

# Evaluation of English to Spanish MT Output of Tourism 2.0 Consumer-Generated Reviews with Post-Editing Purposes

**Miguel Ángel Candel-Mora**  
Universitat Politècnica de València  
mcandel@upv.es

## Abstract

Consumer-generated reviews (CGR) entail a significant potential business volume in terms of translation and post-editing, however travel review platforms usually rely solely on raw machine translation. As a new digital genre, CGR require specific post-editing guidelines, therefore, this paper focuses on the analysis of a corpus of Spanish machine translation output of hotel reviews in order to identify error patterns and their effects on quality with the aim of designing a post-editing strategy adapted to this particular type of text.

## 1 Introduction

Internet users have evolved from being passive observers to active participants in Web 2.0. According to studies (Schemmann, 2011) on consumer-generated reviews (CGR), seven in every ten Internet users worldwide trust consumer opinions and peer recommendations posted online. Likewise, according to the most recent statistics published by the Spanish Tourist Movement Survey (Familitur, 2013) of the Spanish Institute of Tourism Studies, Internet use increased over 29%: almost all users (99.2%) used it to search for information, 76.5% to make a reservation and 52.4% for payment of services.

Despite this significant potential business volume, travel review platforms usually rely solely on raw machine translations of consumer reviews without further processing or revision, therefore this paper focuses on the analysis of a corpus of machine translation output of hotel reviews in order to identify error patterns and their effects on text quality with a view to implement a post-editing strategy. This study is part of the ProjecTA research project funded by the Spanish Ministry of Economy and Competitiveness (FFI2013-46041-R), aimed at exploring the effects of the implementation of MT-related services on the professional profile of translators.

More specifically, the objectives of this paper are twofold: to define the characteristics of this new digital genre to determine its level of text quality and acceptability, and identify and classify error patterns.

In order to reach these objectives, a corpus of one hundred user reviews originally written in English was compiled from TripAdvisor, the leading online travel review platform in terms of use and content available that operates in 45 countries and in 28 languages. Currently, TripAdvisor stores more than 200 million reviews and opinions from travelers around the world on more than 4.5 million businesses and properties in more than 147,000 destinations.

This research work is structured in three parts: firstly, it briefly approaches the literature on consumer-generated reviews in order to identify their characteristics and pragmatic purpose and consequently suggest the need to implement new methods of analysis to reflect and look into its distinctive features. Secondly, upon clearly defining the conventionalized patterns and

textual artifacts of consumer reviews, the focus turns on the quality and evaluation of their machine translation output to suggest a post-editing strategy that would best suit this new digital genre. Finally, the analysis and discussion section illustrates with examples from the corpus the most effective post-editing strategies to increase readability, reliability and quality aspects of consumer reviews generated by machine translation.

## **2 Consumer-Generated Reviews as a New Digital Genre**

From the point of view of discourse analysis, recent years have witnessed a shift in the approach to the study of digital genres mainly due to the emergence of new platforms and communication forms: after the appearance of email and blogs as digital genres per se, the real expansion started with social networking sites and the active participation of Web 2.0 users with consumer-generated content and product reviews.

Tourism 2.0 consumer-generated reviews have thus opened new lines of research for linguists: from the approach to specialized terminology and new text types, to the influence of the translation of tourism 2.0 on the target language, and the paradigm shift in the translation model: the active participation of the user in the translation process.

The main criteria in the definition of a genre include the existence of a shared set of communicative purposes (Swales, 1990) and its conventionalized textual artifacts “in the context of specific institutional and disciplinary practices” within a specific discourse community (Bathia, 2002: 6). Research work on online reviews is relatively new, as evidenced by the variety of designations found in the literature: “electronic word of mouth” or “eWOW” (Pollach, 2006), “online consumer reviews” (Vásquez, 2012) “user generated product reviews”, “product reviews” or “user opinions” (Ricci & Wietsma, 2006), to refer to the evaluation of users posted on a travel review site on their experience. According to the definition by Ricci & Wietsma, (2006: 297): “Product reviews can be described as a subjective piece of non-structured text describing the user's product knowledge, experiences and opinions, together with a final product rating.”

With regards to research lines, Vásquez (2014) states that online reviews have been studied in fields such as marketing, economics, tourism, computing and information sciences. Research topics range from the potential roles of product reviews in the decision process (Ricci & Wietsma, 2006), the involvement of reviewers (Vásquez, 2014) and the characterization of online reviews (Shemmann, 2011), to the improvement of review websites (Pollach, 2006).

