

Survey Report



**Translation Technology Use
in Regional or Minority
Languages**

The Survey

The aim of this survey is gathering information about the actual use of translation technology in the context of minority languages as defined by the European Charter for Regional or Minority Languages (ECRML available on-line at http://www.coe.int/t/dg4/education/minlang/textcharter/default_en.asp), which applies only to languages

1. traditionally used by citizens of a state and thus explicitly excludes languages used by recent immigrants from other states, and
2. which significantly differ from the majority or official language - disregarding local dialects of majority languages, and
3. either have a territorial basis within the state or are used by linguistic minorities within the state as a whole (including Yiddish and Romani).

A major concern for the author of this survey is the assessment and evaluation of what kind of translation technology is most appreciated by translators and which kind of tools are deemed necessary and important for translation in the setting of linguistic minorities. In addition, information about free and open source tools used and developed by and for linguistic minorities is collected. In short, the survey had the following primary objectives:

- to collect data regarding the practical use of translation technology tools within Regional or Minority Languages (RML),
- to collect data regarding the use of free and open source translation technology software packages within RML,
- to collect data that allow for the evaluation of the importance of different translation technology modules.

The survey is aimed at all persons dealing with language and translation within a Regional or Minority Language (RML) setting: translators – employed or free-lance, terminologists, legal experts and administrative staff, civil servants, language planners. It is, to our knowledge, the first attempt to address the use and pervasiveness of translation technology by and within RML.

The questionnaire

The link to the following questionnaire was communicated to about 200 e-mail addresses, including members of translators associations in minority regions (South Tirol, Catalonia, Basque region, South Africa), official translation services in minority settings (South Tirol, Canada, Catalonia), associations and organizations concerned with regional linguistic minorities, as well as posted to relevant discussion groups in social media such as academia.edu, LinkedIn (“translators, interpreters, linguists, linguistic mediators and language professionals”, “Translation Automation”, “CAT tools & translation technologies: MT, TM, Google and others”, “Translation Management System (TMS)”, Yahoogroups (“Innovations in Translator’s Training”, “CATMT”). Overall, more than 10.000 persons should have received information about the survey, either

personally or through interest/discussion groups, but only 92 questionnaires were answered.

Translation Technology Use in Regional or Minority Languages

Information regarding the respondent

This survey (8 questions) is aimed at all persons dealing with language and translation within a Regional or Minority Language (RMS) setting; we would like to collect data regarding the practical use of translation technology tools and especially free and open source software in such settings. Don't forget to click on the button NEXT at the bottom of the page to complete the 2nd page of the survey. Please forward this survey to colleagues who deal with language and translations in a minority setting.

*1. Which Regional or Minority Language do your answers refer to?

Country	<input type="text"/>
Region	<input type="text"/>
Language(s) or language pair	<input type="text"/>

*2. Regarding your status/person, you are a

- official / civil servant
 translator employed by a public institution
 translator employed by a company
 free-lance translator
 terminologist employed by a public institution
 terminologist employed by a company
 lawyer employed by a public institution
 lawyer employed by a company
 researcher
 other

please specify

3. Do you use the following translation technology tools? (multiple selection possible)

- translation memory
 text alignment
 terminology management
 terminology extraction
 machine translation system (minority language / majority language)
 subtitling
 corpus management
 software localization
 none

others, please specify

Next

4. Do you have at your disposal or use the following services in your community? (multiple selections possible)

- translation memory repository
- terminology repository
- phraseology repository
- central machine translation system (minority language / majority language)
- central translation management
- bilingual searchable corpus repository
- central repository for official documents

others, please specify

5. Are you involved in or do you contribute data to one of the following services? (multiple selections possible)

- translation memory repository
- terminology repository
- phraseology repository
- central machine translation system (minority language / majority language)
- central translation management
- bilingual searchable corpus repository
- central repository for official documents

others, please specify

6. How would you rate the importance of the following modules?

	weak	important	essential
translation memory repository	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
central text alignment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
terminology repository	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
terminology extraction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
phraseology repository	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
central machine translation system (minority language / majority language)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
central translation management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
bilingual searchable corpus repository	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
central repository for official documents	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

others, please specify

7. Does your community use Free and Open Source Software for one of the following services? Please specify which software

translation memory repository	<input type="text"/>
translation memory management	<input type="text"/>
text alignment	<input type="text"/>
terminology repository	<input type="text"/>
terminology management	<input type="text"/>
terminology extraction	<input type="text"/>
phraseology management	<input type="text"/>
central machine translation system (minority language / majority language)	<input type="text"/>
central translation management	<input type="text"/>
bilingual searchable corpus	<input type="text"/>
central repository for official documents	<input type="text"/>
others	<input type="text"/>

