

Document downloaded from:

<http://hdl.handle.net/10251/112978>

This paper must be cited as:

Candel-Mora, MÁ.; González Pastor, DM. (2017). Subtitling for a global audience. Handling the translation of culture-specific items in TEDx talks. FORUM. Revue internationale d'interprétation et de traduction. International Journal of Interpretation and Translation. 15(2):288-304. doi:10.1075/forum.15.2.07can



The final publication is available at

<https://doi.org/10.1075/forum.15.2.07can>

Copyright John Benjamins Publishing Company

Additional Information

## **Subtitling for a Global Audience: Handling the Translation of Culture-Specific Items in TEDx Talks**

FORUM 15:2 (2017), 288–304. doi 10.1075/forum.15.2.07can

Miguel Ángel Candel-Mora - Universitat Politècnica de València

Diana María González-Pastor - Universitat de València

### **Abstract**

TED.com is a platform to share ideas through influential talks in video format on topics that range from science and technology to business that engages volunteers from all over the world to help transcribe, subtitle and translate their scripts in more than 100 languages. The justification to engage volunteer transcribers is that transcribed talks can reach a wider audience because they are accessible for hearing impaired individuals, can be indexed in search engines and can achieve TED's mission of spreading ideas by making transcripts available for translation through TED's Open Translation Project.

Therefore, talks transcribers play a crucial role in the overall translation workflow and dissemination process as they are responsible for transcribing the contents and foundations of what will be later on translated into different languages. The objective of this paper is to analyse a corpus of talks originally delivered in different variants of Spanish to identify the most common

strategies used by volunteer transcribers to handle local or idiomatic expressions and culturally biased items to reach the maximum audience possible and facilitate translation.

keywords: subtitling, translation, TEDx, localization, neutral Spanish

## **Résumé**

TED.com est une plateforme qui sert à partager des idées à travers des conférences influentes en format vidéo sur des sujets allant de la science et la technologie aux affaires, grâce aux traducteurs bénévoles du monde entier qui aident à transcrire, à sous-titrer et à traduire leurs scripts dans plus de 100 langues. Les transcribers volontaires s'engagent à transcrire les présentations afin que celles-ci puissent atteindre un public plus large, notamment la communauté malentendante, qu'elles puissent être indexées dans les moteurs de recherche et puissent également se conformer à l'objectif de TED, qui consiste à propager des idées en rendant les transcriptions disponibles pour traduction au travers de TED Open Translation Project.

De ce fait, les transcribers jouent un rôle essentiel dans le processus global de traduction, car ils sont chargés de transcrire le contenu et les fondements de ce qui sera plus tard traduit en différentes langues. Dans cet article, on se propose d'analyser un corpus de conférences, initialement diffusées dans différentes variantes de l'espagnol, avec le but d'identifier les stratégies les plus couramment utilisées par les transcribers volontaires pour traiter les

expressions locales ou idiomatiques, ainsi que les expressions culturellement marquées pour atteindre le maximum de public possible et faciliter la traduction.

**Mots clés :**

Sous-titrage, traduction, TEDx, localisation, espagnol neutre

**1. Introduction**

TED talks are delivered by expert speakers on topics that range from education, entertainment and design to science and technology, global development and gender equality. Within the spirit of TED's "ideas worth spreading" mission, TEDx supports independent organizers to create TED-like events in their own community and deliver talks in languages other than English. In the case of languages with different variants like Spanish, volunteer transcribers face a further challenge as they not only have to pave the way for translations into other languages, but also find an appropriate global Spanish version that can be understood by the majority of Spanish speakers.

TED talks are translated into 113 languages by 25,001 translators totalling 101,877 translations through the Open Translation Project (OTP). Likewise, the TEDx library contains approximately 30,000 videos, 3,000 of them

originally in Spanish, which is an interesting phenomenon from the perspective of translation research. In sum, OTP provides a platform that helps volunteers from all over the world to transcribe, subtitle and translate the scripts of TED talks into more than 100 languages.

In spite of the large number of videos subtitled and translated into several languages, research studies on the relationship between TED talks and translation are relatively scarce and basically focus on the automation of subtitling and translation (Paul et al. 2010), the motivation of volunteer translators (Olohan 2014), and above all, the different approaches to community translation (Cronin 2010; Díaz-Cintas and Muñoz 2006; Fernandez 2011; Kelly et al. 2011; O'Brien 2011; O'Hagan and Ashworth 2002; O'Hagan 2009 and 2011; Perrino 2009).

