

Beyond prescription: What empirical studies are telling us about localization crowdsourcing

Miguel A. Jiménez Crespo
Rutgers University

ABSTRACT

Translation crowdsourcing represents a new and quickly evolving phenomenon that has attracted the attention of industry experts and scholars alike. During recent years the industry has released a number of publications, mainly case studies and best-practice reports, while academic disciplines such as Computational Linguistics and Translation Studies (TS) have primarily focused on empirical studies. This paper attempts to compare and critically analyze research produced from both perspectives and locate these different approaches within the wider cycle of applied and theoretical/descriptive research. The findings of empirical studies on volunteer motivation and quality in TS will be contrasted with the best practices in the industry. This analysis will show a potential avenue to engage both perspectives to collaborate towards closing the existing research gap.

1. Introduction

During the last two decades, translation has experienced a digital revolution that has given rise to new phenomena and practices, such different translation technologies or translation crowdsourcing that are reshaping both industry practices as well as societal views and theories of translation (O'Hagan 2013; Jiménez-Crespo 2013a). Over the years, different stakeholders in the study of translation have followed different paths due to diverging objectives. These objectives range from the more prescriptive and applied industry approaches to theoretical or empirical studies. Industry research often appears in response to the rapid development of technologies and the need to quickly adapt to an ever-evolving field. For example, the industry has tried to rapidly understand, harness and exploit the power of the crowd to produce translations (Jimenez-Crespo 2011). This means that industry experts normally produce applied research at a much quicker rate than academic disciplines (O'Hagan 2013). On the other hand, Translation Studies (TS) often trails behind industry research, adopting industrial de-facto models and conceptualizations that result from an applied and prescriptive approach (Jimenez-Crespo 2013a). This paper argues that crowdsourcing represents a prime example of an exciting new phenomenon that can help us assess and understand why the "gap" between both fields exists while simultaneously helping us to be more aware of possible synergies between both fields.

This study is partly motivated by the existing need in the industry to identify best practices for crowdsourcing in the rapidly developing world of crowdsourcing. Desilets and van de Meer indicate that “there is a clear need for a more concise, summative body of knowledge that captures recurrent best practices” (2011: 29). The authors also mention that current practitioners are the most suitable subjects for creating them: “we advocate the building of such a compendium, [...] which could be written collectively by practitioners of collaborative translation.” (Ibid: 29). Prescriptive collections of best practices can be found in the different publications by Desilets (Desilets 2011; Desilets and Van de Meer 2011), as well as work by DePalma and Kelly (2011), among others. It is often the case that the research gap rests on pressure from experts and professionals to convert research findings into applicable “how to” knowledge. Nevertheless, both perspectives do feed into each other. After all, prescriptive practices recommended in industry publications and existing case studies can help develop studies and testable hypotheses in the descriptive branch and develop theoretical models. Similarly, empirical and theoretical research can help shape best practices.

It should be mentioned at this point that according to the canonical map of TS as a discipline (Holmes 1984), research can fall under the Theoretical/Descriptive or the Applied branches. The latter branch focuses its attention on the work of professionals and practitioners, while the Theoretical/Descriptive branch is largely the realm of scholars and researchers. Both branches represent a global cycle in which the transfer of knowledge in both directions represents one of the main engines of evolution of the discipline and the production of knowledge about existing phenomena (Rabadán 2010). That is, both branches feed into each other and therefore help refine theories, models and applied practices. Obviously, the several stakeholders interested in the advancement of research (namely professionals and scholars), can have different objectives, tempos and research agendas, but both can and should cooperate towards a common goal.

The following sections review empirical research into crowdsourcing in TS and related disciplines and connect findings from this research with best practices recommended by the industry.