Different authors (Pollach, 2006; Vásquez, 2012) confirm the existence of a new digital genre with special characteristics, and highlight the lack of research on online consumer reviews from the linguistic point of view, probably because this type of texts did not exist previously in written format as they were transmitted orally and without a specific structure. However, with the emergence of travel review sites, but primarily due to the large amount of comments and reviews posted online by users in recent times, it can be regarded as a digital genre in its own right.

Schemmann (2011) identifies twelve different types of presentation for CGR classified into three broad categories: (1) service evaluation functions, (2) feedback and interactive functions and, (3) matching and search performance functions. Service evaluation includes free-style text and structured text –the most common in travel review sites; ratings, where overall performance can be rated on a scale; pictures and videos, review summaries and trend analysis. The other two broad classifications (2 and 3), focus on feedback of readers in forums and communities, or on the integration of reviews and ratings from other platforms, and therefore not so much based on textual resources, and thus beyond the scope of this study.

This research work concentrates on CGR within the service evaluation functions and more specifically on reviews provided by means of free-style text and structured text, with different styles and lengths.

From the literature consulted on CGR as a digital genre, the research conducted by Pollach (2006) and Vásquez (2012) provide the most valuable guidance as to the structural text features, analysis methodologies and identification of the most remarkable characteristics and resources to connect reviewers with their readers.

Vásquez's work (2012) focuses on involvement and the resources that authors use to engage their audience in their narratives based on a corpus of negative comments exclusively. On the other hand, Pollach (2006) proposes the improvement of consumer opinion web sites upon an extensive analysis of 358 product reviews from an online product forum. Their insights and research framework proved extremely useful and paved the way for the design of the research methodology used of this paper.

According to Vásquez (2012: 109) due to the extended temporal experience of staying in a hotel, reviews are usually written in a chronological sequence of events that follow a linear narrative structure in eight phases: ranging from the planning phase, the first encounter with the room, to check out and follow up communication with hotel. From these eight phases, reviewers are selective and only include in their reviews a discussion of some of them. Therefore, for Vásquez (2012) the main structural text features are summary, background (reason to travel, travel companion ...), explicit evaluation, interactions with hotel staff, resolution (check-out/cost) and personal advice, suggestions or warnings. However, this author adds that for this opinion to be reliable and have credibility, some features associated with involvement in discourse must be taken into account: reported speech, story prefaces, deictic shifts, which are ultimately responsible for the connection among participants. Among the resources used to engage with their readers, reviewers make use of humor, detail and personal experience. Finally, Vásquez (2012: 107) acknowledges the constraints of carrying out this type of research relying solely on language, since there are other nonlinguistic cues that also play an important role.

With a similar approach, through corpus linguistics techniques and textual analysis, Pollach (2006) also refers to the importance of the rules and conventions established by the genre community and focuses her work on the analysis of structure, content, audience appeals, sentence style, and word choice.

The definition of the pragmatic purpose in consumer reviews is especially interesting for this research work, which according to Pollach (2006: 3): "...is to inform potential buyers of the strengths and weaknesses of consumer products." Thus, the key is to share an experience that can help other users make decisions and that on many occasions the reviewer becomes a kind of expert on the matter based on features such as credibility and expertise. In the same vein, Vásquez (2012: 111) states that "the main purpose of online consumer reviews is to rate, evaluate, describe, and, on that basis, to provide recommendations to others for or against a particular product or service."

Finally, other genre-specific features include intertextuality – or reference to previous comments, the personal profile of the reviewer and paralinguistic elements, mainly "orthographic strategies designed to compensate the impersonality of written discourse" (Pollach, 2006: 8) such as capitalization, spelling, and punctuation. Most probably, here lies the key to the reliability and credibility of consumer reviews, i.e., how to express the emotions and emphasis that the MT output cannot convey. Among the elements that Pollach (2006) notes are emoticons, the use of capital letters, overuse of punctuation marks and acronyms. However, Pollach also insists that the use of non-verbal cues was not too common

in the corpus perhaps because reviewers take their tasks seriously, and use a neutral, non-emotive language.