The justification to engage volunteer transcribers is that transcribed TEDx talks can reach a wider audience because they are accessible for hearing impaired individuals, can be indexed on search engines and can achieve TED's mission of spreading ideas by making transcripts available for translation in TED's OTP. Therefore, TEDx talks transcribers play a crucial role in the overall translation workflow process as they are responsible for transcribing the contents and basis of what will be later on translated into different languages.

Therefore, the objective of this paper is to identify the most common strategies used by volunteer transcribers to handle local or idiomatic

expressions and culturally biased expressions from a corpus of TEDx talks originally delivered in different variants of Spanish. One of the objectives of transcribing TED talks is to reach the maximum audience possible and facilitate translation, thus, the underlying idea of this work is to identify transcription and subtitling strategies that take into consideration translation into other languages from the beginning. One of these strategies, the use of a universal and neutral version of the language, which is also suggested in TED's guidelines for volunteer transcribers, would allow the participation of more volunteers from a wider range of nationalities since the language that is common to a group, can be hard to understand or translate for another.

To this end, this work has been structured into four sections: after this introduction to contextualize the work, the first section focuses on the relationship between translation and TEDx talks and the role of translation in the dissemination of information aimed by this organization. Special emphasis is made on the phenomenon of community translation which in the literature appears under terms such as "community translation"(Kelly et al. 2011; O'Hagan 2011), "collaborative translation" (O'Brien 2011), "volunteer translation" (Olohan 2014), "crowdsourcing" (Cronin 2010; Fernandez 2011; O'Hagan 2009), "user-generated translation" (O'Hagan 2009; Perrino 2009) and "fan-based translation" or "fansubbing" (Díaz-Cintas and Muñoz 2006). This section concludes with the most common subtitling guidelines and the specific guidelines for TEDx talks in particular - which will justify some of

the decision-making strategies (Guardini 1998) used in the subtitles object of this study. This section also includes a concise review of the technical aspects of subtitling and the main currents of study of this discipline within audiovisual translation, because the subtitling of TEDx talks might throw light on a new subcategory of subtitling not taken into consideration in the literature up to now (Díaz-Cintas and Remael 2014).

Then, a brief revision of the literature on subtitling and translation of culture-specific items, or extralinguistic cultural references made. Section four presents the methodology used for this research work and the analysis and discussion of the results found in a corpus of fifty lectures compiled specifically for this study. In order to obtain evidence of the potential difficulties of using different varieties of Spanish, a corpus of 50 TEDx talks with a total of 82,502 words originally delivered in different varieties of Spanish, from Argentina, Cuba, Mexico, Guatemala and Spain was compiled to allow the identification and extraction of samples for the subsequent analysis and discussion of findings.

Finally, the conclusions and some ideas for further research are presented.

## **2. TED Talks and community translation**

From the computational linguistics perspective and more specifically, the spoken language translation, Paul et al. (2010) report research on the automation of subtitling and then translating TED talks from English into

French. Their challenge was the processing of text segmentation used in the reference transcripts (Paul et al. 2010, 4), and the wide variety of topics dealt with in TED talks, which hindered the availability of training data for the machine translation engine.

The other important line of research on the translation of TED talks comes from the motivation factors that lead volunteers to participate in community translation projects (Fernandez 2011; Olohan 2014; O'Brien 2011). For Fernandez (2011, 92), one of the main motivations, specifically for fansubbers, has its origins in the lack of translations from Japanese into Western languages of manga series in order to guarantee access to other followers around the world: "The personal desire to contribute to the spread of information or to allow other people to read and access materials in their own language." In addition, we should also consider altruistic aims and willingness to collaborate with NGOs.

O'Brien (2011, 18) distinguishes three motivators for collaborative translation: commercial, social and personal. Among the specific personal motivations pointed out by O'Brien (2011) are the desire to gain experience, learn new skills, and network. Among the most prominent implications of this mass collaborative crowdsourced translation are the effects on the translate-edit-proofread traditional professional translation process model (Kelly et al. 2011), and more specifically on the translation quality assessment, the blurring of boundaries between professional and non-professional translation



(O'Brien 2011), which consequently leads to a debate on the ethics of the profession (O'Hagan 2009).

Authors such as Kelly et al. (2011) and Fernandez (2011) point out that the benefits of reducing time in a translation project through the participation of a large number of translators can compromise quality control as there is a shift from correcting errors, TEP model, to an error prevention model in the case of collaborative translation. Kelly et al. (2011) also point out that this translation environment is still not very popular with language service providers and professional translation.

Fernandez (2011), on the other hand, analyses the phenomenon of collaborative translation from the point of view of amateur translation, non-professional translation and the threat for translation schools and future professional translators, and encourages more scientific research in this amateur field where there are no clear quality criteria established (Fernandez 2011, 87).

