The Use of Machine Translation and Post-editing among Language Service Providers in Spain

Olga Torres-Hostench
Universitat Autònoma de Barcelona
olga.torres.hostench@uab.cat

Celia Rico
Universidad Europea de Madrid
celia.rico@uem.es

Miguel Á. Candel-Mora
Universitat Politécnica de Valencia
mcandel@upvnet.upv.es

Anna Aguilar-Amat
Universitat Autònoma de Barcelona
anna.aguilar-amat@uab.cat

Amparo Alcina-Claudet
Universitat Jaume I
alcina@trad.uji.es

Pilar Cid-Leal
Universitat Autònoma de Barcelona
pilar.cid@uab.cat

Adrià Martín-Mor
Universitat Autònoma de Barcelona
adria.martin@uab.cat

Ramon Piqué Huerta
Universitat Autònoma de Barcelona
ramon.pique@uab.cat

Marisa Presas
Universitat Autònoma de Barcelona
marisa.presas@uab.cat

Pilar Sánchez-Gijón
Universitat Autònoma de Barcelona
pilar.sanchez.gijon@uab.cat

Abstract

This article presents the results of a market survey carried out on the use of machine translation (MT) and MT post-editing (PE) among translation service providers (TSPs) in Spain. This market survey is part of a research project called ProjecTA, funded by the Spanish Ministry of Economy and Competitiveness (Ref. FFI2013-46041-R), which attempts to analyse the flow of MT+PE work in the professional translation sector in Spain and develop guidelines for professional translators to implement MT translation projects. The specific aim of this market survey is to systematically collect, analyse and disseminate information on the use of MT and PE in order to help TSPs make decisions on how to incorporate MT and PE into this business and also to help Spanish universities make decisions on incorporating MT and PE into their degree programmes. Our initial hypothesis was that MT was not implemented evenly across Spanish companies. Quantitative data were collected through an online survey, which was sent to 189 Spanish TSPs in January and February 2015. The results from the survey show that almost 50% of the Spanish companies that participated in the survey use machine translation and carry out post-editing, albeit on a limited basis.

1. Introduction

This study is part of ProjecTA, a project financed by the Spanish Ministry of Economy and Competitiveness (FFI2013-46041-R). ProjecTA works from the premise that the progressive implementation of MT-related services and processes in companies is radically changing the

Proceedings of the 37th Conference Translating and the Computer, pages 90–95,
London, UK, November 26-27, 2015. ©2015 AsLing

profile of professional translators. Market research at an international level has shown that MT is increasingly offered by language service companies. Data from the latest survey on the MT market (Van deer Meer and Ruopp, 2014) reveals that MT post-editing output accounted for 2.47% of revenues (US $828.02 million), with 38.63% of the 1,119 survey respondents reporting that they offered post-editing services.

Our study collected data from a survey on the use of MT systems in Spanish language service providers which are related to the company profile, i.e. turnover and the sectors they work in. It also collated the impressions and attitudes of the companies and their staff regarding MT-related tasks.

2 Methodology

ProjecTA works from the hypothesis that MT has been implemented very unevenly in Spanish companies, and the following data was collected in order to corroborate or challenge this hypothesis:

- Quantitative data. These were collected between January and February 2015 from an online survey sent to 187 Spanish language service companies or those companies with a main office in Spain. The objective of this survey was to discover to what degree these companies employ MT and post-editing.
- Qualitative data. This was collected via three methods: open-ended telephone conversations, in-depth interviews with experts and an expert focus group. All of these were carried out during the second quarter of 2015.

The sample of 187 companies was based on a previous list also drawn up by the ProjecTA research group in the absence of comprehensive directories for language service companies in Spain. The survey was revised by experts in statistics and by representatives from the professional translation sector. In addition, a pilot test was carried out in two companies. A total of 57 surveys were received, however two were ruled out was they were duplicated leaving 55 companies as our data source, which corresponds to 29.4% of the initial sample. Data evaluation and mining was carried out between May and June of 2015 and the survey design allowed us to extract three types of information:

1. Data to provide a basic profile of the companies.
2. The most commonly offered services and languages; sectors they work in and types of clients.
3. Degree to which MT and post-editing is used in these companies.