8. Does your community produce Free and Open Source Software for one of the following services? Please specify which software

translation memory repository	<input type="text"/>
translation memory management	<input type="text"/>
text alignment	<input type="text"/>
terminology repository	<input type="text"/>
terminology management	<input type="text"/>
terminology extraction	<input type="text"/>
phraseology management	<input type="text"/>
central machine translation system (minority language / majority language)	<input type="text"/>
central translation management	<input type="text"/>
bilingual searchable corpus	<input type="text"/>
central repository for official documents	<input type="text"/>
others, please specify	<input type="text"/>

9. Do you have any other comments, questions, or concerns?

10. Thank you for your kind cooperation!

If you want to be informed about the results of this survey please specify your email address below:

The survey results are presented as descriptive statistics that outline the situation on the ground as reported by survey interviewees. Due to the limited number of responses the survey results are not representative as a whole but merely of the groups that were surveyed in the locations where they were surveyed.

This survey does not aim at offering prescriptive suggestions for policy responses or policy action in the light of its findings, though an overall translation technology policy would be desirable for any linguistic community.

As such, the survey faced a number of methodological challenges, including: lack of recent and reliable statistical information about language policies and translation offices set up by RML as well as difficult access to minority communities, to name just a few examples.

Questions and Answers

Geographical distribution of answers, interviewees and corresponding minorities is rather focused on Europe with a few respondents in other continents as visualized in the following map:



PAGE 1: Information regarding the respondent

1 Which Regional or Minority Language do your answers refer to?

The first question is answered by 100 % of respondents who come from 22 different nation states: Afghanistan, Algeria, Austria, Belgium, Brazil, Colombia, Denmark, Ecuador, Finland, France, Iraq, Ireland, Israel, Italy, Norway, Pakistan, Portugal, Russia, Slovenia, Spain, UK, USA, for a total of 21 minorities and 25 languages concerned.

2 Regarding your status/person, you are a

With regard to the profession of respondents the survey revealed a satisfactory coverage of RMLs offering answers from translators and public servants working to sustain official multilingualism: apart from 24 % who were researchers, all others are translators either free-lance or employed by a language service provider or a public institution, public servants or lawyers working in the public administration with only one terminologist.



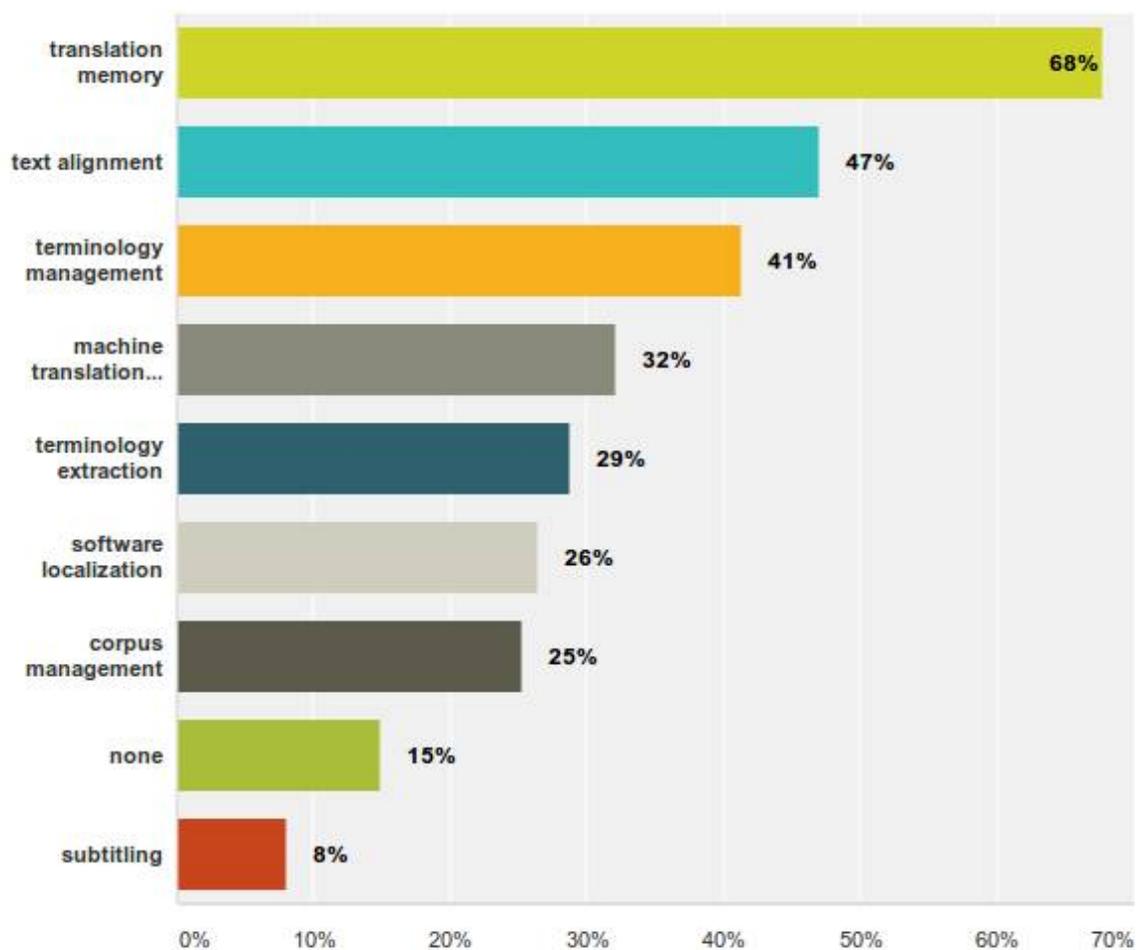
Other professions mentioned under ,others‘ were: trainer, university lecturer, court or conference interpreter, language industry director, language planner.