The literature on community translation focuses mainly on its development thanks to the technological advances and the motivators that lead to carry out user-generated translation provided in a voluntary manner without monetary reward (Kelly et al. 2011; O'Hagan 2009; O'Brien 2011; Perrino 2009).

This phenomenon, which arises from the active participation of web 2.0 users, entails that a translation project is carried out by different translators who

participate on disinterested and voluntary basis. As pointed out by authors such as O'Hagan (2009, 94), this phenomenon has been taking place since the 1980s, and it is after web 2.0 platforms when they have become a global phenomenon. Since then, it has been designated with different denominations like "community translation", "collaborative translation", "volunteer translation", "fan-based translation", "fan subbing", "user-generated translation", or "crowdsourcing". This variety of denominations is due to the different objectives, motivations or even thematic or technological resources used by participants in translation projects. Nevertheless, this terminology contains slight nuances that make these terms not completely interchangeable: for example, collaborative translation is carried out simultaneously on a translation project, such as Wikipedia or the translation of Facebook (O'Brien 2011,19; Kelly et al. 2011); while fansubbing usually refers to the translation carried out by fans, and almost always circumscribed to the scope of manga, and comics. As Fernandez (2011, 92) puts it:

“Several differences between these activities are evident since they belong to different fields and scopes. Nevertheless, there is one common feature underlying all these processes which means the basis for a translation can be considered “collaborative”: the fact that they are all adaptations made by fans for fans.”

The common thread between all these terms is that translation is done by a group of individuals working together on a volunteer basis, and it is expected that each small contribution paves the way to make the audiovisual product in question more accessible and help subsequent translations.

On the other hand, Perrino (2009) and O'Hagan (2009, 97) prefer the term user-generated translation (UGT) “to mean a wide range of translation, carried out based on free user participation in digital media spaces where translation is undertaken by unspecified self-selected individuals.”

Finally, fansubbing (O'Brien 2011, 17) is primarily concerned with interlingual subtitles where the script is adapted from the source into the target language, as opposed to intralingual subtitles (or captioning) in which the on-screen text is in the same language as the dialogues.

O'Brien (2011, 17) attributes the rise of collaborative translation to technological developments, especially computer-aided translation tools, the creation of virtual communities on the Internet, and the Open Source movement, since the increase of user-generated content involves translation demand. This statement is also supported by authors such as Kelly et al. (2011, 84): In the collaborative translation world, translations are carried out more efficiently because of technologies that enable these communities to work together seamlessly.

Of all the literature consulted, Perrino (2009, 63) focuses exclusively on the possibilities of online translation tools that have favoured the development of new translation activities such as collaborative translation in particular "UGT tools providing the means to translate digital texts: Professional translation networks, wikis and user-generated dictionaries."

Among the tools highlighted by O'Brien (2011) are what she calls Collaborative Translation Platforms, which combine terminological management functions, translation memories, machine translation and project management. For Kelly et al. (2011, 85):

"In a collaborative model, translation communities rely on commonly used technology like instant messaging (including via social media platforms), SMS, discussion groups, RSS feeds, social tagging, content categorization, and voting."

## **2.1. Subtitling**

The different modes of audiovisual translation are usually classified by the type of transfer (Chaume 2013, 111): between two languages and cultures (interlingual) or within the same language and culture (intralingual), and by the way in which the translation is displayed, that is, subtitling or dubbing. Within the former category, in which the written translation is inserted into

the screen (subtitling or captioning), Chaume (2013, 112) distinguishes between subtitling, surtitling, respeaking, subtitling for the deaf and hard of hearing and fansubbing.

For Chaume (2013, 112) subtitling “consists of incorporating a written text (subtitles) in the target language on the screen where an original film is shown, such that the subtitles coincide approximately with the screen actors' dialogues” highlighting the space constraints for the subtitles and the academic discussion about the speed with which the human eye can read, which depending on the language ranges from 180 to 300 words per minute. According to Díaz-Cintas and Remael (2014, 8):

“Subtitling may be defined as a translation practice that consists of presenting a written text, generally on the lower part of the screen, that endeavours to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image (letters, inserts, graffiti, inscriptions, placards, and the like), and the information that is contained on the soundtrack (songs, voices off)”.

Therefore, the text of the subtitles must appear synchronously with image and dialogue, the message must be adequately conveyed as it appears in the

original dialogue, and stay on screen long enough for the audience to read them (Díaz-Cintas and Remael, 2014)

According to Díaz-Cintas and Remael (2014) there are five types of subtitling that are classified based on linguistic criteria, preparation time, technical aspects, projection method, or distribution format. Furthermore, these five types contain several subcategories, such as open and closed subtitles, chemical and mechanical subtitling in the methods of projection category, or in the linguistic section, bilingual, intralingual or interlingual subtitles. Within this type of linguistic subtitling it should be noted that after an exhaustive analysis of the mechanics, procedures, and translation philosophy of TED talks, we conclude that a new fourth type of subcategory should be added to the classification by Díaz-Cintas and Remael (2014): subtitling to facilitate translation, which precisely has its origins in community translation.

