance of cross-cultural exchange for Slovenia, providing resources for Slovene is as vital as ensuring the ongoing production of high quality interlingual resources (and research), especially for the leading European languages, including English as the *lingua franca*. The production of high quality interlingual resources is vital due to the significance of foreign language use and instruction **for Slovenian speakers, as well as foreign users/learners of Slovene**.

Here we will focus on the attitudes of (Slovenian) foreign language users with respect to both the available language resources and those that are lacking, particularly dictionaries and other lexical resources, as well as translation tools.

2.1 The changed role of translators (and dictionaries) in the

(semi)automated translation business

In this section, we highlight two (sociolinguistically) relevant factors demonstrating the impact that technological advances, especially in automated translation, have had on the way we now perceive the professional field of linguistic mediation and the translator's role in it. Rapid technological advances have enabled numerous (semi)automated processes whereby human translators are declared (semi)redundant, perhaps one of the most common being the widespread practice of automated website translation. If on the one hand, a general leniency towards clearly inadequate but increasingly widespread fully automated translations of web content can even be detected in academic settings, which by definition (would be expected to) deal with both the theory and practice of translation, notably translation studies, it should not be overlooked that minor language speakers, in particular, are expected to accept linguistic degradation as part of the presumably necessary collateral damage of technical progress. The consequences for the development and status of minor languages are yet to be fully understood. Technological (individual) initiatives along the lines of the multilingual, partly crowdsourced web dictionary **Glosbe** (www.glosbe.com) enable users to access multiple international multilingual databases, which is essentially a positive development, as they aim to improve the level of (human) translation. Paradoxically, however, the low quality of automated website content demonstrates a surprisingly high level of tolerance for linguistic inadequacy.

A third factor, related to the two factors mentioned above, should really be addressed here in order to give a more complex and therefore more adequate picture of the

translation “market”: the intertwining of academic interests and the increasingly commercialised framework to which translation and interpreting as a professional field has been assigned. For reasons of space, however, we have to leave this issue aside.

2.2 The motivation and rationale behind the research of user needs and habits

An overview of research to date reveals at least three different settings in which studies of user habits take place, each with its own set of objectives and motivation. Research on monolingual dictionaries and the role of dictionaries in the teaching process seems to be carried out by: 1) faculty members and doctoral or postdoctoral students at the Department of Slovene Studies, within special projects and in collaboration with other research/academic partners; individual minor scale studies are undertaken occasionally by graduate or master’s students (such as Čebulj 2013 for general monolingual resources, etc.); 2) the central cultural institution traditionally in charge of language resources (the aforementioned Fran Ramovš Institute of the Slovenian Language of ZRC SAZU), which partly operates on continuous financial support at the national level; and 3) recently formed and relatively exclusive initiatives with an explicit interest in the research and production of digital language resources involving language technologies and crowdsourcing, combining public and private initiatives. Whereas the second group is composed mainly of linguists and lexicographers, in the third group the presence of “practicing lexicographers” is notably modest. Nevertheless, the third group have largely been driven by aspirations to compile an already envisaged new corpus-based dictionary of Slovene.

3. Survey and Analysis of Studies to Date and the Relevant

Literature

There has been no comprehensive study to date in Slovenia covering both monolingual and interlingual resources, and including more than one or two specific user groups.

3.1 Monolingual studies

Most studies regarding general **monolingual resources for Slovene** have been carried out in the context of formal education amongst primary and secondary school students. Mostly, the role of the dictionary as a basic tool in the teaching/learning process has been examined, focusing on the comprehensiveness and accessibility of dictionary data. Specifically, the use of the Dictionary of Standard Slovene has been examined as well as its inclusion in teaching Slovenian language at school (Stabej et al., 2008; Rozman et al., 2010; Čebulj 2013). A more detailed overview of the research into monolingual dictionary use in teaching can be found in Rozman et al. (2015). As

found in another attempt at a (monolingual) dictionary survey (Arhar Holdt et al., 2015), the specifics of “professional dictionary use” have not been sufficiently examined compared to dictionary use for pedagogical purposes. There is also a lack of research in the field of Slovene as a second or foreign language (Rozman et al., 2015). Despite its limited data on actual dictionary use, the survey on Slovene language teaching (Rozman et al., 2010) has identified/highlighted the growing use of ICT amongst students, thus suggesting that a similar trend could be expected for foreign language resources for Slovene.