### **3 Quality, Evaluation of Machine Translation and Post-editing**

Different variables determine the approach to the assessment of quality of machine translation output and it is very complex to find common ground that serves as a starting point for proposing universal quality evaluation criteria. Most authors highlight that quality is conditioned by the purpose of the MT product (Allen, 2003; TAUS, 2010b), i.e., if the translation is intended to be published and disseminated, or if the translation is only aimed at guiding the reader on its overall meaning.

Post-editing is not a new phenomenon, what is new is machine translation technology and the new types of digital genres that have emerged with the evolution of information and communications technology, exemplified by the emergence of social networks and Web 2.0 user participation.

On the other hand, the development of machine translation technology, especially since the advent of corpus-based statistical machine translation systems, has resulted in varying degrees in the quality of MT output: the quality of the output of recent machine translation engines is substantially improved as the size of the corpus increases.

Research on MT post-editing has been approached from different points of view: quality (Aramberri, 2014; Koby et al., 2014; Specia et al., 2010), evaluation guidelines (Babych, 2014), productivity gain (O'Brien, 2011), cognitive effort (O'Brien, 2005; Porro et al. 2014), the acceptability of MT output (Gorög, 2014), or a combination of strategies such as pre-editing and use of controlled languages to improve translatability (Temnikova, 2010).

Given the novelty of this field, there exists a limited number of methodologies and criteria on how to train post-editors or perform post-editing tasks, and frequently internal post-editing criteria are not accessible for confidentiality reasons, which hinders the possibility of a more general overview on existing post-editing guidelines. All this leads us to reflect on the changing nature of post-editing, and the obstacles to propose a universal tool applicable in any context.

Allen (2003: 300) quite accurately depicts the use of MT in the context of Web 2.0 and user participation as he notes that in recent years there is a “change in expectations with regard to the type and quality of translated material.” Traditionally, translation was considered a high quality text product for important documents on user safety or commercial information, for example, but currently there is an increased demand for gisting translation, users just need to understand the main idea of the text in their own language.

With regards to post-editing levels, there are different factors like the specifications of the client, the volume of documentation expected to be processed, or the expectations with regard to the level of quality for reading the final draft of the translated product, among others (Allen, 2003: 301). In sum, each case is different and should be studied individually as “differing percentages of MT accuracy have even been found when applied to different subdomains and different document types within the same technical domain” (Allen, 2003: 303) which corroborates the initial hypothesis of our work on the need to study in detail the characteristics of each text genre and develop customized post-editing guidelines accordingly.

In general, Allen (2003) distinguishes two types of translation activities: inbound or outbound, depending on whether it is translation for assimilation (inbound) or translation to be disseminated and published (outbound). Thus, for each type of MT post-editing he distinguishes different levels ranging from “no post-editing” (gisting) to “rapid post-editing”,

for restricted circulation documents. This minimal editing focuses exclusively on eliminating flagrant or important errors, and stylistic aspects are not taken into account. For the second type, outbound, he also distinguishes “zero post-editing”, “minimal post-editing” and “complete post-editing”. Therefore, the problem is how to quantify the amount of post-editing. What seems to be clear is that the typical human translation editing workflow process is completely different from a machine translation post-editing process, and research efforts should aim at developing post-editing methodologies and training actions.

Finally, with reference to post-editing guidelines and criteria, Allen (2003: 306), highlights the lack of concrete data on specific post-editing criteria, linguistic categories to be revised, or quality control scales used, among others, possibly due to the fact that most post-editing guidelines are for internal use, company-specific and proprietary, and cannot be disclosed, or they refer to specific translation systems and therefore not applicable to the rest of MT systems: post-editing guidelines vary whether they are oriented to a rule-based or statistical machine translation system, or to a hybrid system.

According to the literature consulted (Guzmán, 2007; Mitchell et al., 2014; SAE International, 2001; TAUS, 2010b), among the most common categories of errors are terminology errors, lexical ambiguity, syntax, omission, word agreement error or punctuation errors; and depending on the type of metrics used, with different weights for each error. However, the literature seems to emphasize that in addition to specific grammar and lexical criteria there are general criteria such as readability and acceptability of MT output, but especially if the objectives of the text type are met (TAUS, 2010a; Stymne and Ahrenberg, 2012).

In this same line, Mitchell et al. (2014) propose three quality evaluation methods: an error annotation, evaluation of fluency and fidelity by domain specialists, and evaluation of fluency by community members. For our work, the contribution by Mitchell et al. (2014) is particularly interesting because it advocates the need to implement new assessment methods to the new paradigm of user-generated content. In the research work carried out by Mitchell et al. (2014) on community post-editing, the types of error categorization considered were: accuracy errors (additional information, missing information, untranslated information, mistranslated information) language errors and format errors.