2. Empirical research into crowdsourcing

Empirical research into translation crowdsourcing has emerged mainly from two related perspectives, (1) Computational Linguistics / Machine Translation and from (2) TS. In the first case, research has focused on the development of workflows and models to harness the knowledge of the crowd (i.e. Shimoata et al. 2001; Morera-Mesa and Filip 2013), sometimes comparing professionals vs. crowdsourced translations to feed MT engines (Zaidan and Calliston Burch 2012). Empirical research in TS has mainly focused on two main research questions: motivation of volunteers to participate in translation initiatives and translation quality in crowdsourced texts. Both research issues, motivation and quality, also appear to be primary concerns in industry publications. For example, Desilets and Van de Meer (2011) indicate in their collection of best practices that emerged after a TAUS 2011 meeting that “[m]otivation issues are most critical in crowdsourcing scenarios, and this is possibly the main reason why it has yet to become widespread.” (p.32). Similarly, DePalma and Kelly indicate that it is necessary to discover volunteer motivations, and organizations need to “keep nurturing them with rewards and incentives” (2011: 401). Motivation also predominantly appears in Common Sense Advisory

publications (DePalma and Kelly 2008; Kelly and Ray 2011). For its part, and despite the enormous importance of translation quality, crowdsourcing quality seems to be less of a concern in industry publications than motivation:

Quality Control issues tend to resolve themselves, provided that enough of the “right” people can be enticed to participate and that you provide them with lightweight tools and processes by which they can spot and fix errors. (Desilets and van de Meer 2011: 41).

In any case, TS studies have also paid less attention to quality issues in crowdsourcing, with a lower number of empirical studies on this topic as the following sections will show.

3. Empirical Research into motivation in TS

Since 2010, a growing number of empirical studies have appeared in TS related to the motivation of volunteers. Theoretically, these studies have mostly departed from what is known as “sociological approaches” to translation (Wolf 2010). According to Chesterman topics of interest in sociological approaches relevant for crowdsourcing research are “the social role of the translators and the translators’ profession, translation as a social practice...” (2007: 173-174). The main research questions that have been the object of empirical inquiry have been (1) what are the motivations of volunteers, (2) what are their profiles? and (3) how are these volunteers organized? (Orrego-Carmona 2012). This section will focus on the results of volunteer motivation to participate in crowdsourcing initiatives.

The methodologies for these studies are mostly interventionist ones such as online surveys. Studies have focused on motivation to participate in Wikipedia (McDonough Dolmaya 2012), Facebook (Dombek 2013; Mesipuu 2012), TED open translation initiative (Camara forthcoming), Skype (Mesipuu 2012), or non-profits such as the Rosetta Foundation (O’Brian and Schäler 2010). The following table summarizes the studies, initiatives and the number of respondents in the surveys.

Researcher(s)	Initiative	N. of subjects in survey
O’Brien and Schäler (2010)	Rosetta Foundation	139
Mesipuu (2012)	Facebook and Skype	10 each (20 total)
McDonough-Dolmaya (2012)	Wikipedia	75
Dombek (2013)	Facebook / Poland	19 + 20
Camara (forthcoming)	TED	177

In order to compare the results of these studies with the best practices in the industry, the results from these studies were summarized and critically analyzed. Even when all the above-mentioned studies depart from slightly different perspectives and different theoretical foundations, it was possible to identify similar formulations of survey questions and underlying motives. Most studies separate between two basic notions in existing theories of motivation: the fact that they can be intrinsic or extrinsic (Frey 1997). Intrinsic motivations are those related not to financial compensation or reward but rather to personal enjoyment or a feeling of obligation to a specific community. Examples of this motivation type in the survey questions in studies are “Found the project intellectually stimulating” or “Help make information available in other

languages". Extrinsic motivations are related to direct or indirect rewards, such as gaining more clients, getting presents or the potential to attract customers.