3.2 Foreign language studies

The use of foreign language resources has been investigated mainly in the rather narrow and specific field of formal education amongst university students of translation. In fact, the first of the two most recent studies was conducted on trends in the use of language resources (sources and tools) amongst trainee translators (Hirci 2013), while the second focused on translation queries performed by users of the “Translators, help!” Internet forum (Čibej et al., 2015).

A few earlier studies carried out at the Translation Department, Faculty of Arts, University of Ljubljana, need to be pointed out; namely, Hirci (2007; 2009), Mikolič Južnič (2009), Pisanski Peterlin (2003) and Vintar (e.g., 1999). All of these studies were conducted in the context of university translator training with a focus on the use of text corpora. However, there have been no studies involving various groups of more general users, including language experts of various backgrounds, which would provide a more objective picture of the current state of affairs. In view of the above, the ZRC SAZU survey—more precisely, the section on interlingual infrastructure—foresees a systematic analysis of the actual needs and attitudes of the various professional groups and actors in translation and interpreting, especially as resulting from and conditioned by their professional affiliations and status. In any case, the analysis of interlingual resources was designed to minimise biased interpretations of users’ needs relying on specific groups, such as students of particular subjects, with a maximum dispersion of target groups in terms of age, professional background, education, etc.

Below we highlight two pieces of research (by faculty members) that shed some light on translation practice (and translation practicalities) in Slovenia in the past two decades. The first focuses on the development and use of new translation resources amongst translation students, while the second summarises changes in the translation market that have radically reshaped the profession of translating and interpreting.

3.3 On the application of translation resources

The results from the questionnaires of 2005 and 2012 show changing trends in the application of translation tools (Hirci, 2013: 154–158). While the results show a stable

use of bilingual and general monolingual dictionaries as resources in first and second places respectively, a change is evident in that, in 2012, electronic dictionaries were used almost exclusively, unlike in 2005, when paper and electronic dictionaries were used on a much more equal basis. The vast majority of the respondents in 2012 thus reported using only those resources that can be accessed electronically. Furthermore, the proportion of those respondents who regularly consult text corpora and parallel texts found on Google had considerably increased by 2012 (Hirci, 2013: 155). Another research question showed that the consultation of dictionaries, glossaries, encyclopaedias remains stable, albeit now almost exclusively in digital form (in fact, the proportion of those who consult bilingual dictionaries increases to virtually 100%), as does the use of the Internet (parallel texts). The use of corpora (monolingual and bilingual) is on the increase (from 12 out of 20 in 2005 to 18 in 2012). Perhaps the most striking difference is seen in the decrease in the use of mobile phones as a platform for accessing linguistic information, which in the 2012 survey is not reported at all.¹ On the other hand, there are more users of CAT software (4 out of 20 in 2012 as opposed to 2 in 2005). It can be concluded that the structure of, and familiarity with, the resources used in the examined period is largely unchanged, but the resources themselves are increasingly electronic, i.e., digital. The results of both questionnaires are also highly consistent on the issue of the usefulness of translation resources: roughly 30% of respondents in 2005 and 2012 indicated dictionaries, glossaries and encyclopaedias as the most useful resources, followed by the Internet (parallel texts), with 22% in 2005 and 26% in 2012. An increase is seen in the benefits ascribed to various computer corpora (15% in 2005 vs. 25% in 2012). Very similar results are yielded by reporting on the frequency of use of the listed categories of resources. A change is identified in the use of various monolingual and bilingual corpora (with an increase from 14% in 2005 to 26% in 2012), while a serious drop is also detected in the use of e-mail and translation forums for seeking advice from friends/experts (from 14% in 2005 to 6% in 2012). There is a considerable increase in the use of CAT systems (Hirci, 2013: 156–158). Whereas, in 2005 only 11 (out of 20) respondents believed that their translation work was considerably influenced by the use of electronic tools, in 2012 19 out of 20 believed that to be the case (*ibid.*: 158–159).