3 Do you use the following translation technology tools? (multiple selection possible)

With this query we wanted to investigate the general impact of translation technology which was rather good. In contrast to question 4 this refers to the tools used on the desktop by single translators.

With 68 % translation memory systems are leading all other tools. Text alignment as an auxiliary tool to translation memory management is used by 47 % of respondents and terminology management by 41 %.

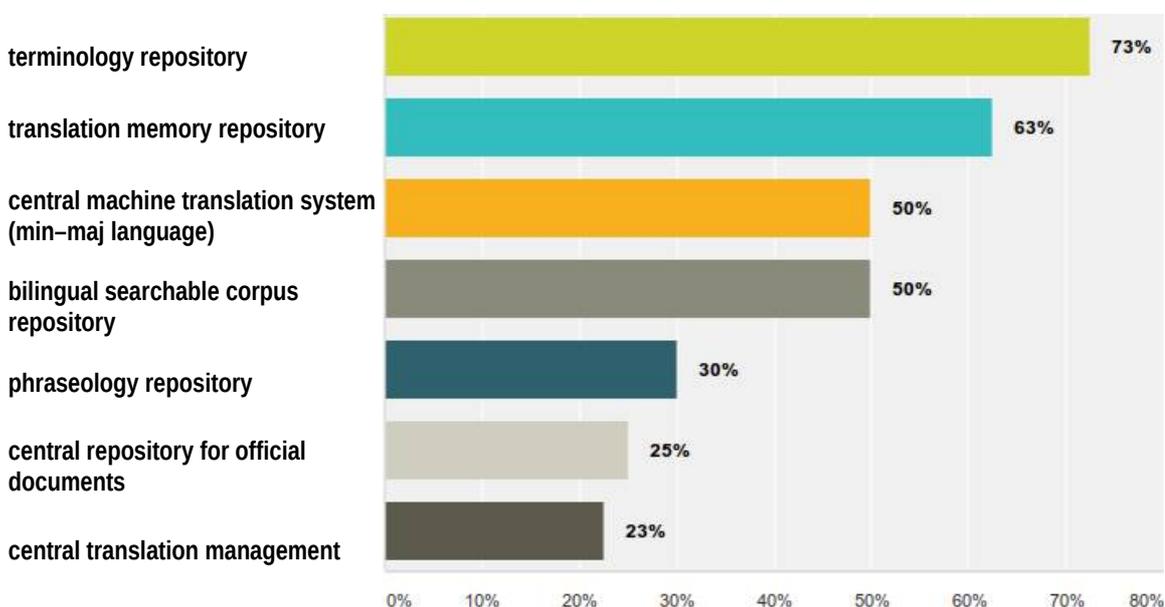
Interestingly, multimedia translation is not very common, with subtitling tools used by only 8 % of respondents and software localization at 26 %. 15 % do not use any translation technology at all; this could be regarded as an astonishing figure but it means on the other hand that 85 % of respondents make use of translation technology. In addition many smaller minority languages do not have any support or the necessary means to develop language-specific technology.