Comparability was somewhat made difficult by the (1) different measuring scales, such as the likert scale of the O'Brien and Schäler (2010) compared to the multiple choice options in McDonough Dolmaya (2012), or (2) the differences in formulation of the potential motives. It was decided for comparability purposes, to rank the results from all studies numerically and then subsequently aggregate and compare them. The outcome of this analysis yielded three tiers or groups of motivations, from the first tier of motivations, namely those that consistently seem to appear at the top of most studies, to those less important for the volunteers. The first group or main tier of motivations includes exclusively intrinsic motivations such as:

1. Making information in other languages accessible to others.
2. Helping the organization with their mission or a belief in the organizations' principles.
3. Receiving intellectual satisfaction, probably related to the notion of 'cognitive surplus' (Shirky 2010).

The second tier as reported by participants includes both intrinsic and extrinsic motivations:

4. The desire to practice the second language.
5. The need to gain professional translation experience or increase one's reputation

Finally, a range of other motivations that appear consistently at the lower end of the results are:

6. The desire to support less known languages.
7. The satisfaction of completing something for the good of the community.
8. For fun.
9. The desire to be part of a network.

To some extent, it is surprising that the community component of this participation, that is, being part of a network, tends to be at the bottom of the motivations reported by users. This was the main motivation finding in the study of the close community of Skype volunteer translators (Mesipuu 2012). This finding may point to different motivations in cases of open or closed translation communities. It should also be kept in mind that all studies, in tune with findings about motivations in other crowdsourcing and volunteering areas, have concluded that a combination of motives, rather than any single one, underlies volunteer motivation. According to Olohan "volunteers are often motivated by a combination of factors and can be seen as behaving simultaneously altruistically and egoistically" (2014: 19). In any case, the only study that separated between professional and non-professional translators, that of McDonough-Dolmaya, identified that the main difference between both populations is the greater significance of extrinsic motivations for translation professionals, i.e. reputation, attract clients, etc. Another difference of interest between professionals and non-professionals, even when professionals make around 7 to 16 % of volunteers in the studies on Wikipedia and TED talks, is that professionals are also more attracted to initiatives that they perceive to have "greater cultural or symbolic values" or "more prestigious activities" (McDonough-Dolmaya 2012: 188). This brings up the question of the role of professionals in these initiatives. It is often the case that best practices reports indicate

that professionals should be involved where needed (DePalma and Kelly 2011), both by conducting in-house reviews or attempting to motivate them to volunteer. In this case, only those initiatives perceived by professionals with this higher symbolic value or prestige will be able to attract them as uncompensated participants. As the case of LinkedIn showed, requesting exclusively professionals to participate can backfire, but certain filters such as exams or evaluations can help bring to the initiative participants with a sufficient degree what is known as “professional expertise”.

4. Motivation in best practices publications

As previously mentioned, the publications by Common Sense Advisory (i.e. DePalma and Kelly 2009, 2011) and those resulting from the TAUS 2011 meeting (Desilets 2011; Desilets and van de Meer 2011; Collaborative Translation Patterns 2011)¹ are examples of prescriptive best practices developed within the industry. Both publications include a list of similar areas. The TAUS report includes a compendium of the most commonly used decision-making patterns, previously identified issues in the implementation of crowdsourcing during a meeting with industry experts and recommendations for how to best tackle each of them. The areas of interest include the following sections:

1. Planning and scoping
2. Community Motivation
3. Translation Quality
4. Contributor Career Path
5. Community right -sizing
6. Tools and Processes