3.4 Changes in the “translation market”

The second study is more recent and addresses the problem of the radical reshaping of the translation market, which bears considerably on translator training programmes and, in particular, on the status of professional translators in Slovenia. In addition to the latest developments in lexicology, lexicography and translation studies, radical shifts in translation practice and the use of language resources have been caused by the

¹ Caution is needed when researching the use of mobile devices as sources of linguistic information. At least two very distinct scenarios are at work here: contacting people/experts as sources of information or using mobile applications, such as dictionaries.

drastically changed translation market. This has resulted in the deteriorating status of professional translators in Slovenia and worldwide. The number of translation agencies and companies has gone up since 2004 when Slovenia joined the EU, as has the need for translations. Until about that time, however, in a steadily growing Slovenian (and European) translation market, accompanied by an ever greater accessibility of contemporary translation tools, the relationships and roles of all the stakeholders remained basically unchanged. Kocijančič-Pokorn recently made a repeat of a survey carried out in 2007 by Fišer of the situation in the translation market, concluding that the translation market is still on the rise and the trends established in 2006 still valid: only a few translation companies/agencies seemed to be reaping the fruits of this growth, despite the fact that since 2004 the number of (small) businesses and (self-employed) individuals engaged in translation activity has grown considerably (Kocijančič-Pokorn, 2016: 5; data are based on business entities stating translation as their key activity in the Business Registry of the Agency for Public Legal Records and Related Services, AJPES, for 2014). As implied in the cited article, the growing market and the increased automation of (some aspects) of translation work have caused a disconcerting degradation and led to the increasingly precarious status of the profession.

Technological advances resulting in translation memories, applications for editing terminology databases and automated translation project management have further increased the individual translator's dependence on larger teams and translation agencies. This is corroborated by the fact that even seriously underpaid literary translators, who are only marginally, if at all, replaceable by machine translation software, must often seek additional financial means (from European or national funds) in order to ensure fair payment for their work, thus sharing the fate of their Western European counterparts (see Kocijančič-Pokorn, 2016). The new situation is characterised by demands for virtually instant translations, often into more than one language, which means that complex or larger translation jobs are only manageable by large translation teams. Individual translators are unable to meet the demands of such clients (*ibid.*: 13). Despite ever greater demands on translators in terms of the speed of their services and the quantity of texts, human translating has become increasingly undervalued, with translators increasingly hired for the so-called (full/partial) 'post-editing' of large portions of machine translated texts.

3.5 General studies on user attitudes towards language resources and language policy

In May 2017, a European survey on dictionary use was launched in 29 countries with the support of the European network for e-lexicography (<http://www.elexicography.eu/events/european-survey-on-dictionary-use/>), which was partly aimed at users regardless of their country of origin and partly

country/language-specific. The survey “aims to explore the attitude of language users towards general monolingual dictionaries of their native language” (Corpora list, 9 May 2017). This survey appears to be the first international survey of native language user needs and attitudes of its kind that—in addition to the anticipated similarities in attitudes related to the technology-driven changes in the use of language resources—might provide an insight into potential culture-specific differences in the attitudes of the respondents. The initiative has resulted from ENeL activities aimed at unifying and standardising cross-linguistic lexicographic tools and infrastructures across Europe.

In October 2016, a comprehensive and systematic (sociolinguistic) national study was launched by the FRISL called **Slovenian Language Policy and User Needs** (hence: ZRC SAZU CRP 2016 Study, <http://isjfr.zrc-sazu.si/sl/programi-in-projekti/jezikovna-politika-republike-slovenije-in-potrebe-uporabnikov#v>). A part of this study deals with language resources, and, within that, **multilingual resources**² addressing primarily, but not exclusively, groups of language experts—translators, interpreters and other language professionals using at least one foreign language, including language teachers—with their established daily working routines and strategies for dealing with professional challenges. Naturally, an insight into the use of monolingual resources is of crucial importance and therefore was not excluded from the section on interlingual resources.

More generally speaking, the interdisciplinary research project, which involves many experts, such as legal experts, educationalists, etc.,³ focuses on the language needs of four main categories: speakers of Slovene as their mother tongue; Slovenian minorities living across the border in Italy, Austria, Hungary and Croatia, with their specific linguistic and cultural backgrounds (bilingualism); users/learners of foreign languages; and users with special needs. All these categories are investigated from the perspective of the legal framework regulating language use in individual fields, communicative practices, empirical evaluation of user needs and attitudes and, of particular importance for the present contribution, the current state-of-the-art of language infrastructure, including language technologies and digitisation. The results of the survey will provide an overview of the sociolinguistic situation in Slovenia as well as a description of user needs to help create a platform for the new national language policy agenda. In the following chapters, special prominence is given to some aspects of the interlingual resources survey.