PAGE 2: Use of Translation Technology in Regional or Minority Languages

The response rate for page n° 2 was lower than for page n° 1. It seems that some respondents only compiled the first page without going to the second page of the survey.

4 Do you have at your disposal or use the following services in your community? (multiple selections possible)



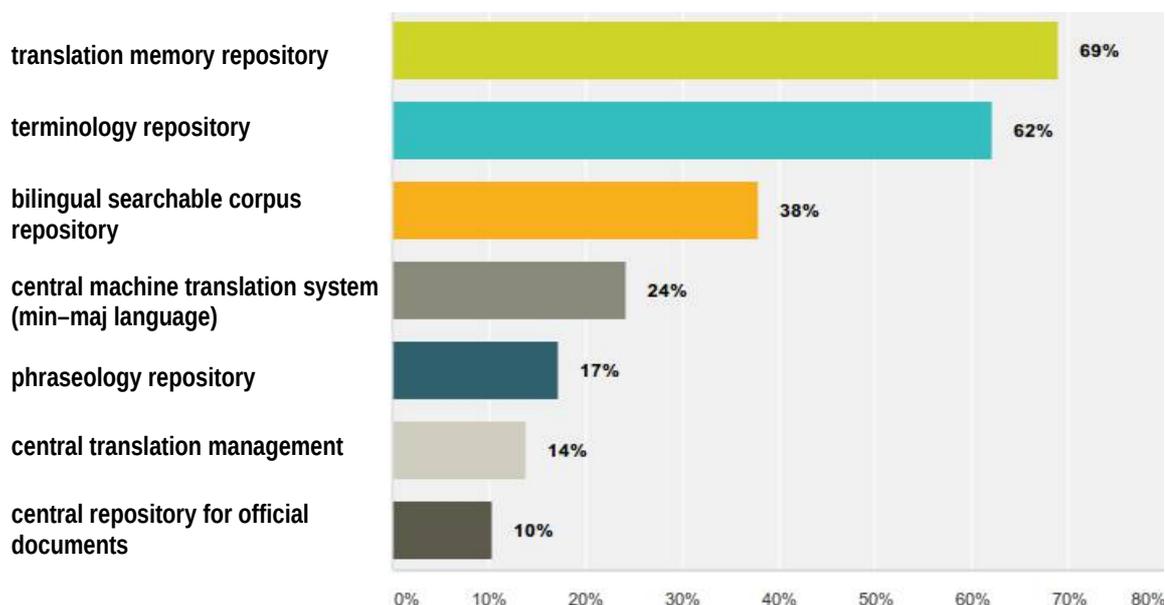
Networking and sharing resources is of overall importance for translation as a social activity and central repositories for language resources constitute the tool of choice to guarantee general accessibility of resources to translators. Rather than desktop oriented tools the attention here lies on networking services which support efficiency and harmonization of translation efforts in RML communities.

Terminology proves to be one of the more common resources together with translation memory. Surprisingly, half of the respondents use a central machine translation system, a figure in part explained by the large use of Apertium for minority languages in Spain and the relative high rate of respondents coming from this country.

As a whole we may say that the use of translation technology is widespread with regional or minority languages.

5 Are you involved in or do you contribute data to one of the following services? (multiple selections possible)

A majority of respondents contribute data to a central translation memory repository which is rather obvious for translators, but as many as 62 % add terms to a terminology repository. About a quarter are also able to contribute their translations to a machine translation system, while phraseology data play a minor role.