As in the other authors consulted, the starting point for TAUS guidelines (2010b) lies in the impossibility of developing a set of guidelines that apply to all scenarios. TAUS (2010b) also distinguishes two levels of post editing determined by two main criteria: the quality of the MT raw output and the expected end quality of the content. These levels are “good enough” quality, and quality “similar or equal to human translation”. TAUS “good enough” level is defined as comprehensible and accurate but not very convincing with respect to style.

Finally, Vilar et al. (2006) propose another classification of errors and acknowledge that this is a controversial and unambiguous task. However, they propose a hierarchical structure in which the first level includes the following five major classifications: missing words, word order, incorrect words, unknown words and punctuation errors.

#### **4 Analysis and Discussion**

Thus, having concluded that there is no universal post-editing strategy and MT output quality-assessment scales cannot be used directly on any type of text, this paper attempts a novel approach which consists in the design of a classification of errors based on the observation of error patterns identified after a manual revision by expert linguist of the Spanish MT output of a corpus of 100 hotel reviews.

One of the reasons for proposing specific PE guidelines is based on the nature of the reviews. Common metrics of translation quality include error annotation and calculation of proportion of errors with the total amount of words in the translated text, however in the case of consumer reviews, with an average of 144 words per review (See Table 1), the error proportion would be higher and low quality translation would be more noticeable.

In addition to this error annotation proposal, the specific features of online consumer reviews of hotels such as involvement or credibility/expertise of reviewer, intertextuality, structural text features and paralinguistic features are specially taken into consideration during the design of the PE strategy.

The methodology followed in this work can be summarized in three steps: design of a corpus of CGR of reference to validate the genre characteristics and perform PE tasks, manual PE of Spanish MT output by expert linguist, and identification of recurrent errors and correspondence with digital genre features.

The corpus is composed of 100 consumer-generated reviews with a total of 14,528 words in English and 14,818 words in the Spanish MT output. Reviews were selected for the following criteria: originally written in English, written about the same hotel, and posted online on TripAdvisor during the period January-June, 2015. Only reviews originally written by native speakers of English were selected. This was determined first, by the place of origin of the reviewers (UK, USA and Australia) and then, by the degree of linguistic accuracy of the texts. In order to obtain representative data of this textual genre in Spanish, a small reference corpus was compiled with all the reviews that were written originally in Spanish during that period on the same hotel, and posted on TripAdvisor: a total of 34, totaling 1,532 words. This corpus of reference would help to compare the results obtained from the analysis of the Spanish MT output corpus, with what is found naturally in reviews originally written in Spanish.

	English corpus	Spanish MT output corpus	Spanish reference corpus
Average review length	144	146.72	69.63
Longest review	424	420	228
Shortest review	38	42	31
Average sentence length	17.46	17.70	15.78
Longest sentence	58	57	60
Shortest sentence	2	2	1

Table 1. Average number of words and sentence length in reviews.

At first sight, the length of reviews (see Table 1) is very similar in English and Spanish, which contrasts with the analysis of the reference corpus originally written in Spanish, with an average of 69.63 words per review.

The corpus of reviews was then fragmented into sentences and aligned with their corresponding Spanish MT output to facilitate manual revision. During the first stage, aligned segments were labeled as *unacceptable* (message not accurate due to incorrect grammar or lexical usage, unusual syntax or due to mistranslation), *acceptable* (accurate but not fully convincing or with minor errors) and *correct* (without any error). As Table 2 shows, only 183 (22%) segments were labeled as unacceptable.

correct	acceptable	unacceptable	total segments
305	324	183	812

Table 2. Initial classification of MT output quality.

Secondly, the first two categories, *unacceptable* and *acceptable*, underwent a second thorough revision work to identify specific recurrent error patterns. In order, to facilitate data processing, errors were grouped in two categories: 1) grammatical errors and 2) mistranslations. Within the first category, the following recurrent errors were identified: word agreement, use of articles, word order, verb tenses, and collocations and phraseology. The second type, mistranslations, included omissions, spelling mistakes in original, terminology issues, ambiguity, and problems concerning proper names and brand names.

Finally, revision also concentrated on verifying compliance with genre specific features of consumer generated reviews such as textual artifacts, intertextuality, structure and format, and paralinguistic elements.