² In the actual study, the term “multilingual (society)” is used to mean the ability of a group of speakers to communicate in more than one language, but the term “interlingual” is used instead in the present paper in the context of language resources to denote the type of both bilingual and multilingual resources.

³ To name just a few participating partners: Academy for Theatre, Radio, Film and Television, two Law faculties, Faculty of Arts, Institute for Ethnic Studies, Centre for Slovenian as a foreign/second language, Pedagogical Institute, etc.

4. The ZRC SAZU CRP 2016 Study - Interlingual Resources:

Content and Method

The **online survey** on interlingual resources has been designed with a view to exploring the **actual needs, practices and attitudes** of language users that can be aggregated to help identify the potential need to amend legislation regulating language use and speaker rights. In this paper, we present the design of the section focusing on key general and specialised resources for foreign languages. Overall, the aim of the section on interlingual resources is to give an illustrative insight into how users themselves reflect on their use of language resources, particularly with regard to the various categories of these resources.

4.1 Target groups

As stated above, and in view of research carried out to date, such as Hirci (2013) focusing on students of translation and Čibej et al. (2015) examining the habits of professional translators, we have sought to design from scratch first and foremost a survey of the use of language resources on the part of: **a)** professional translators, interpreters; **b)** other language experts using foreign languages professionally on a regular basis in the production of written and spoken texts for public use; and **c)** general users. While the section on interlingual resources is very much focused on professional use, both in opposition to private use and non-expert use, the part of the ZRC SAZU CRP 2016 Study investigating monolingual resources for Slovene seeks to investigate language issues from the perspective of field experts as well as non-experts. The differences between expert and non-expert users are, in fact, in themselves an interesting research topic, and we expect questions to arise in the process.

As stated above, the survey on interlingual resources investigates the habits and needs of translators and interpreters, proofreaders and language editors, journalists, publicists, legal document compilers, business and public administrators, etc. Also invited to participate in the survey, specifically through their professional associations, are the Slovenian Scientific and Technical Translators, Slovenian Association for Permanent Court Interpreters and Translators, Slovenian Association of Literary Translators, Slovenian Association of Conference Interpreters, Association of Slovenian Film and Television Translators, Slovenian Proofreaders' Association, Slovene Association of Journalists, Translation Services with the Government Secretariat-General, etc. In addition, potential respondents will be invited to participate in the survey through the appropriate mailing lists, forums and language-related websites, as the aim is to include as many general users as possible.

4.2 A note on Slovenian language policy drafting: the need for an adequate language technology taxonomy

The popular discourse on the need to develop language technologies for Slovene, particularly vociferous on the part of (technology-oriented) stakeholders/project partners, is often very general, disregarding the specifics of individual infrastructural units and the actual needs and attitudes of their potential users. From this point, and for the purposes of more efficient language policy drafting, it would be necessary to adopt a functional taxonomy of language technologies as well as setting priorities according to a clear set of criteria.

In the narrow sense, language technologies (henceforth: LT) are generally believed to include all forms of language processing and pre-processing (tokenisation, named-entity recognition, etc.), tagging, parsing, semantic analysis, (morphological and phonetic) lexicons, etc., while speech technologies include speech recognition and synthesis and other speech-related technological products. While these listings/facts are relevant for field experts, from a general user's point of view, LT can fundamentally (and more intuitively) be divided into: 1) tools serving the compilation of digital dictionaries, corpora, other manuals, etc. (mainly tools designed for researchers and experts in the field); and 2) tools designed to solve language-related problems encountered by general users (machine translation, data summarisation, speech recognition and synthesis, etc.). Probably the most widely used LT applications at this stage are grammar and spellcheckers (Krek, 2012: 14). On the other hand, LT include speech technologies, technologies for users with special needs, specialised technologies, such as those for translators, interpreters, etc. Another related distinction that should be made when talking about language resources is between so-called "applied" resources and LT applications, a distinction that is also important in terms of financial transparency. For one thing, lumping "applied" bilingual (general and specialised) dictionaries, parallel and monolingual corpora with taggers, parsers, machine-readable syntactic or semantic lexicons, etc., is manipulative in view of the fact that the two groups have different end users and do not serve the same purpose. In other words, it should be clear which resources are designed primarily for NLP and which are primarily for human users.