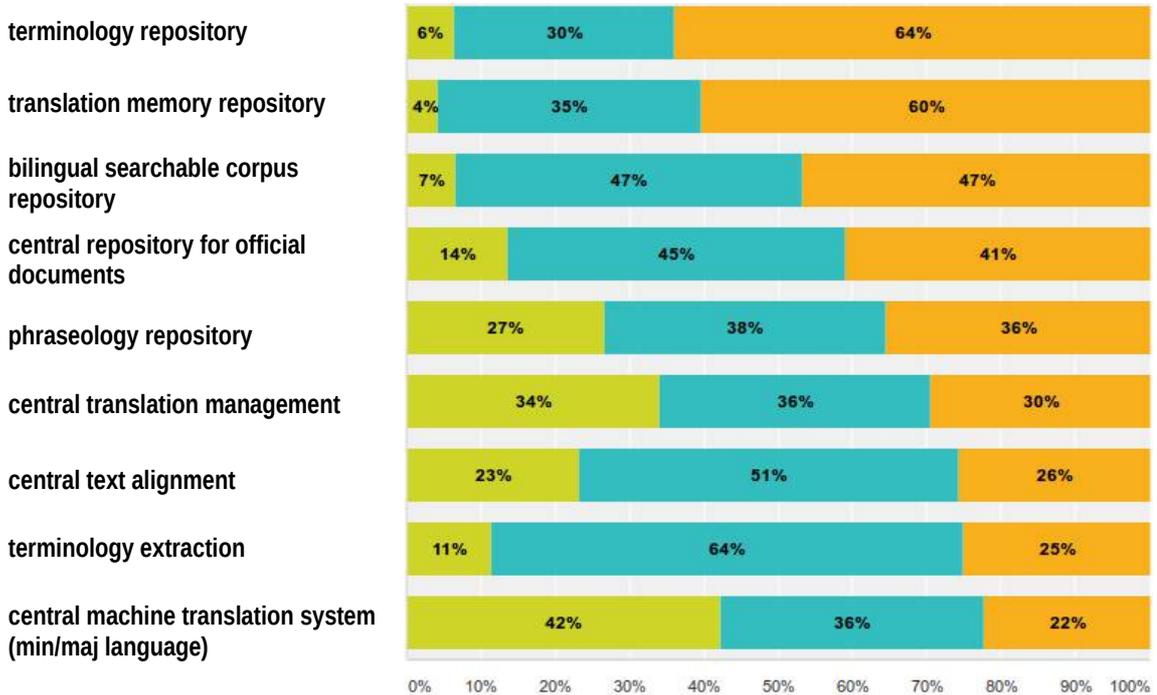


6 How would you rate the importance of the following modules?

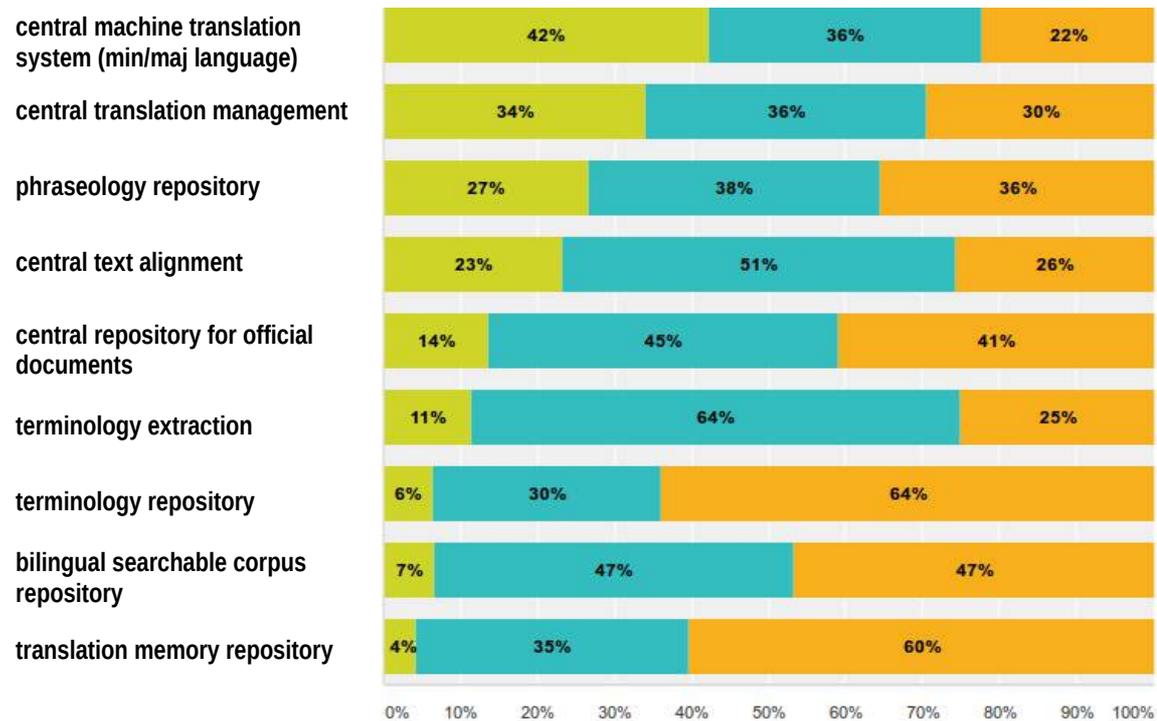
Overall, we have a relatively conservative assessment of language resources by the respondents: a terminology repository is deemed essential by 64 %, important by another 30 % and of weak importance by only 6 %; in second place the translation memory repository with 47 %, 47 % and 7 %, followed by a central repository for official documents and a phraseology repository. Although 50 % of respondents (Q4) have at their disposal a central machine translation system only a meager 22 % believe it is essential.