From the literature consulted, it seems that until now most classifications focus on the audiovisual product and the way to make it accessible, either because it is a translation, or a subtitling or audio description, for a hearing impaired audience, and not on subtitling as part of a larger process, such as the subsequent translation into other languages.

From the purely technical side, TED uses the Amara system to allow volunteers subtitle and synchronize the scripts of the talks in an extremely user-friendly interface, with the minimum amount of functions and concentrating on effectively transcribing and synchronizing the script of the

talk. To help achieve this, OTP and Amara provide several video tutorials and guidelines to learn about subtitling best practices, collaboration tips and style guidelines. These guidelines include technical specifications and style and language recommendations, among them, the preference to the global over the regional use of language and to the choice of words and phrases that can be universally understood among all language variants and how to handle idiomatic expressions instead of a word-for-word representation.

From the platform's point of view, the complete subtitling process consists of five phases: transcription (including synchronization), revision of subtitles, approval by another more experienced volunteer user, publication of subtitles, and availability for its translation into other languages.

### **3. Culture-specific items in translation**

Since the main purpose of our study is to identify the most common techniques to translate culturally biased expressions, it is convenient at this point to offer a brief summary of the literature on culture-specific items.

There is a great diversity of definitions to designate cultural-specific items from the different translation schools and there are several taxonomies to classify them. The terminological variety found in the literature reflects the lack of consensus in the research community to identify and define cultural elements which entails a problem of ambiguity and the need for the explanation of its meaning.

Vinay and Darbelnet (1965) and Vázquez Ayora (1977) represent the comparative theoretical current by which cultural elements are defined as metalinguistic divergences when culturally translating and adapting a text.

Newmark (1981, 2) refers to “cultural terms”, and defines them as "... token-words which first add local colour to any description of their countries of origin, and may have to be explained, depending on the readership and the type of text". Newmark also coined the term “cultural focus” to express a phenomenon in which a community focuses its attention on a particular topic. Some examples are the terms on cricket in English, terminology of wines and cheese in French and in German, sausages. The cultural distance is the main cause of translation problems in those cases.

In line with this, Reiss (1996) coined the term extralinguistic determinants to refer to those words that determine the culture of a community itself or its specific history, that is, realia. And in *Culture Bumps: An Empirical Approach to the Translation of Allusions* (1997), Leppihalme uses the denomination “culture bumps” to indicate any situation where the reader of a target text has a problem understanding a source-cultural allusion.

Vlakhov and Florin (1970) use the term realia, with four dimensions: geographical and ethnographic, folk and mythological, everyday objects, social and historical. Nida (1964, 216) refers to the problems when translating elements associated with a cultural context: “Terms associated with social culture pose numerous problems, not only because the basic systems are often



so different, but also because the extensions of meaning appropriate to one system rarely work in another.” Nida also establishes several areas of cultural differences: ecology, social culture, religious culture and linguistic culture, with additional categories: phonology, morphology, syntax and lexicon.

Newmark (1988) adapts what Nida indicated by dividing what he calls cultural words: ecology, including flora, fauna, geography, topography, etc.; material culture, with subcategories such as food, clothing, homes and populations, and transport; social culture, which includes work and leisure; organizations, customs, activities, procedures, concepts, which are divided into political and administrative, religious, and gestures and habits.

Finally, Katan (1999) adds that cultural elements cannot be found in one or several levels in isolation, but they function as a process and operate at all levels. For Katan, there are six logical levels of culture that operate at the same time: environment, which includes the physical and ideological context, climate, space, housing and constructions, clothing, smells, food, and time frames, behaviour, which includes the actions and reactions perceived in any culture, skills, rituals, strategies and skills to communicate, and other hidden characteristics, which include the values of a society and its hierarchy, beliefs and identity.

After considering the different cultural areas with their different classifications and denominations, the criteria and techniques to solve the translation of cultural references should be pointed out. One of the most

recent trends is internationalization. Jiménez Crespo (2013) suggests that it is necessary to produce and design global products that can easily be adapted to different regions. This involves neutralising any elements rooted in language or culture.”