Even though it makes sense to conceive of the various tools as part of a broader category of LT, we would like to emphasise that, from the perspective of Slovenian language policy, each tool requires individual treatment: it is assigned a place in the priority list with regards to its design and objectives, as well as in relation to language policy as a whole. Although perhaps not so important from a purely technological perspective, this analysis is crucial from the point of view of meeting the needs of the various user groups. To give an example: unless it is to be an end in itself, listing a machine translation system as a priority should be supported with empirical evidence

on the types of text such a system should be developed for: general language or, perhaps, specialised texts with a high proportion of terminologies, such as MT@EC,⁴ the online machine translation service provided by the European Commission. These tools have been specifically developed and used in combination with specialised translation memories. Currently free of charge, the conditions of use will eventually become “part of the sustainability plan for the new EU Automated Translation platform (eTranslation), which is funded through the Connecting Europe Facility programme”.

Moreover, the rapidly developing relationship between public-private initiatives and the technology-driven economy, on the one hand, and the (digital) humanities and LT, on the other, is accompanied by a potential lack of transparency in determining the goals and priorities of language policy. In view of this, there is a need to draw a clear line between the requirements and expectations of the LT community as a professional field—which has its own (commercial) interests, whether in public research institutions or private institutions—and that which represents the common interests of the community at large, which will contribute to the development of LT indirectly, in the form of taxes through the state budget.

4.3 The significance of the production and development of interlingual

resources for Slovene

According to the Eurobarometer 386 of 2012, as many as 92% of Slovenians (aged 25–64) speak at least one foreign language, which places Slovenia in fifth place among EU countries (the average for EU is 54%) (Krek, 2012: 1): “37.2% of these 92% can use two and 34.1% even three or more languages. In the 50 plus population 27.8% speak English, the proportion rising to 50% in the 35 to 49 age group and to as high as 75.5% in the lowest age group (25 to 34)” (ibid.: 14). On the other hand, knowledge of German, French and Italian is more constant, with the first at around 30% and the last at around 10%. We can safely predict that English is by far the leading foreign language and perhaps soon to become a second language (L2) in Slovenia, as is already the case in the Netherlands, for example, which for a while has prioritised English as L2. However, if these data seem encouraging we cannot say the same for language resources supporting multilingualism (SAPIR: 12). As rightly established in *The Slovene Language in the Digital Age study* (Krek, 2012; <http://www.meta-net.eu/whitepapers/e-book/slovene.pdf>), translators and

⁴ See MT@EC:

https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-mtec_en (Accessed 24 May 2017)

interpreters use dictionaries, corpora and translation memories, which need to meet the desired standards in terms of quality and scope (ibid.). According to the above study, Slovene is rated very low on the scale of MT development, particularly with regard to resource and technology enablers, which include general and specialised bilingual dictionaries and lexicons, parallel and comparable corpora, taggers, syntactic and semantic parsers, etc.

4.4 Type of inquiry in the ZRC SAZU CRP 2016 Study

There are four main categories of questions, of which not exclusively but mainly category d), in particular the part on language infrastructure, is addressed in this paper:

- a) attitudes related to language use;
- b) communicative practices and usages;
- c) language users' needs;
- d) language description and language resources.

The aim is to verify some of the generally accepted truths and assumptions regarding the state of needs in the area of interlingual resources against the responses acquired in the survey. These data should assist in determining the possible priorities for a 5–10 year action plan at the national level.

4.4.1 Question type profiling

As a number of drawbacks can be identified in studies based solely on data analysis (e.g., log-files) or social media (e.g., translation forums), we have opted at this point to conduct an online questionnaire survey, which includes mainly closed-ended but also open-ended questions (as a free comment). On the whole, according to Müller-Spitzer (2012: 5), these are expected to elicit more informative responses and are actually more appropriate “in web surveys than in paper surveys, especially when the response field is large”. The idea is for the results to be later complemented with data/query analysis, such as that found in translation forums, e.g., Facebook groups, mailing lists, online chat rooms, etc.

One of the open-ended questions is dedicated to language pairs that users prioritise in the scheme of planning publicly (co-)funded revisions or new dictionary editions. More importantly, the questions also investigate attitudes, albeit indirectly, towards actions that were drafted in the SAPIR 2014–2018; for instance, respondents' views with regards to investing public money in the digitisation of out-of-date dictionaries with the aim of integrating them into multilingual portals, as drafted in the SAPIR. Moreover, we aspire in this study to establish the current trends and the relationship between the role of traditional (electronic) resources and those of CAT and MT

systems. This includes obtaining information on the type of source or tool, such as a CD-ROM, a web dictionary, a CAT or MT system.