The evaluation of the subjective importance of the single centrally managed services for translators brings up the following preferences (weak – important –

essential) sorted by essential:



If we change the perspective and view the answers according to the weak rating we get the following chart.



7 Does your community use Free and Open Source Software for one of the following services?

For the same centrally managed services we wanted to know if the communities use existing free or open source software. It seems that the awareness for this kind of software tools is rather low given that this question has a response rate of only 26 %. Furthermore, some respondents mention a commercial tool or a free web based tool like Google Translators Toolkit which is not open source.

Transifex and Tstream are mentioned as translation memory repositories but they are also not open source applications, but in fact commercial web-based tools. Desktop translation memory systems like OmegaT and Virtaal are open source but do not offer centrally managed translation memory repositories.

We have a similar problem with terminology repositories where also some commercial tools are mentioned. Web-based community terminology repositories like Euskalterm and Termofis do not qualify as open source software packages, they are free to use on the web but, as far as I know, you cannot download the software to install and use it in your own community.

Answer Choices	Responses	Responses
translation memory repository	Responses	50%
terminology repository	Responses	38%
central machine translation system (minority language / majority language)	Responses	38%
translation memory management	Responses	33%
text alignment	Responses	33%
others	Responses	29%
terminology management	Responses	25%
terminology extraction	Responses	17%
central translation management	Responses	17%
bilingual searchable corpus	Responses	13%
central repository for official documents	Responses	13%
phraseology management	Responses	8%

Apertium and Matxin are mentioned as open source machine translation systems for the Iberian languages. In the category of translation memory management the only true open source software mentioned are the Okapi Localization tools, for text alignment BiText2TMX and the LF-Aligner, for terminology extraction the TES Terminology Extraction Suite, and for central translation management Pootle.

8 Does your community produce Free and Open Source Software for one of the following services?

Answer Choices	Responses	Responses
translation memory repository	Responses	43%
translation memory management	Responses	43%
text alignment	Responses	21%
terminology repository	Responses	29%
terminology management	Responses	36%
terminology extraction	Responses	36%
phraseology management	Responses	21%
central machine translation system (minority language / majority language)	Responses	36%
central translation management	Responses	14%
bilingual searchable corpus	Responses	36%
central repository for official documents	Responses	29%
others, please specify	Responses	43%

With only 15 % question 8 has an even lower response rate. Worth mentioning are the Basque Elhuyar foundation and the Basque Langune institution who develop language technologies for the respective communities. All in all, a slightly disappointing result, as no new tools emerged than the ones already present in tuxtrans (www.tuxtrans.org) our linux system for translators.

9 Do you have any other comments, questions, or concerns?

Apart from personal annotations and useful link suggestions there is nothing specific to comment here.

A total of 35 % of respondents left their e-mail addresses and will receive this report. Please feel free to comment, send suggestions, proposals or remarks to mail@petersandrini.net.

This survey report constitutes a preliminary evaluation of results. Insights and findings of this survey will be integrated into a more detailed study of translation technology for regional or linguistic minority languages to be published later on.

produced with

tuxtrans GNU/Linux <http://www.tuxtrans.org>
LibreOffice 4.2 <http://www.libre.office.org/>
SurveyMonkey <http://www.surveymonkey.com>

University of Innsbruck, Translation Department

please cite as:

Sandrini, Peter (2014): Translation Technology Use in Regional or Minority Languages.
Preliminary results of the survey (April-May 2014).
<http://www.uibk.ac.at/downloads/trans/publik/survey-report.pdf>