Overall, internationalization procedures are designed to ensure that the document or application is not culturally limited, but it is independent of the language of development and is easy to translate without major technical modifications (Jiménez Crespo 2013). Pérez Fernández (2010) conceives internationalization from the translation perspective as a process of designing a product so that it can be more easily adapted, linguistically and culturally, to another target market and without engineering changes.

From the point of view of subtitling, Pedersen (2011, 43) uses the term "extralinguistic cultural references" which he defines as:

“any cultural linguistic expression, which refers to an extralinguistic entity or process. The referent of the said expression may prototypically be assumed to be identifiable to a relevant audience as this referent is within the encyclopaedic knowledge of this audience.”

Within this category of extralinguistic cultural references, Pedersen (2011) includes references to places, people, institutions, customs, food etc., which one may not know even mastering the language. Pedersen (2011) also points out that some elements are more difficult to subtitle than others and that a conscious use of subtitling strategies is required.

In our study, the translation of TED Talks poses a serious challenge to volunteer transcribers, as they need to activate their cross-cultural competence to be able to translate any cultural items belonging to the different language-cultures. Some of these cultural items are easily recognisable, such as food, clothing, typical houses or famous people, but other concealed and intangible elements may also be present. These can be cultural biased expressions reflecting society values and behaviours, as well as other linguistic expressions involving more controversial issues like ideology or beliefs.

#### **4. Analysis and discussion**

With all this context in mind, a representative corpus was compiled to find and identify examples of challenging situations for translation that could be addressed from the subtitling stage. The first step consisted in identifying 50 representative TEDx talks delivered in 5 variants of Spanish: Argentina, Cuba, Mexico, Guatemala and Spain. All in all, a corpus with a total of 82,502

words and approximately 350 minutes of TEDx talks video recordings was processed.

One of the criteria for the selection of the talks was that they had to be translated into English to facilitate the identification of the subtitling-for-translation strategies. Then, the Spanish source text was aligned with its translation into English to carry out a manual review and identify the type of cultural elements that could cause difficulties to translators or to a potential audience if cultural items were not handled properly.

In sum, for simplification purposes, culture specific items found in the corpus were classified into three general categories based on their frequency of occurrence: Greetings and ways of addressing the audience, Culture specific references (including geographical), and Idiomatic expressions.

This work proposes a qualitative study to pave the way for further research, firstly because the main objective of this work is to contribute to research on the approach to subtitling cultural elements performed by volunteers; and secondly, because the identification was carried out manually by expert linguists with the idea of extracting the most significant examples and not a detailed compilation of all of the occurrences.

According to recent statistics (Fernández-Vitores 2016), more than 472 million people speak Spanish as their mother tongue, and in total, Spanish is spoken as an official language in 21 countries. Although the main differences

between the varieties of Spanish occur between Latin American Spanish and Castilian Spanish, it is also true that there is no standard form of Latin American Spanish. Overall, the main differences between all varieties of Spanish are pronunciation, vocabulary and grammar. For this work, focused on the written text of subtitling, emphasis is placed on lexical and grammatical differences, and on the cultural specific references.

#### **4.1. Greetings and ways of addressing the audience**

The mode of addressing and greeting is part of cultural rituals as seen in section 3. They are socially essential and may include formulaic expressions that form part of more elaborate language rituals that vary in the various cultures.

At the grammatical level, the main differences between Latin American and Peninsular Spanish are the use of “tú” and “usted” for the second person singular (you), and some slight differences in the use of tenses.

Among the examples found in the corpus, there is a clear distinction between the use of “usted” in Latin American Spanish as compared to the use of “tú” in peninsular Spanish, as in the examples found in Mexico (1) “Sin duda, probablemente, alguno de **Uds.**.../ *Surely, one of you...*”; (2) “Hoy voy a compartir con **Uds.** / *Today I will share with you*”; (3) “¿Por qué **les** digo esto? / *Why do I say this?*”.

However, although the use of “tú” is more common in Spain, there are some examples in the corpus from Mexico and Guatemala that use the “tú” form, as in (4) “Y ¿**sabes** / *And you know what?*”; (5) “¿Cuál es tu propósito? / *What's your purpose?*”; (6) “**Tú** eres mi inspiración, / *You are my inspiration.*”

Nevertheless, as expected, the plural form of you (vosotros) was only found in the TEDx talks from Spain: (7) “¿Os **imagináis** encontraros con un monstruo así en un callejón oscuro a media noche? / *Can you imagine bumping into a monster like this one in a dark alley at midnight?* ; (8) “¿Y **sabéis** que? / *And you know what?*”; (9) “Como todos **sabéis**... / *As you all know,* “

In the above mentioned examples the neutralisation of the different ways of addressing the audience in Peninsular and Latin American Spanish in the subtitled text is due to the intrinsic target language structure, which restricts the various addressing formulae in Spanish to use of “you” in English.