As far as the Community motivation is concerned, industry experts suggest that these issues can be potentially be solved fostering of intrinsic motivations (12 recommendations out of 13), while only one of them relates to handing out branded products. Only two studies asked translators whether they would be motivated if gifts were handed out. In the case of Skype (Mesipuu 2012), it was found that community events, getting together in beta releases, was a more powerful motivator according to participants than handing out t-shirts or other merchandise. Similarly, in the study for the non-profit Rosetta foundation, merchandise, gifts or monetary compensation came at the bottom of the list in the survey, while intrinsic motivations such as feedback from qualified translators or clients, as well as invitations to events were reported twice as often as free gifts or payments. Additionally, subjects indicated that the least attractive incentives to motivate them in the future were practices such as translator of the month profiles or monthly top-ten lists. Practically all best-practice reports include these types of incentives and to some extent this finding contradicts these recommendations in industry publications. For example, DePalma and Kelly identify the main incentive to motivating participants is to “keep nurturing them [volunteer translators] with rewards and incentives...Something as simple as a certificate can be a powerful form of recognition” (2011: 403). They also indicate the value of “Highlight[ing] and showcase[ing] member contributions.

¹ (<http://collaborative-translation-patterns.wiki4us.com/tiki-index.php>)

Companies in the article find the “leaderboard” to be an effective tool”.² However, the study by O’Brien and Schäler (2010) also found out that top lists could be somewhat detrimental to the engagement of “lurkers”, those with little time to volunteer (Nielsen 2006)³. The authors indicated that:

Some volunteers [...] mention factors that would demotivate them. In particular, turning their activities into a competition by making them bid against each other or simply compete for positions on leadership boards was highlighted... (n.p.)

It is therefore necessary to put our fingers on the existing discrepancies between initial industry practices to motivate volunteers and the opinions of those participating in motivation-related surveys. The practice of including a leaderboard might, nevertheless, be recommended in the industry since it seems directed towards recruiting and retaining the low percentage of highly active participants that volunteer beyond the average weekly average of 2 to 5 hours identified in studies (McDonough-Dolmaya 2012; Camara forthcoming).

5. What about empirical studies on crowdsourcing quality?

The issue of quality in crowdsourcing has been the focus of a number of studies in Machine Translation with the goal of feeding MT engines (i.e. Yan et al. 2014; Zaidan and Callison-Burch 2012). Within TS, two empirical studies have focused on aspects related to quality of crowdsourced translations (Jimenez-Crespo 2013b, forthcoming). The research objectives of these studies are to identify whether crowdsourcing can produce ‘naturally’ sounding translations and whether the actual crowdsourcing process has an impact on the result.⁴ In Jimenez-Crespo (2013b) a corpus based approach to research crowdsourcing was taken to research whether the Spanish version of Facebook included the most conventional terminology and phraseology that appeared more frequently in non-translated Spanish social networking sites. The non-translational section of the corpus included all the interactive and navigation segments in the 25 most popular social networking sites in Spain in 2009. The results showed that the localized version of Facebook included the most terminology and phraseology identified in the non-translated sites. This study therefore confirmed that the translation workflow used by this company, as with many other TS theoretical proposals over the years (Jimenez-Crespo 2011), is effective in order to achieve texts similar to non-translated ones. The study also concluded that this method is more effective for producing a localized website that resembles non-translated ones than the average professional localization process.

The best practices repository and the workflow studies by Morera-Mesa (Morera-Mesa and Filip 2013; Morera Mesa 2014) document a range of possible procedures employed to guarantee

² DePalma and Kelly (2011) indicate also that it is necessary to remember that all volunteers do not have the same amount of time and it is necessary to recognize them all.

³ According to Nielsen 90% of participants in crowdsourcing efforts are ‘lurkers’ who never contribute, while 9% contribute a little and 1% of participants account for all of the activity.

⁴ Dombek (2013) in her study on the interaction between the Facebook Translate platform and volunteers concluded that subjects expressed the actual configuration of the process was detrimental to their translation process.

quality in crowdsourced translations.⁵ Many of them depend on the type of workflow in place. For example, organizations such as Kiva include entry exams, while many others apply automatic reputation management such as Cucumis. The same occurs when translation alternatives are open or closed, such as Asia Online (Morera and Filip 2013). In any case, the implications of the above mentioned studies for the industry point at two facts of interest: (1) Provided that enough volunteers participate in the open voting process implemented by Facebook, translations can result in more natural texts similar to those spontaneously produced in-country, and (2) the process implemented to produce the translations does have an impact on the final configuration of the target texts (Jimenez-Crespo 2013b). In this sense, more research should be conducted to compare different crowdsourcing workflow models and their resulting translations, such as open or closed alternatives followed by voting, translation plus in house revision, publish then revise, etc.