Furthermore, drawing on extensive lexicographic practice, the authors of this paper can safely conclude that language policy in general and language planning documents in particular focus almost exclusively on the infrastructure for teaching foreign languages; therefore, the needs of other types of users have been prioritised in the survey in order to bridge the gap.

4.4.2 The questionnaire

The questions targeting language experts as compiled in the first version of the online survey include⁵:

- 1) the type of texts they most frequently translate:
 - a. *Literary*
 - b. *Journalism/media*
 - c. *Commercial*
 - d. *Sworn court translations*
 - e. *Expert and academic/scholarly*
 - f. *Technical*
 - g. *Other*

- 2) the language resources and translation tools they most frequently use:
 - a. *Bi- or multilingual dictionaries*
 - b. *Machine translation*
 - c. *Monolingual dictionaries of the target language(s)*
 - d. *General lexical and terminological databases on the Web*
 - e. *Translation memories*
 - f. *Bi- or multilingual text corpora*
 - g. *Monolingual text corpora*
 - h. *Translation forums and other social media*
 - i. *Speech repositories (e.g., of the EU Directorate-General for Interpreting)*
 - j. *Other*

- 3) the role of social media:

⁵ The questions have been translated from Slovene into English by the author for the purposes of this article.

Is using a translation forum or other social network your first choice or do you consult the forum only when you cannot find the answer in a standard lexical resource, such as a dictionary or lexical database?

- a. A translators' forum or other social network is my first choice followed by standard language resources*
- b. Standard language resources are my first choice followed by a translators' forum or other social network*

4) the role of translation technologies (memories, MT, etc.):

What percentage of your work is completed via a translation desktop, i.e., a translation memory and a machine translation system (MT)?

- a. Translation memory: less than 20%/between 20% and 40%/between 50% and 70%/over 70%*
- b. MT: less than 20%/between 20% and 40%/between 50% and 70%/over 70%*

5) the use of native language (i.e., Slovene) dictionaries, corpora, etc.

6) attitudes and opinions showing user priorities in the development of language infrastructure and its financing:

Which of the below resources in the field of interlingual resources for Slovenian users should, in your opinion, become priority in the next 5–10 years in terms of upgrade or development funded with public money (you can choose up to 2 resources)?

- a. Bilingual text corpora, i.e., parallel or comparable text/translation corpora*
- b. Machine translation*
- c. Bilingual dictionaries and bilingual lexical databases for the prioritised language pairs*
- d. Slovenian Wikidictionary, Slovenian Wikipedia, Wikisource and other collaborative interlingual resources*
- e. Terminological databases and a terminological portal*
- f. Multilingual information portal (with links to the existing sources and tools)*
- g. Other*

Comments:

7) users' favourite IT platforms for accessing multilingual information, and similar:

Which information-communicative platforms do you normally use to access information in/on foreign languages? (Mark with 1 to 7 whereby the most frequently used platform is 1 and the least frequently used one is 7.)

- a. Google*
- b. Websites*
- c. Mobile applications*

- d. *Social media (Facebook, Instagram, Pinterest, Twitter and similar)*
- e. *Electronic resources (e.g., CD-ROM)*
- f. *Paper resources*
- g. *Other*

Comments:

- 8) what source and/or tool do you miss the most when you are working in a multilingual context within the field of your expertise? (Please, list your language combination(s))
- 9) you perform your language services:
- a. *As a self-employed language expert with privately owned language resources*
 - b. *Employed in a private company/agency with access to language and translation technologies*
 - c. *Employed as a civil servant or public administrator with access to language and translation technologies*
 - d. *Other*

5. Conclusion

The questions are designed so that they enable the assessment of the efficiency and actual benefits of some of the already financed public projects for the development of language resources. Any future action plan needs to take into account the limited financial means allocated to the development of language infrastructure and the fact that Slovenian speakers are yet to see the compilation and publication of some of the most basic corpus-based (monolingual and interlingual) resources, such as a comprehensive monolingual dictionary of contemporary Slovene, a pedagogical monolingual dictionary, an SFL (Slovene as a Foreign Language) dictionary, a Slovenian–English dictionary, etc., calling for a sensible judgement on which of all the possible language resources, including language technologies, are truly urgently needed in the most imminent future. The missing resources are, in fact, as has often been pointed out, critical for the development of language technologies for Slovene.