#### **4.2. Culture-specific references**

The topics covered in TED Talks make use of a technical-specialized language, thus not many culture-specific referents have been found, as they belong to a more colloquial register. Nevertheless, as mentioned by Pedersen (2011), most referents common in subtitling fall into the category of names of people, places and institutions.

As previously reviewed in section 3, the use of references to famous people contributes to shape the cultural identity. In some cases, examples of explicitation in names, origins and profession of the person mentioned are found—which does not appear in the original since it is assumed that the audience of the TEDx talk is familiar with that name: (10) “Por supuesto, bastaría con pensar en todo el registro que hace **Ernesto Oroza** / *It'll be enough to think of a Cuban designer, Ernesto Oroza's study.*” In other cases, given the relevance of the characters in the Spanish-speaking world, they remain the same as in the original: (11) “Tiene su parte romántica, de lo que murieron **Miguel Hernández**,. / *It has its romantic side, the cause of death of Miguel Hernández,...*” And finally, there are also examples in which the name of the person is known only in that country and is reproduced directly without further explanations, as in the following case taken from TEDx Guatemala and a reference to a very famous local sportsman, unknown to the rest of the Hispanic world: (12) “Si salías a correr con walkman terminabas haciendo puro **Julio Martínez**. / If you went jogging with a Walkman, you would end up like **Julio Martinez.**”

With regard to the name of institutions, the strategies found coincide with the strategies use to handle proper names seen above. Sometimes, perhaps because there is more context, the acronym remains the same in English as can be seen in this example from TEDx Habana: (13) “estuve 12 años becado por la Escuela Elemental de Arte, la escuela de nivel medio de arte de

Camagüey y el **ISA** ... / *I was granted scholarships many times, from the Elementary School of the Arts, the Secondary School of the Arts in Camagüey and ISA...*”; (14) “De acuerdo a cifras del **INEGI** del 2013 / *According to the INEGI figures for 2013...*” although in other cases the spelled-out form in English is preferred: (15) “Luego me fui a la ENA, / *Then I went to the National School of Arts.*”, or the explanation of the abbreviation: (16) “Durante toda la **EGB** yo había ido... / *Throughout all of primary school,*”

### 4.3. Idiomatic expressions and lexical differences

At the lexical level, Latin American Spanish is influenced by its contact with indigenous languages, or in the case of Mexico, also by the English influence due to the proximity with the United States.

The following examples from TEDx Tijuana are clear examples of the uniqueness of Mexican Spanish: (17) “El trabajo es el ¡**Órale!**!” / *Work is the just do it!*”; (18) “**No hay de otra.** / *There's no way around it.*”; (19) “he sido **mercadólogo**... / *I've been a marketing specialist*”; (20) “Pues he tenido empresas dedicadas al **import & export** / *I've had import-export companies*”; (21) “lo podemos usar como balón de **básquetbol** / *or even use them as a basketball ball.*”; (22) “Y quiero que Uds. le echen una mirada a su **clóset**. / *...and would like you to take a look at your closet.*”



On the other hand, the lexical and cultural richness of those 21 countries where Spanish is spoken gives rise to different denominations for the same concept, which is sometimes a source of lexical ambiguities. For example, a common verb like "parar" means "stand up" in Mexico, "stop" in Spain, "wake up" in Venezuela, etc.

The word "milonga" may lead to ambiguity to a non-Argentinian speaker as it has up to 8 meanings, among them, "musical composition", "dance", "song", and "mess" just to mention a few: (23) "Comencé a bailar cuando tenía 17 años pero mi permanencia en la milonga duró poco/ *I started to dance when I was 17 years old but my milonga experience was short-lived.*" However, in most of the cases, the context clarifies immediately this ambiguity.

In other cases, we have identified the lexical richness of terms such as (24) "Lavarropas, ascensores / *Washing machines, elevators*"; (25) "Pero además, esta computadora, que tengo acá en mi mano, / *In addition, this computer that I have here in my hand,*" as well as other words that appear frequently in the common language in informal contexts (26) "para que Uds. entiendan qué **lindo** que es... / *so that you understand its beauty*"; (27) "Sabemos que es **chiquita**. / *We know it's tiny.*"; (28) "Bueno, sinceramente yo debo comentarles que no hubiese querido estar **acá**. / *I should confess that I wish I didn't have to be here.*"; (29) "Pero no solo **comida chatarra**, también hay

trigo transgénico. / But there's not only **junk food**, but also transgenic food, soya and wheat.”