#### 4.1 Error Pattern Identification in Consumer Reviews: Grammatical Errors

A total of 354 errors were identified within this category (see Table 3). Although some errors were not highly noticeable and sometimes did not affect comprehension of the text, the occurrence of several errors within the same sentence or within one review interferes to a large extent with the overall readability of the text and thus affects the main features of this type of text, namely reliability and credibility.

Category	Number of errors
Word agreement	99
Word order	57
Articles	53
Collocations and phraseology	49
Personal pronouns	43
Verb tense	37
Relative pronouns	8
Passive voice	8
Total grammatical errors	354

Table 3. Error pattern identification in consumer reviews: grammatical errors.

Word agreement is by far the most recurrent error, probably because it includes three different types of errors, plural vs. singular agreement in nouns, masculine vs. feminine in adjectives, and subject-verb agreement. Some examples to illustrate this are:

ST: ... a stay here is not **cheap**.

MT: ...una estancia aquí no es **barato**.

ST: We had two rooms and both were **perfect** in every way.

MT: Teníamos dos habitaciones y ambas eran **perfecto** en todos los sentidos.

ST: The hotel also **booked** theatre tickets for me.

MT: El hotel también **reservamos** billetes de teatro para mí.

Similar interference on readability is found in errors related to word order and use of articles. Although, the analysis of the causes of errors was beyond the scope of this research work, in the case of word-order errors, it was very noticeable that the main source of errors came from the attempt to translate structures in parallel, and the majority of word-order errors (34) were detected in sentences with a length of more than 20 words or sentences that were separated by commas or conjunctions. For the rest of errors, a specific pattern was not found. With regards to errors in articles, the most frequent error was found when the name of the hotel was used in the review, as in Spanish definite article is required.

ST: Hotel 41 has very good service.  
 MT: [E1] Hotel 41 tiene un muy buen servicio.

ST: Thank you all at 41.  
 MT: Gracias a todos en [e1] 41.

ST: It's London centre after all.  
 MT: Es [el] centro de Londres después de todo.

## 4.2 Error Pattern Identification in Consumer Reviews: Mistranslations

As shown in Table 4, most errors were caused by the incorrect handling of the MT system of ambiguous forms, which in some cases correspond to very frequent words found in hotel reviews, such as bar (establishment / counter / candy), play (sport / theater), ticket (train / theater), glass (receptacle / material), or in common English verbs that have two forms in Spanish, such as *to be*, *to have*, *to miss*, as shown in the examples below:

ST: I had selected a few **plays**...  
 MT: Había seleccionado **algunos juega obras de teatro**...

ST: ...they know what you have **had**.  
 MT: ...saben lo que hemos **tenido-tomado**

ST: My phone only **charges** with that charger  
 MT: mi teléfono sólo **cobra carga** con ese cargador

Category	Number of errors
Ambiguity	58
Terminology	45
Omissions	27
Proper names / brands	13
Spelling mistakes in original	9
Total mistranslations	152

Table 4. Error pattern identification in consumer reviews: mistranslations.

Authors like Vásquez (2012) note that reviewers construct their expertise through the use of specialized terminology, therefore accuracy in the use of specialized terminology should be regarded as essential in a PE strategy for CGR. In this research, hotel and catering industry terminology seems accurately translated when it appears in its standard form such as stay as *estancia*, lounge as *sala de estar*, room as *habitación*, check-in as *registro*, suite as *suite*. However, when these terms are used in combination with other words, errors are more frequent: “conservatory suite” was mistranslated as *la suite invernadero* or *el Conservatory Suite*; executive lounge had up to four different versions: *salón de ejecutivos*, *sala de estar ejecutiva*, *salón ejecutivo*, *Executive Lounge*.

Finally, it should be noted that in the case of proprietary hotel terminology, which sometimes appears in inverted commas, MT output reaches its lowest quality results:

ST: ... upgraded us to a **split-level suite**...  
 MT: ...nos pasaron a una **separación de niveles suite**...

ST: We booked a mid range room to splash out with the **Romantic Turn Down option**  
 MT: Reservamos una habitación de gama media **tira la casa por la ventana con la romántica por opción**

ST: I've been to other hotels with "**plunder the pantry**" style offerings...  
 MT: He estado en otros hoteles con "**latrocinio las ofrendas de estilo**" en la despensa



Some of these errors would be resolved if the MT system had a corpus of texts from the same area, or from a corpus of hotel reviews. However, in proprietary and differentiating hotel terminology mistranslations would still remain unsolved.