3 Data Mining

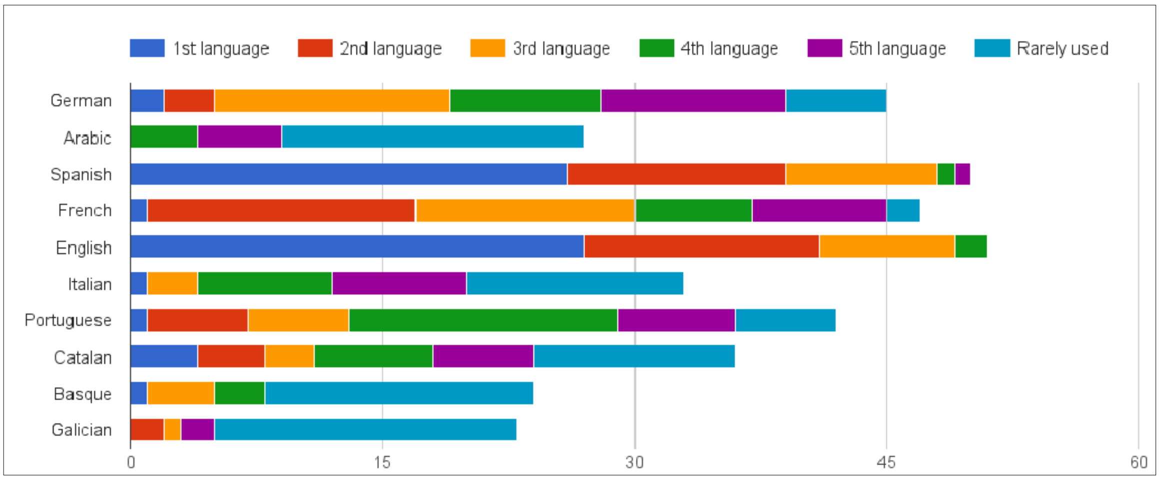
3.1 Company Profile

Although the 55 companies that responded to our survey are spread throughout Spain, the majority are centred in Madrid and Barcelona. Our data shows that small companies are a dominant feature in this sector: 61.8% are microenterprises (up to 9 staff), 23.6% are small companies (10-49 employees) and 10.9% are single member companies. Medium size companies (50-250 employees) account for only 3.7%. The vast majority are companies which have been formed relatively recently: 85.5% since 1991.

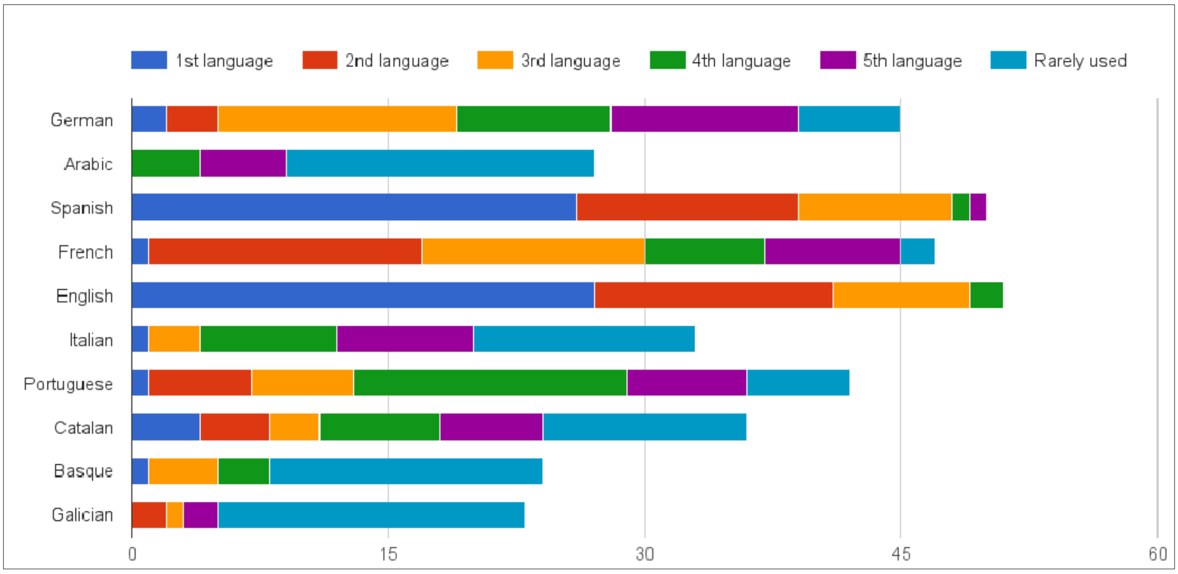
Although not all the companies answered the question regarding yearly turnover, our data show that only 25.5% invoice more than €500,000 per year, 9.1% between €300,000 and €500,000, 31% between €100,000 and €300,000, while 22% have a turnover of less than €100,000.