6. Conclusions

The uncertainty of working with volunteers and therefore the dependency on their motivation present interesting challenges that both the industry and TS are attempting to uncover. This paper has argued that this area presents an interesting case for bringing together industry and TS research since both interested parties are working in interrelated areas of a global research cycle, applied and prescriptive vs. theoretical and descriptive, and both feed into each other. Desilets and van de Meer concluded in their paper that “most practices [...] are not that different from best-practices which are being used for crowdsourcing in other domains” (2011: 41), and ask whether translation requires a set of best-practices. However, research by McDonough-Dolmaya (2012) identified clear differences in the motivations and types of participants if translation crowdsourcing is compared with studies in Free and Open Software. Other studies have identified a great variety of existing workflows that apply exclusively to crowdsourced translation (Morera-Mesa 2014). This means that the potential to research and identify best practices through the global cycle of research, whether it starts in the applied or the theoretical-descriptive side is wide open. It is hoped that this paper helps bridge the gap between the industry and academic research and starts to provide a foundation for joint research projects in this fascinating area.

References

- CAMARA, L. (forthcoming). “Motivation for Collaboration in TED Open Translation Project”, *International Journal of Web Based Communities*.
- CHESTERMAN, A. (2007). “Bridge concepts in Translation Sociology”. In M. Wolf, Michaela and A. Fukari (eds.), *Constructing a Sociology of Translation*. Amsterdam-Philadelphia: John Benjamins, pp. 171–183.
- DEPALMA, D. AND N. KELLY (2008). “Translation of, for and by the people”. Common Sense Advisory. Online. Available:
http://www.commonsenseadvisory.com/Portals/_default/Knowledgebase/ArticleImages/081218_R_community_trans_Preview.pdf

⁵ The TAUS report includes: Entry Exam, Peer Review, Automatic Reputation Management, Random Spot-Checking, In-House Revision, Users as Translators, Voting, Transparent Quality Level, Publish then Revise, Refining Redundancy, Open Alternatives, Hidden Alternatives