While in a systematic analysis we study the actual needs and habits of members of all of the major language-related professional associations, we also ask questions in the online survey that will give an insight into the daily (social) reality of individual language experts: forms of employment, working conditions, the degree of professional autonomy, social status, etc. This will fill the gap in data regarding the needs and expectations beyond the system of formal education, particularly of professional groups who are the most actively involved in language mediation and in the production of texts for public use in foreign languages and in Slovene. On the basis of the final results (projected for September 2017), it will be possible to plan the development of language resources and LT, whereby adequate, more intuitive

functional distinctions within the field of LT, as suggested in the present paper, will serve the purposes of greater transparency in language planning.

The online survey on interlingual resources is part of a broader interdisciplinary research project, the aim of which is to provide data on the sociolinguistic situation and user needs in Slovenia for the compilers of the key Slovenian language policy documents (resolutions and action plans). Sociolinguistic as well as legal aspects will be examined due to the fact that any language policy as a public policy in the interest of all the speakers must necessarily be adequately legislated. The comprehensive on-line survey analyses Slovenian speaker attitudes, communicative practices and language infrastructure, bringing all of these into a meaningful relationship with the need to develop language technologies. Ultimately, this project is to produce a comprehensive and empirically based study of key sociolinguistic issues, including the attitudes of language users towards the existing language infrastructure and that which is lacking, for the new *National Language Policy Programme* and a future *Slovenian Action Plan (for Interlingual Resources)*.

6. Acknowledgements

This contribution is based on research conducted within the *Slovenian Language Policy and User Needs* national project (V6-1647) and partly financed by the Faculty of Arts, University of Ljubljana.

The authors would also like to thank the anonymous reviewers for their useful and much appreciated comments.

7. References

- Arhar Holdt, Š. (2015). Uporabniške raziskave za potrebe slovenskega slovaropisja: prvi koraki. In V. Gorjanc, P. Gantar, I. Kosem, S. Krek (eds.) *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL, pp. 136–149.
- Arhar Holdt, Š., Čibej, J. & Zwitter Vitez, A. (2016). Value of language-related questions and comments in digital media for lexicographical user research. *International Journal of Lexicography*. doi:10.1093/ijl/ecw017.
- Arhar Holdt, Š., Čibej, J., Zwitter Vitez, A. (2015). S pomočjo uporabniških jezikovnih vprašanj in mnenj do boljšega slovarja. In V. Gorjanc, P. Gantar, I. Kosem & S. Krek (eds.) *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL, pp. 196–214.
- Arhar Holdt, Š., Kosem, I. & Gantar, P. (2016). The Dictionary Typology: the Case of Slovene. In T. Margalitadze & G. Meladze (eds.) *Lexicography and linguistic diversity: proceedings of the XVII EURALEX International Congress, 6-10 September, 2016*. Tbilisi: Ivane Javakhishvili Tbilisi State University, pp. 179–187.