All the language service providers in our survey offer translation services. Between 54% and 83.6% offer text proofing services (proofing originals, concept review and post-editing). In a third group of services related to translation technology we find translation memory and bilingual parallel text alignment (49%), followed by database and terminology base creation and management (47%). Among terminology services offered, 29% of the companies offer terminology concordance services. Other services consist of client-provided MT post-editing (30%) and pre-editing (23.6%). These companies mention that they also offer services such as interpreting, localisation, subtitling, page layout, proofing galleys, certified translations and transcribing.

English and Spanish are the target languages most in demand, followed by French and German. Other languages also in demand, albeit to a lesser degree, are Portuguese, Italian, Arabic and the other official languages in Spain (Catalan, Basque and Galician), as shown in the graphs 1 and 2.



Graph 1. Source languages companies work with



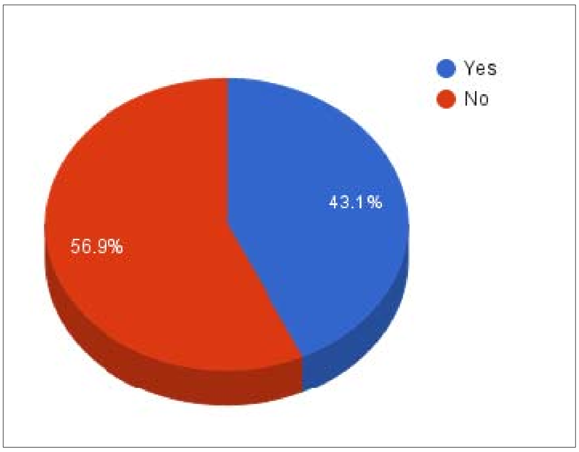
Graph 2. Target languages companies work with

As regards the profile of these companies' clients, the majority are private companies: Spanish (98%) and foreign (81%). The latter figure reflects the high degree of international projection and integration of Spanish language service companies in the international market. Ranked in order of importance, private clients account for 63%, and public administration institutions (central state or from the autonomous regions) 61.8%. 14.5% of the companies surveyed provide services for international institutions and another 14.5% have clients corresponding to European Union institutions. The high degree of business generated between these companies should also be noted given that 43% of their clients are other Spanish language service companies while 47% are foreign.

Our data shows a wide range of sectors. In order of volume generated the main sector is industrial/technical (87%), followed by a block comprising technology/telecommunications, legal and advertising (78.1%), economic/financial (76.3%), tourism/leisure (67.2%) and health/pharmaceutical (63.6%). Real estate and construction industries account for 43% of translation commissions, with publishers last on the list (16.3%), possibly because they outsource directly to freelance translators rather than language service companies. Although to a lesser degree, the companies surveyed mentioned other sectors such as insurance, automobile industry and administration.

3.2 Use of Machine Translation in Companies

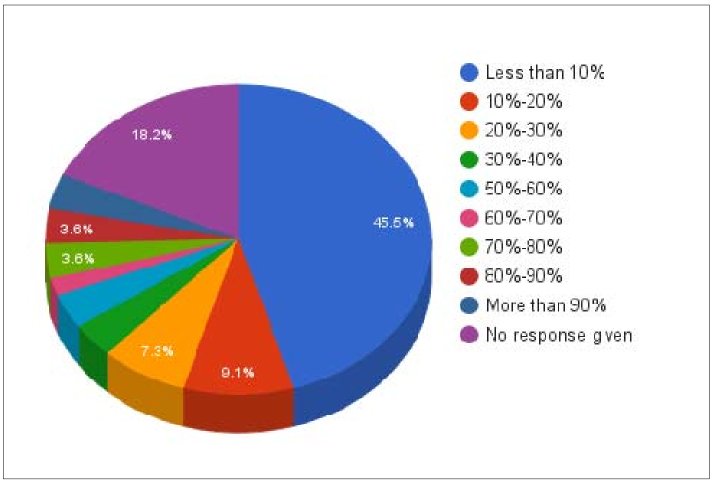
From the 55 companies who responded to the survey, 56.9% do not use MT in their workflow, as opposed to 43.1% who do.