- DEPALMA, D. AND N. KELLY (2011). "Project Management for Crowdsourced Translation: How User-Translated Content Projects Work in Real Life". In K. Dunne and E. Dunne (eds), *Translation and Localization Project Management: The Art of the Possible*. Amsterdam-Philadelphia: John Benjamins, pp. 379-408.
- DESILETS, A. (2011). "Wanted: Best Practices for Collaborative Translation: A TAS report". Online. Available: <https://www.taus.net/downloads/finish/56-public-reports/482-wanted-best-practices-for-collaborative-translation>.
- DÉSILETS, A. AND VAN DER MEER, J. (2011). "Co-creating a repository of best-practices for collaborative translators". In O'Hagan, M. (ed.) *Linguistica Antverpiensia New Series – Themes in Translation Studies. Translation as a Social Activity*, 10/2011, pp. 27-45.
- DOMBEK, M. (2013). *A Study into the Motivations of Internet Users Contributing to Translation Crowdsourcing: The Case of Polish Facebook User-Translators*. Unpublished doctoral dissertation, Dublin City University.
- DUNNE, K. (2006) "A Copernican Revolution", in Dunne, K. (ed.) *Perspectives on Localization*. Amsterdam-Philadelphia: John Benjamins, pp. 1-11.
- FREY, B. (1997). *Not just for the money. An Economic Theory of Personal Motivation*. Cheltenham and Brookfield: Edward Elgar Publishing.
- HOLMES, J. (1988). "The Name and Nature of Translation Studies". In J. Holmes (ed.), *Translated! Papers on Literary Translation and Translation Studies* (pp. 67-80). Amsterdam: Rodopi.
- JIMÉNEZ-CRESPO, M. A. (2013a). *Translation and Web Localization*. New York/London: Routledge.
- JIMÉNEZ-CRESPO, M. A. (2013b). "Crowdsourcing, Corpus Use, and the Search for Translation Naturalness: A Comparable Corpus Study of Facebook and Non-Translated Social Networking Sites". *TIS: Translation and Interpreting Studies* 8: 23-49.
- JIMÉNEZ-CRESPO, M. A. (2011). "The future of general tendencies in translation: Explicitation in web localization". *Target*, 23 (1): 3-25.
- MCDONOUGH DOLMAYA, J. (2012). "Analyzing the Crowdsourcing Model and its Impact on Public Perceptions of TRANSLATION". *THE TRANSLATOR*, 18(2): 167-191.
- MESIPUU, M. (2012). "Translation crowdsourcing and user-translator motivation at Facebook and Skype". *Translation Spaces* 1:33-53.
- MORERA-MESA, J.J. (2014). *Crowdsourced Translation Practices from the Process Flow Perspective*. Doctoral Dissertation, University of Limerick, Ireland
- MORERA-MESA, J.J. AND D. FILIP. (2013). "Selected Crowdsourced Translation Practices". *ASLIB Translating and the Computer* 35, 2013.
- NIELSEN, J. (2006). "The 90-9-1 Rule for Participation Inequality in social Media and Online Communities". Online. Available: <http://www.nngroup.com/articles/participation-inequality/>
- O'BRIEN, S. AND SCHÄLER, R. (2010). "Next Generation Translation and Localization. Users are Taking Charge", *Translating and the Computer Conference*, 17-18 November 2010, London. Online. Available http://doras.dcu.ie/16695/1/Paper_6.pdf.

- O'HAGAN, M. (2013). "The Impact of New Technologies on Translation Studies: A Technological Turn?" In C. Millán-Varela and F. Bartrina (ed), *Routledge Handbook of Translation Studies*. London: Routledge, pp. 503-518.
- OLOHAN, M. (2014). "Why do you Translate? Motivation to Volunteer and TED translation". *Perspectives: Studies in Translatology* 7: 17-33.
- ORREGO-CARMONA, D. (2012). "Internal Structures and Workflows in Collaborative Subtitling". Paper delivered to the *First International Conference on Non-professional Interpreting and Translation*. Università di Bologna, May 17-19. Online. Available: http://isg.urv.es/publicity/doctorate/research/documents/Orrego/Orrego-Carmona_Structures-Workflows_NPIT1.pdf
- RABADÁN, R. (2008) "Refining the idea of 'applied extensions' ". In A. Pym, M. Shlesinger and D. Simeoni (eds), *Beyond Descriptive Translation Studies*, Amsterdam-Philadelphia, John Benjamins, pp. 103-117.
- RAY, R. AND N. KELLY (2011). *Crowdsourced Translation: Best Practices for Implementation*. Lowell, Ma: Common Sense Advisory.
- SHIMOATA, S. ET AL. (2001). "Collaborative Translation Environment on the Web". *Proceedings from Machine MT Summit 8*, pp. 331-334.
- SHIRKY, C. (2010). *Cognitive Surplus: Creativity and Generosity in a Connected Age*. New York: Penguin Press.
- YAN, (2014). "Are Two Heads Better than One? Crowdsourced Translations via a Two Step Collaboration of Non-Professional Translators and Editors". *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1134-1144. Baltimore, Maryland, USA, June 23-25 2014.
- ZAIDAN, O. F AND CALLISTON-BURCH, C. (2011). "Crowdsourcing Translation: Professional Quality from Non-Professionals". *Proceedings of the 49th Annual Meeting of the Association of Computational Linguistics*, pp. 1120-1129.