The degree of omissions found varies from cases where the meaning is completely altered to instances where only quantifiers are removed, without any effects on the final meaning.

ST: I **highly** recommend Hotel 41.

MT: Recomiendo el hotel 41.

Use of proper names and brand names in reviews clearly contribute to the expertise of the reviewer, however when translated into Spanish, two differentiated cases are found: contextual information is not given as in the case of the location of the hotel near Victoria [station], thus leading to poor MT output, or when the brand is unknown to target text reader, contributing to an even more confusing text:

ST: ...I asked the reception for a fine-nib **sharpie**.

MT: ... ~~pregunté a la recepción por un elegante incluía imprentas~~ **sharpie**.

One last phenomenon already mentioned in the PE literature is when mistranslations occur due to spelling mistakes in the source text. The MT solution is also different depending on the case, sometimes it omits completely the misspelled form; in a couple of cases it reproduces the same word as in the original, with the same spelling mistake (If you are looking for perfect refined service from **interetsing** people... *Si estás buscando el servicio refinado perfecto de personas **interetsing**...*) and in one case it fixes the problem and provides the spelling in Spanish (...and only mentioned it to the **consierge**... - *...y sólo se lo mencioné al **conserje**...*)

### 4.3 Compliance with CGR Genre-Specific Features

With a couple of exceptions, Spanish MT output of key structural artifacts such as evaluation, thanks, reference to other reviews and advice was outstanding, without any doubt due to the simple syntax used in these structural artifacts. As it can be concluded from the literature, credibility and reliability are essential features in CGR and the purpose of this genre basically focuses on evaluation of hotel experience and reviewer's advice, therefore post-editing guidelines for consumer reviews should prioritize that these artifacts do not look like they were generated by a computer, or at least contribute to reviewers' expertise with added fluency.

CGR specific features	Occurrences
Evaluation	65
Thanks	80
Advice	73
Reference to reviews - intertextuality	48
Paralinguistic features	5

Table 5. CGR genre-specific features

A key keyword analysis with Wordsmith Tools, revealed among its 30 most frequent keywords words such as wonderful (42), amazing (33), perfect (29), lovely (26) and excellent (22) and its Spanish equivalence in the analysis of the Spanish corpus: *especial* (36), *increíble* (32), *maravilloso* (25) *excelente* (20), *perfecto* (19), *encantador* (16).

Finally, there is a lack of paralinguistic features, probably because reviewers are careful not appear unprofessional. No emoticons or punctuation emphasis were found in the corpus, with the exception of the use of several exclamation marks common in digital genres. However, reminiscent of its oral origins, there are several instances of emphasis artifacts common in spoken language.

ST: **Amazing amazing** hospitality  
MT: hospitalidad **totalmente increíble**

ST: Everything is **so So SO** amazing.  
MT: Todo es **tan** increíble.

## 5 Conclusions

Consumer-generated content has become a powerful indication of customer satisfaction, therefore research to analyze this new digital genre would throw light on its peculiarities, especially in terms of improving MT output and contribute to current studies on MT post-editing.

MT quality evaluation has been studied for a while now and most authors seem to agree on one characteristic: MT quality is primarily determined by the purpose and use of the translated text. Likewise, post-editing is not new either, what is new is machine translation technology and the new types of digital genres that emerge as social networks and product review sites evolve. The main features of reviews revolve around reviewer's credibility and reliability, therefore the PE strategy should give priority to these features and their textual artifacts towards achieving a more natural language.

The decision on whether a more or less detailed post-editing effort should be appropriate depends on the use and purpose of the translated document. Thus, it should take into account the characteristics of textual genre and design a PE strategy accordingly. This strategy and the detailed analysis of the textual genre must be taken into consideration when training future post-editors in PE techniques and guidelines.