Graph 3. Use of MT in company workflow

Among those companies that do not use MT, the reasons given are: "We don't find MT reliable" (22.2%); "Our clients don't require it" (20.8%); or "Translators don't accept it" (12.5%). 6.9% of the companies claim other reasons such as "poor results"; "given the current state of this technology time saved is outweighed by time spent correcting the text later"; "because of the format of the source texts"; "we don't have the technology at present" and other comments related to major difficulties tailoring and adapting these systems to their specific needs.

For a more precise picture of how widespread MT is in those companies that use it on a regular basis, they were asked to assess to what degree they exploited it, which is summarised in the graph below.



Graph 4. Percentage of projects in which MT is used

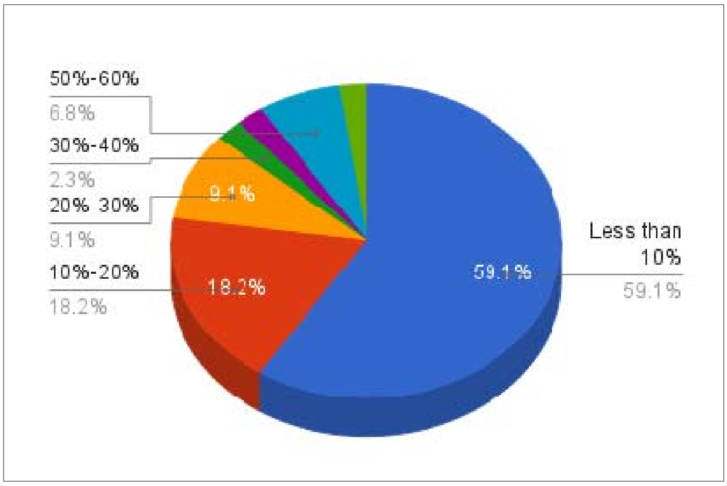
Based on these figures it can be seen that MT use is very low: of the 26 companies that use MT almost half (45.5%) do so in only 10% of their projects.

Scarcely 16% of the companies have their own MT system. Of these, five are statistical systems, two are rule-based and three are hybrids. One of the companies uses various systems.

3.3 Post-Editing

20% of the companies that responded do not offer post-editing services. To a large degree this figure falls in line with the percentage of companies that do not use MT. The difference is due to the fact that some companies do not use MT but receive post-editing commissions directly from the client.

Among the companies which do offer post-editing services (80%) it can be seen that for almost half (47%) post-editing tasks represent less than 10% of the company's total work volume. As illustrated in graph 5, figures for post-editing that represent more than 10% of the total are very low, although it should be pointed out that that in 3 companies post-editing tasks account for between 50% to 60% of their work volume and in another it is as high as 70-80%.



Graph 5. Volume of post-editing commissions in the companies

4 Conclusion

The first difficulty this study had to overcome was to identify a representative sample of the participating companies, given the absence of an official census of this sector. We believe that the final sample is significant and consequently the results provide a good starting point for studying this sector.

The results clearly show that Spanish language service companies are small, offer a wide range of services for different language pairs and work for a very wide range of specialised sectors.

It is noteworthy that almost half of the companies use MT in their workflow. However, it should also be noted that 45.5% of the companies using MT only do so for less than 10% of their projects and only 16% of these companies have their own machine translation system.

Improvements in the quality of MT output in recent statistical translation systems seem to be an incentive to implementing MT in the workflow.

Acknowledgement

This work was supported by a grant from the Spanish Ministry of Economy and Competitiveness. Grant Number FFI2013-46041-R.

References

ProjecTA Project <https://sites.google.com/a/tradumatica.net/projecta/home>

Van deer Meer and Ruopp. 2014. *MT Market Report 2014*. < https://www.taus.net/think-tank/reports/translate-reports/mt-market-report-2014#summary> [Last accessed December 23rd 2014]