Among the characteristics of subtitling is the visual support that accompanies the text. In these cases, the transcriber can even choose to leave the cultural element itself in the source language since the image contributes to provide information about the meaning. This is the case of "almendrón", an old American car, very common in Cuba: (30) “Y también los cubanos hemos creado de manera inconsciente, pudiéramos decir "informal", una red de movilidad que conocemos como boteros, almendrones, o que el mundo acuña como taxi en común. / *We Cubans have inadvertently established an informal mobility network made up of what we call 'boteros' [taxi drivers] and 'almendrones' [old American cars]*”. In the same talk, a few minutes later, “almendrón” is left as such: (31) “Cada 5 personas que vamos en un almendrón estamos reduciendo un congestionamiento vehicular. / *that each almendrón with capacity of up to five passengers helps reduce traffic jams.*”

Finally, the main lexical differences found in the corpus consist in the use of prefixes and suffixes that are completely neutralized in the English translation, and which are part of the idiosyncrasy of Spanish, both in Latin America and in Spain, and have not been universalized in the original transcript: (32) “fue **súper** difícil poder llegar a un lugar... / it was **very** difficult to get to a place”; (33) “No es lo ideal para tomarse un **cafecito** / ... which is not ideal for drinking **coffee**.” Other uses of suffixes, prefixes and

examples of lexical peculiarities include: (34) *ahorita* (now), (35) *chiquitos* (little), (36) *acá* (here).

Further examples of use of informal colloquial register in TED talks include some slang words and diminutives as in: (37) “es un trabajo que **mola** mucho”; (38) “a tomarnos unas **cañitas**”; (39) Mirad el “**dibujito**”; (40) “un montón de cosas así”; (41) “¿Por qué os estoy pegando la **chapa** sobre los líderes?”; (42) “**puñetero**”; (43) “**proyectito**”.

## 5. Conclusions

The role of transcribers in the overall dissemination process of TED talks is unquestionable. First of all, because the processing of speech to text facilitates to a large extent the subsequent translation into other languages- in addition to the benefits of indexing and accessibility for the hearing impaired stated by TED.

However, in addition to the technical constraints of subtitling, in the case of Spanish, the first obstacle in the case lies in the significant differences between Latin American and peninsular Spanish, mainly with regards to pronunciation and vocabulary.

Thus, a corpus of TED talks originally delivered in different variants of Spanish was compiled to identify the main potential difficulties for the

translation into other languages and consequently, the dissemination of TED knowledge.

The most common differences that illustrate potential subtitling and translation difficulties focus on greetings and ways to address the audience, culture-specific references and idiomatic expressions and lexical differences.

Overall, the strategies used by volunteer transcribers follow TED subtitling guidelines and make use of a universal version of Spanish, however, it should be noted that language neutralization is seen more clearly in the translation into English, as most of the times visual aids contribute to clarify the meaning. Other strategies include the explanation—transcribers add information to those lexical items which are only recognisable by those individuals sharing the same culture and Spanish variant.

The translation to other languages is carried out by volunteer translators who are not necessarily familiar with the specific Spanish variant in which the speech is delivered, therefore we advocate a use of a neutral or global Spanish that would give access to a greater number of volunteer translators. However, this neutralization presents advantages and disadvantages: among the advantages is the speed and the greater availability of translators and the elimination of potential ambiguity in the subtitled text; and among the disadvantages of neutralization or domestication would be the loss of that original regional flavour—which would nonetheless remain in the audiovisual version of the talk.

Finally, we present some examples identified in the 50 presentations of the corpus that can contribute to initiate the discussion on the appropriateness of the neutralization of subtitles, and then provide some conclusions and ideas for further research on this topic.

Due to the emergence of collaborative translation, within the linguistic subtitling classification, a fourth type of subcategory not included up to now should be added: subtitling to facilitate translation into other languages.

Further research would include the study of the adequacy of using a pivot language to translate from lesser-known languages, i.e., whether subsequent translations are carried out from Spanish as source language, or from the English translation and whether the same issues with the subtitling of cultural referents appear in other languages with different geographical variants such as French or English.

## 6. References

Chaume, Frederic. 2013. "The Turn of Audiovisual Translation: New Audiences and New Technologies." *Translation Spaces* 2 (1): 105-123.

Cronin, Michael. 2010. "The Translation Crowd." *Tradumàtica: tecnologies de la traducció* 8: 1-7.

Díaz-Cintas, Jorge, and Pablo Muñoz-Sánchez. 2006. "Fansubs: Audiovisual Translation in an amateur environment." *The Journal of Specialised Translation* 6: 37-52.

Díaz Cintas, Jorge, and Aline Remael. 2014. *Audiovisual Translation: subtitling*. New York: Routledge.

Fernandez Costales, Alberto. 2011. "Crowdsourcing and collaborative translation: Mass phenomena of silent threat to translation studies." *Hermeneus* 15: 85-110.

Fernández Vítóres, David. 2016. *El español: Una lengua viva. Informe 2016*. Madrid: Instituto Cervantes. Accessed July 10, 2016. <http://www.cervantes.es/imagenes/File/prensa/EspanolLenguaViva16.pdf>

Guardini, Paola. 1998. "Decision-making in subtitling." *Perspectives* 6 (1): 91-112.

Jiménez Crespo, Miguel Ángel. 2013. *Translation and Web Localization*. New York: Routledge.

Katan, David. 1999. *Translating Cultures. An introduction for Translators, Interpreters and Mediators*. Manchester: St. Jerome Publishing.

Kelly, Nataly, Rebecca Ray, and Donald A. DePalma. 2011. "From crawling to sprinting: Community translation goes mainstream." *Linguistica Antverpiensia, New Series—Themes in Translation Studies* 10: 75-94.

Leppihalme, Ritva. 1997. *Culture bumps. An empiric approach to the translation of allusions*. Clevedon: Multilingual Matters.

Newmark, Peter. 1981. *Approaches to Translation*. Oxford: Pergamon Press.

Newmark, Peter. 1988. *A Textbook on Translation*. Hertfordshire: Prentice-Hall.

Nida, Eugene. 1964. *Towards a Science of Translating with special reference to principles and procedures involved in Bible translating*. Leiden: E.J. Brill.

O'Brien, Sharon. 2011. "Collaborative translation" In *Handbook of Translation Studies* 2, ed by Yves Gambier, and Luc van Doorslaer, 17-20.

O'Hagan, Minako and David Ashworth. 2002. *Translation Mediated Communication in a Digital World: Facing the challenges of globalization and localization*. Clevedon: Multilingual Matters.

O'Hagan, Minako. 2009. "Evolution of User-generated Translation: Fansubs, Translation Hacking and Crowdsourcing." *Journal of Internationalisation and Localisation* 1 (1): 94–121.

O'Hagan, Minako. 2011. "Community Translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 10: 11-23.

Olohan, Maeve. 2014. "Why do you translate? Motivation to volunteer and TED translation." *Translation Studies* 7 (1): 17-33

Open Translation Project (OTP). Accessed May 20, 2016. <https://www.ted.com/about/programs-initiatives/ted-translators>

Paul, Michael, Marcello Federico, and Sebastian Stüker. 2010. "Overview of the IWSLT 2010 evaluation campaign." In *Proceedings of the 7th International Workshop on Spoken Language Translation* 10: 3-27.

Pedersen, Jan. 2007. "Cultural interchangeability: The effects of substituting cultural references in subtitling." *Perspectives: Studies in Translatology* 15 (1): 30-48.

Pedersen, Jan. 2011. *Subtitling Norms for Television: An exploration focussing on extralinguistic cultural references*. Amsterdam: John Benjamins.

Pérez Fernández, Lucila María. 2010. "La localización de videojuegos: una propuesta didáctica" In *La traducción en contextos especializados. Propuestas didácticas*, ed by Esperanza Alarcón Navío, Granada: Atrio, 277-286.



Perrino, Saverio. 2009. "User-generated Translation: The Future of Translation in a Web 2.0 Environment." *The Journal of Specialised Translation* 12: 55-78.

Reiss, Katharina and Hans Vermeer. 1996. *Fundamentos para una teoría fundacional de la traducción*. Madrid: Akal.

Vázquez Ayora, Gerardo. 1977. *Introducción a la traductología. Curso básico de traducción*. Washington: Georgetown University Press.

Vinay, Jean-Paul and Jean Darbelnet. 1965. *Stylistique comparée du français et de l'anglais. Méthode de traduction*. Paris : Didier.

Vlakhov, Sergei and Sider Florin. 1970. "Neperevodimoe v perevode: realii," [The untranslatable in translation, realia] In *Mastersvo perevoda*. Moscú: Sovetskii pisatel, 432-456.

Miguel Ángel Candel-Mora is associate professor in the Applied Linguistics Department of Universitat Politècnica de València, Spain, where he teaches language and technology courses. MA in Translation from Kent State University and PhD from Universitat de València. His academic interests focus on translation-oriented bilingual terminology management, computer-assisted translation, information retrieval, corpus-based translation and text mining.

Diana María González-Pastor is lecturer in the Department of Philology, Translation and Communication of Universitat de València, Spain. She holds a PhD in translation and culture. Her research focuses on the language of tourism, its multimodal dimension and its translation. Her pedagogical interests include intercultural communication and educational innovation.