- Čebulj, M. (2013). *Raba slovarja v 1. in 2. triletju osnovne šole*. University graduation thesis. University of Ljubljana: Faculty of Education.
- Čibej, J., Gorjanc, V. & Popič, D. (2015). Vloga jezikovnih vprašanj prevajalcev pri načrtovanju novega enojezičnega slovarja. In V. Gorjanc, P. Gantar, I. Kosem & S. Krek (eds.) *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL, pp. 168–181.
- Fišer, D. (2008). Recent trends in the translation industry in Slovenia. *The Journal of Specialised Translation*, Issue 10. http://www.jotrans.org/issue10/art_fiser.pdf (18 February 2016).
- Hirci, N. (2003). Prevajanje danes in jutri: delo s sodobnimi prevajalskimi viri in orodji. *Jezik in slovstvo* 3-4/48, pp. 89–102.
- Hirci, N. (2007). *Učinkovitost uporabe sodobnih prevajalskih virov pri prevajanju v nematerni jezik*. PhD Dissertation. University of Ljubljana. Filozofska fakulteta.
- Hirci, N. (2009). Uporaba sodobnih prevajalskih virov pri izobraževanju prevajalcev. In V. Mikolič (ed.) *Jezikovni korpusi v medkulturni komunikaciji*. ZRS Koper: Založba Annales, pp. 57–87.
- Hirci, N. (2013). Changing trends in the use of translation resources: the case of trainee translators in Slovenia. *ELOPE* 10, pp. 149–165.
- Gorjanc, V., Gantar, P., Kosem, I. & Krek, S. (eds.) (2015). *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL.
- Kocijančič-Pokorn, N. (2016). Nič več obljubljen dežela : dinamični premiki na slovenskem prevajalskem trgu in področju izobraževanja prevajalcev. *Vestnik za tuje jezike*, 2016, 8 (1), pp. 9–21 <http://revije.ff.uni-lj.si/Vestnik/article/view/7174/6878>. (31 May 2017).
- Krek, S. (ed.) (2012). *The Slovene Language in the Digital Age/Slovenski jezik v digitalni dobi*. *METANET White Paper Series*. Springer.
- Lew, R. & de Schryver, G.-M. (2014). Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27.4, pp. 341–359. doi:10.1093/ijl/ecu011.
- Mikolič, V. (2015). Slovarski uporabniki – ustvarjalci: ustvarjati v jeziku in z jezikom. In V. Gorjanc, P. Gantar, I. Kosem & S. Krek (eds.) *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL, pp. 182–195.
- Mikolič Južnič, T. (2009). Vzporedni korpus - prevajalsko orodje in orodje za jezikoslovne analize. In V. Mikolič (ed.) *Jezikovni korpusi v medkulturni komunikaciji*. ZRS Koper: Založba Annales, pp. 117–132.
- Mikolič, V. (2015). Slovarski uporabniki – ustvarjalci: ustvarjati v jeziku in z jezikom. In V. Gorjanc, P. Gantar, I. Kosem & S. Krek (eds.) *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL, pp. 182–195.
- Müller-Spitzer, C., Kopleinig, A. & Töpel, A. (2012). Online dictionary use: Key findings from an empirical research project. In S. Granger & M. Paquot (eds.) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 425–457.

- Pisanski Peterlin, A. (2003). Uporaba novih tehnologij pri jezikovnem pouku. *Jezik in slovnstvo* 3-4/48, pp. 103–112.
- Rozman, T., Krapš Vodopivec, I. Stritar, M. & Kosem, I. (2010). *Nova didaktika poučevanja slovenskega jezika - projekt »Sporazumevanje v slovenskem jeziku«* - Kazalnik 15 <http://www.slovenscina.eu/Vsebine/Sl/Kazalniki/K15.aspx> (12 June 2015).
- Rozman, T., Kosem, I., Pirih Svetina, N. & Ferbežar, I. (2015) Slovarji in učenje slovenščine. In V. Gorjanc, P. Gantar, I. Kosem & S. Krek (eds.) *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL, pp. 150–167.
- Stabej, M., Rozman, T., Pirih Svetina, N., Modrijan, N. & Bajec, B. (2008). Jezikovni viri pri jezikovnem pouku v osnovni in srednji šoli: končno poročilo z rezultati dela. Ljubljana: Pedagoški inštitut. <http://www.trojina.si/p/jezikovni-viri-pri-jezikovnem-pouku-v-osnovni-in-srednji-soli/> (22 June 2015).
- Vintar, Š. (1999). Računalniške tehnologije za prevajanje. *Uporabna informatika* VIII/1, pp. 17–24.
- Vintar, Š. (2016). *Prevajalske tehnologije*. University of Ljubljana. Filozofska fakulteta.

Websites:

- European network for e-lexicography (European survey)*. Accessed at: <http://www.elexicography.eu/events/european-survey-on-dictionary-use/>. (20 May 2017)
- Glosbe*. Accessed at: www.glosbe.com. (20 May 2017)
- Meta-Net*. Accessed at: <http://www.meta-net.eu/whitepapers/e-book/slovene.pdf>. (20 May 2017)
- MT@EC*. Accessed at: https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-mtec_en. (24 May 2017)
- ZRC SAZU CRP 2016 Project*. Accessed at: <http://isjfr.zrc-sazu.si/sl/programi-in-projekti/jezikovna-politika-republike-slovenije-in-potrebe-uporabnikov#v> (20 May 2017)

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