## Acknowledgments

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

## References

- Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and Translation. A translator's guide*. 297-317. Amsterdam/Philadelphia: John Benjamins.
- Aramberri, N. (2014). Posedición, productividad y calidad. *TradumàTica: Tecnologies de la Traducció*, 12, 471-477. Available at <http://revistes.uab.cat/tradumatica/article/view/n12-aranberri>
- Babych, B. (2014). Automated MT evaluation metrics and their limitations. *TradumàTica: Tecnologies de la Traducció*, 12, 464-470. Available at <http://revistes.uab.cat/tradumatica/article/view/n12-babych>
- Bhatia, V.K. (2002). Applied Genre Analysis: A Multi-Perspective Model, *Ibérica*, 4, 3-19.
- FAMILITUR. (2012). *Informe Anual 2012*. Instituto de Turismo de España. Available at [www.iet.tourspain.es](http://www.iet.tourspain.es)
- Görög, A. (2014). Quality evaluation today: the dynamic quality framework. in Proceedings of Translating and the computer 36. The International Association for Advancement in Language Technology. London: United Kingdom. 155-164.

- Guzmán, R. (2007). Manual MT Post-Editing: If it's not Broken, don't Fix it. *Translation Journal* 11(4). Available at <http://translationjournal.net/journal/42mt.htm>
- Koby, G., Fields, P., Hague, D., Lommel, A., and Melby, A. (2014). Defining Translation Quality. *TradumàTica: Tecnologies de la Traducció*, 12, 413-420. Available at <http://revistes.uab.cat/tradumatica/article/view/n12-koby-fields-hague-et-al>
- Mitchell, L., O'Brien, S., and Roturier, J. (2014). Quality evaluation in community post-editing. *Machine Translation* 28, 237-262.
- O'Brien, S. (2005). Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability. *Machine Translation*. 19, 37-58.
- O'Brien, S. (2011). Towards Predicting Post-Editing Productivity. *Machine Translation* 25(3), 197-215.
- Pollach, I. (2006). Electronic word of mouth: A genre analysis of product reviews on consumer opinion web sites, in *Proceedings of the 39th Hawaii International Conference on System Sciences*. IEEE Computer Society.
- Porro, V., Gerlach, J., Bouillon, P., and Seretan, V. (2014). Rule-based automatic post-processing of SMT output to reduce human post-editing effort. In *Proceedings of the 36th International Conference on Translating and the Computer*, London, United Kingdom. 66-76. Available at <http://www.mt-archive.info/10/Asling-2014-TOC.htm>
- Ricci, F., and Wietsma, R. (2006). Product reviews in travel decision-making. In M. Hitz, M. Sigala and J. Murphy (Eds.), *Information and communication technologies in tourism*. 296-307. Vienna: Springer.
- SAE International. (2001). *SAE J2450: Translation Quality Metric*. Society of Automotive Engineers. Warrendale, USA.
- Schemmann, B. (2011). A Classification of Presentation Forms of Travel and Tourism-Related Online Consumer Reviews. *e-Review of Tourism Research* 2. Available at <http://ertr.tamu.edu/enter-2011-short-papers>
- Specia, L., Raj, D. and Turchi, M. (2010). Machine translation evaluation versus quality estimation. *Machine Translation*. 24. 39-50.
- Stymne, S. and Ahrenberg, L. (2012). On the practice of error analysis for machine translation evaluation. in *Proceedings of the LREC 2012 Conference*. Istanbul: European Language Resources Association, 1785-1790. Available at [http://www.lrec-conf.org/proceedings/lrec2012/pdf/717\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/717_Paper.pdf)
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*, Cambridge: Cambridge University Press.
- TAUS. (2010a). *Error Typology Guidelines*. Available at <https://www.taus.net/academy/best-practices/evaluate-best-practices/error-typology-guidelines>
- TAUS. (2010b). *MT Post-editing Guidelines*. Available at <https://www.taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines>
- Temnikova, I. (2010). Cognitive Evaluation Approach for a Controlled Language Post-Editing Experiment. In *Proceedings of the LREC 2010 Conference*. Valletta: European Language Resources Association. 17-23. Available at [http://www.lrec-conf.org/proceedings/lrec2010/pdf/437\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf)
- Vásquez, C. (2012). Narrativity and involvement in online consumer reviews. The case of Tripadvisor. *Narrative Enquire*. 22:1 105-121.
- Vásquez, C. (2014). *Online consumer reviews*. New York: Bloomsbury.
- Vilar, D., Xu, J., d'Haro, L. F., and Ney, H. (2006). Error analysis of statistical machine translation output. In *Proceedings of the LREC 2006 Conference*. Genoa: European Language Resources Association, 697-702. Available at [http://www.lrec-conf.org/proceedings/lrec2006/pdf/413\\_pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf).