Univerzitet Crne Gore
Filološki fakultet
Nikšić

Na osnovu člana 37. Pravila doktorskih studija Univerziteta Crne Gore dajem sljedeću

SAGLASNOST

Rad pod nazivom *Audiovizuelno prevođenje i elementi kulture: korpusna analiza transfera sa studijom recepcije u Crnoj Gori*, autora mr Petra Božovića, saradnika u nastavi na Filološkom fakultetu Univerziteta Crne Gore, zadovoljava kriterijume doktorske disertacije propisane Statutom Univerziteta Crne Gore i Pravilima doktorskih studija.

U Ljubljani, 25. 2. 2019   MENTOR

            Prof. dr Vojko Gorjanc

How should culture be rendered in subtitling and dubbing?

A reception study on preferences and attitudes of end-users

Petar Božović
University of Montenegro

Empirical reception research in audiovisual translation (AVT) has long been neglected as most previous studies focused primarily on features of AVT as a product and the producers of AVT, but not on end-users. The importance of these studies is emphasized by the fact that AV content is designed for target audiences. Knowing their preferences, expectations and needs can inform the industry and, thus, increase positive reception, placement and usability of the product. The present contribution aims at answering the question of whether there is a difference in preferences when it comes to translation strategies used for rendering the elements of culture (EC) in two different AVT modalities (subtitling and dubbing) and what could be the underlying reasons for it. For this purpose, we conducted an experiment with 136 participants among the undergraduate students at two universities in Montenegro. Results indicate that end-users have different general preferences and expectations for these modalities. We hypothesize that the observed differences are related to the "vulnerability" of subtitling. Furthermore, fluctuations from the general preference in subtitling have been noted in monocultural ECs in connection with encodedness of the EC in source text humor. This shows that industry's translation guidelines should be modality-specific and that an over-simplistic approach to the treatment of such a complex issue as rendering culture within certain modalities should be avoided.

Keywords: reception studies, subtitling, dubbing, elements of culture

1.   Introduction

Some of the first calls for more empirical reception studies (RS) in audiovisual translation (AVT) were given in the mid-1990s and intensified by a number of researchers in the filed from the beginning of the 2000s onward (Kovačić 1995; Leppihalme 1996; Fuentes 2003; Gambier 2006, 2008; etc.). These studies are seen as a way of closing the circle of AVT research which was, previously, mainly focused on the features of AVT as a product and the process of production, but has often neglected the actual end-users and their needs, responses, expectations and preferences (Di Giovanni 2016).

The importance of such studies is emphasized by the fact that the AV content, whether on the internet, TV, or any other medium, is designed for target audiences. Knowing their needs and expectations can influence greatly the reception, placement and usability of the product. O'Sullivan (2016:270) notes that it is also "fundamental in order to assess the impact of policy" used by media and other stakeholders regarding the relevance of approaches and strategies used in AVT. These studies, hence, give empirical evidence that informs translators as to the choices they make during the translation process, without which "it is impossible to judge the translator's extra-linguistic context adequately" (McAuley 2015:221). In most cases, without such insights, translators would be guided by their own intuition, speculations and a self-constructed image of an intended, or ideal, end-user. Clearly, research in this field contributes to the quality and relevance of translations that both depend on "the awareness of the audience's cognitive and evaluative response to the product itself or to specific aspects of it" (Perego 2016:156).

As for the translatological and linguistic context of Montenegro where we carried out our study, subtitling has been a predominant modality while dubbing is mostly used for cartoons. This is the case with most countries of the region, as well. The AV market is highly specific as it has been characterized by the presence of three different language standards both on national and privately-owned stations: Serbian, Croatian and Montenegrin. These are the successors of Serbo-Croatian, a polycentric language that, under the pressure of political situation and disintegration of the Socialist Federative Republic of Yugoslavia, split into different standards (Bosnian, Croatian, Montenegrin, and Serbian) that underwent recodification (Požgaj Hadži 2013; Hlavac 2015). While prior to Montenegro's independence in 2006 much of the AV material with the translations was imported from bigger regional centers in Belgrade and in Zagreb, in recent years, however, the RTCG (Radio and Television of Montenegro – a public service broadcaster) began producing the subtitles and dubbing cartoons in Montenegrin.

2.   Recent AVT reception research

Recent years have witnessed flourishing of the RS in AVT. Based on the primary focus of research, these empirical studies can be divided into two sub-categories: *translation-induced effect studies* and *user-induced effect studies*. The former focus on the features of translated texts and their effects on reception (e.g. translation strategies that were used and the way they influence the reception), the latter on the effects that user-profile has on the reception (e.g. the influence of the level of listening comprehension proficiency on reception).

The examples illustrating the first group include Orrego-Carmona (2016) who conducted a research on audience reception of subtitled TV series using professional and non-professional subtitles, Denton and Ciampi (2012) who did a study on strategies for rendering culture-specific references in dubbing and the extent to which they allow the transmission of culture-bound issues, Schauffler (2012) who studied the reception of two different strategies for translating English wordplay into German, Caffrey (2009) who conducted a research on the effects of abusive subtitling procedure on reception, Antonini (2005) who did a study on the perception of translated humor, etc. The examples illustrating the second group include Orrego-Carmona (2014) and his study on the influence of listening comprehension skills and reception of subtitles, Touminen (2012) who focused her research on what real viewing situations are like and how viewers watch a subtitled film, etc.

It goes without saying, that the above mentioned effects grouped around the two poles (product vs. user) can be unidirectional (when translations influence the end-users but not the other way round) or bidirectional (the case of mutual interaction between translation practice and end-users when the former influence reception and the latter, in turn, shape the translation practice, which, we could argue, if not in all, at least in many cases might be a desired scenario, especially within the field of AVT). This balance between TT and recipients, though possibly controversial, must remain central to TS RS.

Even though recent years have seen an increase in RS, it is such a vast field that scholars still call for a need "to do much more audience research in order to understand viewer responses to available modalities of translation" (O'Sullivan 2016:270). This papers aims at giving a contribution on shedding light of issues at stake in this field.

3.   Study design

The present research is aimed at answering the following questions: (a) is there a difference in the general preferred translation strategy among the end-users for rendering the elements of culture (EC) between subtitling and dubbing, (b) in case of a difference, what could be some of the underlying reasons.

The theoretical framework relied heavily on Jan Pedersen's descriptive study on rendering the ECs in AVT (Pedersen 2011). Different translation strategies were grouped around the two poles, foreignization and domestication, by following the taxonomy proposed by Pedersen. Hence, retention, specification and direct translation are grouped around the foreignization pole, whereas the generalization, substitution and omission were grouped around the domesticating pole. Particular attention was given to the reception of the cultural substitution as one of the most domestication-oriented translation strategies.

The ECs were classified into the intralinguistic and extralinguistic. The former including diversified language such as slang, as well as idiomatic expressions, proverbs, slogans, while the latter include references to cultural realia that is not part of the language system. Although this taxonomy was originally applied by Pedersen to the extralinguistic elements only, in this study, we applied it successfully to the intralinguistic ones. Furthermore, when describing the nature of ECs according to the level of transculturality, we used Pedersen's classification of these into: transcultural, monocultural and microcultural. Transcultural are those which are assumed to be known by both the source and target culture as part of the encyclopedic knowledge, monocultural are those which are assumed to be unknown by most of the target culture members and as such pose a challenge for translators, and, finally, microcultural ECs which are source-culture bound and unknown to quite a number of members of source culture, also posing a specific challenge for translators.

The main hypothesis is based on the common understanding of the different degrees of "vulnerability" of the modalities in AVT and the level of transculturality of the source language EC. Diaz Cintas and Remael (2007) consider subtitling as one of the most "vulnerable" of all the AVT modalities as the audience is simultaneously exposed both to the source text through visual and auditory channel, and the target text in the written condensed form of subtitles, which is not the case with dubbing, of course. Having this in mind, we assume that the general audience preferences will differ and thus the majority of the participants will prefer the foreignization approach in subtitling and the domestication approach in dubbing. When it comes to the nature of ECs in terms of the level of transculturality, it is assumed that specifically in cases of mono or microcultural ECs, the general preferred approach will be a domesticated one.

3.1 Participants

One hundred and thirty six participants (aged 18–30) took part in this study. They included undergraduate students from both the state and private universities in Montenegro: the University of Montenegro and the University of Donja Gorica.

In order to reduce the potential subconscious bias due to the theoretical knowledge (Schutze and Sprouse 2013), participants were chosen randomly from linguistic (106 participants) and non-linguistic (30 participants) fields of study including international finance and business, geography, psychology while the language-oriented studies include the first and the second year students (thus still beginners in the field) from the English Language and Literature Department, University of Montenegro. Of all the participants, 108 were women (79%), and 28 men (21%). The level of proficiency in English varied from pre-intermediate to upper-intermediate and advanced.

The study was organized as part of the participants' out-of-class and, in some cases, in-class activities for either their English language or Translation course, or as a part of practice in experimental research for the psychology students. The study focusing on dubbing was carried out in May 2016 with 34 participants, while the subtitling study was carried out in March 2017 with 102 participants. Students were given course credits for taking part in this research.

3.2 Stimuli

The subtitling experiment included five sets of two excerpts from the ninth season of the American sitcom *Friends* (Warner Bros. Television, 2002, D. Crane & M. Kauffman). Pedersen's study (2011) of over 100 Anglophone series, movies, etc. suggests that the episode 15, *The One with the Mugging*, is characterized by a high presence of extralinguistic ECs and that is why all of the excerpts were taken from this episode. The average length of each excerpt was 40 seconds. Each set consisted of the same excerpt with two different versions of the subtitles in Montenegrin, one using some of the domesticating, and the other one of the foreignization strategies. The order of appearance will be described below in the procedure section. For the purpose of this study, we used the *iKA*,[1] an open source application that enables services for online surveys.

The dubbing experiment included five excerpts from the animated adventure movie *Incredibles* (*Pixar Animation Studios*, 2004, B. Bird) dubbed into Croatian. Each excerpt was approximately a minute and a half long. They contained diversified language, both standard and non-standard Croatian, including slang and

---

1. www.1ka.si

different regional dialects. Different translation strategies were used in these excerpts, but the dominant one was cultural substitution.

Finding the adequate stimuli for dubbing proved to be a challenging task because in the whole region there is usually just one dubbed version of an animated movie and it's either in Croatian or Serbian. So there aren't different versions with possible different translation strategies at work. It goes without saying that making our own dubbed version would have been financially impossible as we had no grant for this project available. This meant that we had to use a questionnaire with closed and open format questions and an interview trying to find the underlying reasons for their preferences.

3.3   Procedure

For the subtitling experiment, students were asked to fill-in an online questionnaire which consisted of three parts: the basic sociodemographic part consisting of 9 questions including: sex, age, the university they are enrolled at, study program, level of English proficiency (self-assessment), their first language, TV/Internet watching habits, etc. The second part consisted of the actual experiment stimuli: five sets of subtitled excerpts, the results of which are the focus of this paper. Participants read the instructions on screen that informed them that they were about to watch short five sets of two video clips, to imagine that they were relaxing at home and to state their preferences. As indicated above, they watched the excerpt subtitled in two different ways and then had to answer two questions, one closed and the other open format:

1.   Which translation do you prefer?
2.   Why?

We analyzed their answers trying to find the underlying reasons for their preferences. As mentioned before, each set consisted of one clip using one of the domestication strategies (D), and the other one using one of the foreignization strategies (F). They were shown in a random order (see Table 1).

Table 1. Presentation order

	Clip 1	Clip 2
Set 1	F	D
Set 2	D	F
Set 3	D	F
Set 4	F	D
Set 5	F	D

The third part of the questionnaire consisted of three questions about the participant's attitudes towards the AVT practice and policy of the Radio and Television of Montenegro (RTCG), the public service broadcaster of Montenegro, and the preferred AVT modalities to be used for adults and pre-school children. As with the first part, the results of this one will not be the focus of the present paper.

Having completed the questionnaire, each participant was interviewed regarding their preferences in terms of the translation strategies and possible reasons. Later their responses were compared to the ones obtained during the experiment. At the very end, participants were informed about the purpose of this study.

Similar procedure was followed in the dubbing experiment with the exception of using five excerpts for the reasons mentioned above followed by the same closed and open format questions as with the subtitling experiment.

4.   Results and discussion

4.1 H_1: The majority of the participants will prefer the foreignization approach in subtitling and the domestication approach in dubbing

The overall results of the subtitling ($n=102$) and dubbing ($n=34$) experiment, as well as the follow-up interview support this hypothesis. The subtitling experiment results show that in three out of five sets, the majority of participants preferred the subtitles with the foreignization strategy, while the domesticating strategy was preferred for the two sets (see Table 2).

Table 2. Strategy preferences in subtitling (experiment results)

	Set 1	Set 2	Set 3	Set 4	Set 5
Foreignization	42%	59%	58%	23%	67%
Domestication	58%	41%	42%	77%	34%

This coincides with the results obtained from the follow-up interview for the subtitling experiment (see Figure 1) where 70 (69%) participants said they preferred foreignized, while 32 (31%) preferred domesticated translation. The results seem to suggest that there is a general tendency for preference of foreignization strategies in subtitling.
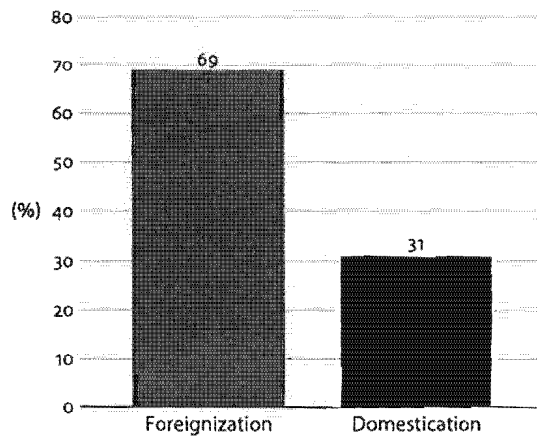
Figure 1. Strategy preferences in subtitling (interview results)

The above preferences are illustrated in the examples that follow. The first one is taken from set 2 and the second one from set 5. The first line in the given examples will be the source text in English, the second one is the subtitled version in Montenegrin, and the third one is a back translation:

(1)   Chandler: I know I'm a little older than you guys, but it's not like I'm Bob Hope.
Znam da sam malo stariji od vas ali nije da sam Čkalja.
Back translation: I know I'm a little older than you guys, but it's not like I'm Čkalja.

In this particular example, Bob Hope, a US comedian that served in the United Service Organization and is well-known especially to older generations in the States, functions as an extralinguistic, monocultural element which is unknown to our participants and not encoded in humor. The domesticated translation by cultural substitution, as shown above, mentions Miodrag Petrović Čkalja, a former Yugoslav comedian that is well-known in the region, especially to older generations. However, most participants preferred the retention strategy ("Bob Houp"), thus a foreignized rendition, over "Čkalja". The majority of answers show that the reason is that the domesticated version does not fit the cultural context of the scene, feels out of place and awkward and is not crucial for appreciating the humor of the scene.

Another example illustrating the preference of the foreignization strategy is the following one:

(2)   Phoebe: No. Actually, you might want to stay away from Jane Street.
      Ne, u stvari, trebalo bi da se kloniš Hercegovačke ulice.
      Back translation: No, actually you should stay away from Hercegovačka Street.

In this example, another instance of extralinguistic element is present, a street in New York (Jane Street), which is mono, or arguably even micro, cultural in its nature and not humor encoded. Even though unknown to our participants, most of them opted for the direct translation strategy ("ulica Džejn") and found the domesticated rendition by cultural substitution given above, "Hercegovačka Street", a well-known street in the capital city of Montenegro, Podgorica, awkward, out of place and not crucial for appreciating the humor of the scene.

In the subtitling experiment, there were some instances of preference of domestication. A closer look at the results and these cases seems to suggest that it is preferred with those elements, both intra and extralinguistic, which are mono-cultural in their nature, and, most importantly, encoded in the ST humor as the following examples illustrate. The first example is taken from set 1 and the second one from set 4:

(3)   Monica: Ross, you did not come up with: "Got milk?"
      Rose, ti nisi smislio slogan: "Moja kravica."
      Back translation: Ross, you did not come up with the slogan: "My little cow!"

In this example, the American advertising campaign slogan "Got milk?" which was created in 1993 by the advertising agency Goodby Agency and Partners from California with the aim of encouraging consumption of cow's milk and featured a number of celebrities in the States, would be unknown to the average viewer of the region and thus functions as a monocultural extralinguistic element. Further-more, it is encoded in the humor of this particular scene when Monica points out that Ross was not the originator of the advertisement. The majority of the answers by the participants show that the foreignizing approach with the direct translation, *Imaš li mlijeka?* was not preferred as this does not recode the humorous effect in the target language. Hence, the issue of humor-encodedness seems to be impor-tant when opting which strategy to use.

Similar situation occurs with the monocultural intralinguistic elements as in the following example:

(4)   Chandler: Well, first of all, they're not called "sneakers" anymore. Apparently, they're called "kicks" or "skids" and I think I heard somebody say "slorps".
      Prvo, više se ne zovu patike. Zovu ih "pumarice", ili "makserice" a čini mi se da je neko nazvao i "tikepa".
      Back translation: First, they are not called sneakers any more. They're called "pumas", or "maxes" and I think someone said "ker-snea".

In this example, the slang words for sneakers, "kicks", "skids" and the non-existing referent to sneakers "slorps", which basically refers to eating and drinking noisily or with a sucking sound, are unknown to the target culture users, and, hence, function as a monocultural, humor-encoded element with no TL equivalent. As for the foreignizing techniques, retention was used. However, most of the viewers preferred the domesticated rendition which is illustrated above in which cultural substitution was used. "Pumarice" and "makserice" are slang words used primarily by younger generations, but understood by all, which refer to specific brands of sneakers. The former refer to the *Puma* and the latter to *Airmax* brands. In the third rendition, a specific feature of permuting syllables in words is used, which is sometimes used in informal speech of younger generations. In this particular example, instead of "patike" meaning "snekares", "tikepa" is used. This is known as *shatra speech* and nowadays is sometimes used in informal speech for fun.

As for the dubbing experiment (see Figure 2), the results and insights from the open-ended questions and interviews show that the majority of participants prefer the domesticating approach in rendering the ECs.
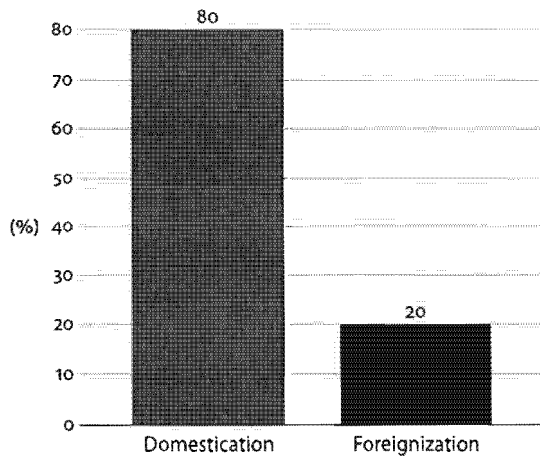
Figure 2. Strategy preferences in dubbing

Insights from the answers and interviews reveal that most of the participants found cultural substitution and the use of diversified language (dialects, specific accents, slang, etc.) extremely effective for transferring the humorous effect, among other things. It was especially with characters that were dubbed with Bosnian accent that the participants found the transfer successful. Here are two examples from *Incredibles* illustrating this:

(5) Laserman's wife: Don't you think about running off doing no derrin'-do.
     Da se ne bi pravio sad malo mačo herojem?
     Back translation: Do you feel like acting as a macho hero now a bit?

(6) Laserman: Tell me where my suit is, woman!
     Đe mi je kostim, kokoško!
     Back translation: Where is my suit, hen!

In this scene, a superhero Laserman, whose voice is dubbed with Bosnian accent, tries to find his super suit as he has to save the city. In rush, he calls his wife and asks where his suit is. Shouting from another room, she replies that she had hidden it and was not going to return it as they had previously planned to go out for a dinner. She uses an idiom ("derring-do") in her answer in Example (5) that is spoken by an African-American in the source text and, thus, functions as an intralinguistic element. In the target text, it is substituted with the Dalmatian dialect and a paraphrase of sense. While in Example (6), Laserman's angry exclamation "woman" is rendered as "kokoško", lit. "hen", a pejorative term used to refer to a woman, especially one who is stubborn and a gossip.

A number of other extralinguistic elements of culture are used in the Croatian dubbing even in places where no such element is used in source text. These include references to some of the ex-Yugoslav rock songs that are well-known in the region (e.g. *To mi radi* by the pop-rock band Crvena jabuka), and even some political figures such as Josip Broz Tito, the president of former Yugoslavia. Personal names and surnames were also domesticated (e.g. Bernie, a primary school teacher, is rendered as *Pero*, a nickname for Petar and a common name in jokes and puns in the region, instead of Bernie). These received a very positive response from the participants of the dubbing experiment. Most of the answers explaining the reasons why they liked the clip refer to the domesticated ECs suggesting that these recode successfully a humorous effect (see Table 3).

Table 3. Positive responses to the use of domesticating techniques in dubbed clips

	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
Percentage of participants with positive responses	86%	78%	60%	33%	93%

As the table shows, each of these results show a high positive response rate to the use of domesticating techniques in dubbing. The only exception is with clip 4. It's worth mentioning, however, that most of the negative responses were due to the fast pace of speaking of the character while at the same time giving positive comments for the use of slang words and local nicknames.

4.2   H2: Specifically in cases of mono or microcultural ECs, the general preferred approach will be a domesticated one

This proved partially to be the case, but only with monocultural ECs. As indicated above, participants preferred being exposed to the original microcultural EC as they felt the domesticated version were out of place and awkward in the given American cultural context of the scene. However, this proved not to be the case with some monocultural humor-encoded ECs. If monocultural EC is not crucial for appreciating humor, then the foreignized version is preferred, as the participants didn't see the need for any intervention other than direct translation, or retention. It is taken for granted that the mono or microcultural intralinguistic ECs (such as idioms) will necessarily have to be domesticated in cases of zero equivalence, otherwise this would not result in a successful communicative transaction.

It's worth mentioning, that the overall results suggest that there isn't a significant difference in preferences neither due to participants' different academic background, nor the level of language proficiency or gender.

5.   Conclusion

To summarize, both the experiment and the interview results of the reception study presented in Section 4 show that: (a) there is a different general tendency in the preferred translation strategy for rendering the ECs between subtitling and dubbing; in subtitling, the general preferred strategy is foreignization, whereas in dubbing domestication; (b) the underlying reasons for the general preferences seem to be:

- the "vulnerability" of the modality (by exposing the end-users to both the ST and TT, subtitling is more "vulnerable" than dubbing where end-users are exposed to the ST version only, which creates different expectations, and users seem to be more judgmental as to the "correctness" of the translation and expect to read what they really hear; furthermore, a number of them said that, apart from enjoying the show, this is a great way of learning the language and expect to see "accurate" translations in subtitles);
- in subtitling, the deviation from the general tendency seems to be conditioned by the degree of encodedness in ST humor of monocultural ECs (if the EC was humor-encoded, the majority of participants preferred the domesticated version, whereas if the EC was not humor-encoded, the majority preferred the foreignized version; this coincides with the results obtained in another study (Luque 2003) which suggest that foreignization reduces the positive reception of verbal humor).

As O'Sullivan (2016: 270) suggests, "[r]eception studies are important as a way of establishing the impact of media policies and potentially as a way of guiding future policy". Thus, the empirical evidence presented here can inform broadcasting companies, marketing and entertainment industry when producing general guidelines for translators. As we have seen, it is important to acknowledge the modality differences and, thus, produce modality-specific, rather than general, translation guidelines, and to avoid the over-simplistic approach to the treatment of such a complex issue as rendering culture within certain modalities. This can help practicing translators to avoid speculations. By following such guidelines informed by audience research product's placement and usability can be enhanced, at least in part affected by the quality of translation.

As a way forward, since this study was limited to a specific genre and audience profile, it would be useful to extend it and carry out a qualitative empirical research with different audiences and genres and see if the results are general tendencies irrespective of age, profession, genre, etc. In particular, it would be useful to shed more light on the raised issues in dubbing with stimuli similar to the ones used in the subtitling part of the experiment, hopefully without the limitations that we have experienced, and thereby deepen our understanding of both translation-induced and user-induced effects on the AV product reception in different modalities.

References

Antonini, Rachele. 2005. "The Perception of Subtitled Humor in Italy". *International Journal of Humor Research* 18 (2): 209–225.
https://doi.org/10.1515/humr.2005.18.2.209

Caffrey, Colm. 2009. Relevant Abuse? Investigating the Effects of an Abusive Procedure on the Perception of TV Anime Using Eye-Tracker and Questionnaire. PhD Thesis. Dublin: Dublin City University.

Denton, John; and Debora Ciampi. 2012. "New Development in Audiovisual Translation Studies: Focus on Target Audience Perception". *LEA – Lingue e Letterature d'Oriente e d'Occidente* 1 (1): 399–422.

Di Giovanni, Elena. 2016. "Closing the Circle of AVT: Analyzing Audiences, Evaluating Reception". Nitra: Paper presented at Audiovisual Translation: Dubbing and Subtitling in the Central European Context (June 15–17).

Diaz Cintas, Jorge; and Aline Remael. 2007. *Audiovisual Translation: Subtitling.* Manchester: St. Jerome.

Fuentes Luque, Adrian. 2003. "An Empirical Approach to the Reception of AV Translated Humour". *The Translator* 9 (2): 293–305.
https://doi.org/10.1080/13556509.2003.10799158

Gambier, Yves. 2006. "Multimodality and Audiovisual Translation". Copenhagen: Paper presented at Marie Curie Euroconferences MuTra (May 1–5).

Gambier, Yves. 2008. "Recent Developments and Challenges in Audiovisual Translation Research". In *Between Text and Image: Updating Research in Screen Translation*, ed. by Delia Chiaro; Christine Heiss; and Chiara Bucaria, 11–13. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.78.03gam

Hlavac, Jim. 2015. "Pre- and post-conflict language designations and language policy: Reconfiguration of language norms amongs translators of Bosnian, Croatian and Serbian languages". *Target* 27 (2): 238–272. https://doi.org/10.1075/target.27.2.04hla

Kovačić, Irena. 1995. "Reception of Subtitles: The Non-Existent Ideal Viewer". *Translation* 14 (3–4): 376–383.

Leppihalme, Rita. 1996. "Caught in the Frame: A Target Culture Viewpoint on Allusive Wordplays". *The Translator* 2 (2): 199–218. https://doi.org/10.1080/13556509.1996.10798974

McAuley, E. Thomas. 2015. "Audience Attitude and Translation Reception". Babel 61 (2): 219–141. https://doi.org/10.1075/babel.61.2.04mac

Orrego-Carmona, David. 2016. "A Reception Study on Non-Professional Subtitling: Do Audiences Notice Any Difference". *Across Languages and Cultures* 17 (2): 163–181. https://doi.org/10.1556/084.2016.17.2.2

Orrego-Carmona, David. 2014. "Where is the Audience? Testing the Audience Reception of Non-Professional Subtitles". *Translation Research Projects* 5: 77–92.

O'Sullivan, Carol. 2016. "Imagined Spectators: The Importance of Policy for Audiovisual Translation Research". *Target* 28 (2): 261–275. https://doi.org/10.1075/target.28.2.07osu

Pedersen, Jan. 2011. *Subtitling Norms for Television: An Exploration Focusing on Extralinguistic Cultural References*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.98

Perego, Elisa. 2016. "History, Development, Challenges, and Opportunities of Empirical Research in Audiovisual Translation". *Across Languages and Cultures* 17 (2): 155–162. https://doi.org/10.1556/084.2016.17.2.1

Požgaj Hadži, Vesna (ed.). 2013. *Jezik između lingvistike i politike* [Language Between Linguistics and Politics]. Belgrade: Biblioteka XX vek.

Schauffler, Svea. 2012. Investigating Subtitling Strategies for the Translation of Wordplay in Wallace and Gromit – An Audience Reception Study. PhD Thesis. Sheffield: University of Sheffield.

Schutze, Carson T.; and Jon Sprouse. 2013. "Chapter 3: Judgment Data". In *Research Methods in Linguistics*, ed. by Robert J. Podesva; and Devyani Sharma, 27–50. New York: CUP.

Touminen, Tiina. 2012. The Art of Accidental Reading and Incidental Listening: An Empirical Study of the Viewing of Subtitled Films. PhD Thesis. Tampere: University of Tampere.

Résumé

La recherche empirique sur la réception dans la traduction audiovisuelle (TAV) a longtemps été négligée, parce que la plupart des études antérieures se concentraient principalement sur les caractéristiques de la TAV en tant que produit et sur les producteurs d'une TAV, et non sur les destinataires finaux. L'importance de ces études a été mise en évidence par le fait que le contenu audiovisuel était conçu pour des publics cibles. Connaitre leurs préférences, leurs attentes et leurs besoins permet d'informer l'industrie et, par conséquent, d'accroître la réception positive, le placement et la convivialité du produit. Le présent article a pour but de répondre à la question de savoir s'il existe des différences de préférences en ce qui concerne les stratégies de traduction utilisées pour rendre les éléments culturels (EC) dans les deux modes de traduction audiovisuelle (sous-titrage et doublage) et quelles en seraient les raisons sous-jacentes. Dans cet objectif, nous avons effectué une expérience avec 136 étudiants de deux universités du Monténégro. Les résultats montrent que les destinataires finaux ont des préférences et des attentes générales différentes pour ces modes. Nous supposons que les différences observées sont liées à la « vulnérabilité » du sous-titrage. En outre, des fluctuations par rapport à la préférence générale du sous-titrage ont été notées dans les EC monoculturels, en ce qui concerne l'encodage des EC dans l'humour du texte source. Cela montre que les directives de traduction de l'industrie doivent être spécifiques au mode et qu'il faut éviter de traiter de manière trop simpliste un problème aussi complexe que le transfert d'éléments culturels.

Mots-clés: études sur la réception, sous-titrage, doublage, éléments culturels

Address for correspondence

Petar Božović
University of Montenegro
Danila Bojovića b.b.
81400 Nikšić
Montenegro
petarb@ugc.ac.me

Zbornik konference
Jezikovne tehnologije in digitalna humanistika

Proceedings of the conference on
Language Technologies & Digital Humanities

Uredila / Edited by:
Darja Fišer, Andrej Pančur

ZBORNIK KONFERENCE
JEZIKOVNE TEHNOLOGIJE IN DIGITALNA HUMANISTIKA

PROCEEDINGS OF THE CONFERENCE ON
LANGUAGE TECHNOLOGIES & DIGITAL HUMANITIES

Opus-MontenegrinSubs 1.0: First electronic corpus of the Montenegrin language

Petar Božović,* Tomaž Erjavec,† Jörg Tiedemann,‡ Nikola Ljubešić,§ Vojko Gorjanc‡

* English Language and Literature Department, Faculty of Philology, University of Montenegro
Danila Bojovića bb, 81400 Nikšić
petarb@ac.me

†Department of Knowledge Technologies, Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana
tomaz.erjavec@ijs.si

‡Department of Linguistics and Philology, Uppsala University
BOX 635, SE-751 26 Uppsala, Sweden
jorg.tiedemann@lingfil.uu.se

§Department of Knowledge Technologies, Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana
nikola.ljubesic@ffzg.hr

‡Department of Translation Studies, Faculty of Arts, University of Ljubljana Aškerčeva 2, 1000 Ljubljana
vojko.gorjanc@ff.uni-lj.si

Abstract

Although recent years have witnessed a growth in the number of computational language resources and tools, a lot still needs to be done, especially with low-density languages. This is the case with all South Slavic languages and especially Montenegrin, the fourth standard of the once Serbo-Croatian language that has been re-codified only recently. Even though it became the official language of Montenegro in 2007, there still isn't any publicly available electronic corpus that would be available for empirical research of linguistic, translatological or any other inquiry. This paper introduces the first publicly available English – Montenegrin parallel corpus of subtitles. It describes the process of corpus compilation, presents linguistic annotation and accessibility of the corpus through web concordancers. Furthermore, it gives a brief overview of linguistic situation in Montenegro with some of the most important recent developments especially in the light of the recent official international recognition of the language which took place in December 2017.

1. Introduction

Recent years have witnessed a growth in the number of machine-readable corpora and language tools for a number of world languages. It is currently estimated that there are 7,097 languages in the world. This is an updated number of officially recognized languages listed by Ethnologue[1] which is to be taken arbitrarily. Out of this number, however, in 2006 there were corpora available for less than 1% of all world languages, and 20-30 of these fall into the category of high-density and medium-density languages, where "density" is understood to represent the number of computational resources available (Maxwell and Hughes, 2006). The first group included a handful of languages only, among which English, German, Arabic, etc. Today, the number of available resources has increased to app. 90 languages, which means app. 1.2% of all world languages having any kind of publicly available computational resource. The majority of these, however, are lower-density languages as resources are rather scarce.

Being a multilingual and integrative society, Europe is estimated to cover more than 80 languages, of which 23 are official and the rest are either minority or immigration languages. However, quite a number of these languages is technologically not supported sufficiently and run the risk of being marginalized or digitally extinct. Thus, various initiatives have been introduced, such as the META-NET Strategic Research Agenda for Multilingual Europe 2020, with the aim of using various language technologies for overcoming language barriers, enabling free flow of information, goods, and innovations, and thus creating a single digital space and marketplace[2].

When it comes to electronic language resources and corpora for some of the major official languages of former Socialist Federal Republic of Yugoslavia, Bosnian, Croatian, Montenegrin, Serbian, Slovenian (BCMSS), the majority of resources are available for Slovenian, followed by Croatian, Serbian, and Bosnian. The last language standard based on Serbo-Croatian or Croato-Serbian language that has recently been re-codified and internationally recognized in December 2017 is Montenegrin. So far, there have not been any electronic public corpora of any kind available for the study of this standard. This all testifies to the fact that there is still a lot of work to be done.

This paper presents Opus-MontenegrinSubs 1.0, the first parallel English-Montenegrin electronic corpus developed as a joint effort of researchers from the University of Montenegro, Jožef Stefan Institute, University of Helsinki, and University of Ljubljana. First we will briefly discuss the potentials and possible applications of parallel corpora, the specifics of subtitle corpora as a sub-type of parallel corpora, followed by an outline of the current available parallel corpora for BCMSS. Since this is the first electronic resource of Montenegrin language that is being presented, we will give a brief overview of the linguistic situation in Montenegro followed by the description of the corpus itself and the first study based on it.

---

[1] http://www.ethnologue.com, accessed: March 15th, 2018, at 11:30 pm.

[2] http://www.meta-net.eu/sra/key-messages, accessed: March 30th, 2018, at 1:00 pm.

2. Parallel subtitle corpora for BCMSS: the potential, current state and specifics

Parallel corpora have found a number of applications in linguistics, translatology, translation practice, and beyond. They proved to be an indispensable tool for a number of contrastive linguistic studies, word sense disambiguation and construction of lexicons, as well as an input for parallel concordancing systems. Furthermore, the use of parallel corpora has especially become a trend in translation and interpreting studies for developing and training statistical machine translation systems (which require large amounts of parallel language data), for the study of regularities of translations and translators, translation teaching and learning, translation practice including terminology extraction, identifying translation equivalents and correspondents, translation quality assessment, etc. (Bywood et al 2013; Hu, 2016). However, as noted by Tiedemann (2007), most of the existing parallel corpora cover the high-density languages and the domains of legislation, administration and technical documentation.

With regards the parallel subtitle corpora for BCMSS, most of them were developed as subcorpora within the Opus2 project (Tiedemann, 2009) and for each of the western South Slavic languages they include:
- Bosnian: subcorpus OpenSubtitles 2011 (tokens 26,491,099, words ~ 20,906,596),
- Croatian: subcorpora OpenSubtitles 2011 (tokens 111,981,881, words ~ 86,600,021), TedTalks (tokens 1,285,011, words ~ 993,749),
- Serbian: subcorpus OpenSubtitles 2011 (tokens 154,063,822, words ~ 119,149,120),
- Slovenian: subcorpus OpenSubtitles 2011 (tokens 109,690,961, words ~ 81,500,854).

Another project that involved the creation of subtitle corpora for Slovenian and Serbian was SUMAT (tokens 1,250,000/ 1,500,000) (Bywood et al., 2013, Fishel et al., 2012).

Apart from the above mentioned possible applications of corpora of this type, subtitles can be used for the study of text compression and summarization. The reason for this are the unique features of subtitles that make them a specific language resource in many ways. Subtitles are usually transcriptions of spontaneous speech with a diversified language (genre, slang, colloquialisms) and they can be classified into several categories: interlingual and intralingual (depending whether they represent a translation from a source to a target (foreign) language, or they are in the same language as the source audiovisual text); monolingual or multilingual (depending on the number of translations into different languages which are shown on the screen); pre-recorded and live. Interlingual subtitling, which we refer to when we use the term in this paper, is a specific form of translation practice since subtitles per se are "a vulnerable modality" for various reasons (Diaz Cintas and Ramael, 2007). This is primarily the case because viewers are exposed both to the source and target text, and there are specific time and space constraints: they are usually shown in one or two lines with 30 – 40 characters, cca. 3 – 7 seconds only with no room for annotation. This calls for specific translation strategies such as: *condensation* (it is estimated that subtitles are 40 – 75% shorter than spoken version), *omission* (especially in cases of redundancies and spoken discourse markers such as exclamations, false starts, repetitions, hesitations, question tags, etc.), cultural substitution, generalization and specification. Moreover, standardization is also used frequently (especially in cases of slang, regionalisms, grammar mistakes, etc.), and occasional censorship. This shows that subtitles should be approached as a specific, yet an important and unique, resource of translated language.

3. Linguistic situation in Montenegro: a brief overview of recent history

As previously mentioned, Montenegrin language is the last out of four re-codified standards that stem from the same linguistic base of the polycentric Serbo-Croatian language. The remaining three include Bosnian, Croatian and Serbian. Similarly to these standards, it is based on the Eastern Herzegovian Shtokavian dialect. As it has been the case with most neighboring countries, the situation in language policy in Montenegro reflected a rather turbulent political situation in the Socialist Federal Republic of Yugoslavia (SFRY), and later the Socialist Republic of Yugoslavia (SRY) and the union of Serbia and Montenegro (SM), only to reach its current state in the post-2006 period when Montenegro became independent.

During the pre-1991 period, Serbo-Croatian was one of the official languages of the SFRY, together with Slovenian, Macedonian and other languages which were constitutionally of an equal status, but the reality was somewhat different as they seem to have been in a position of a "competitive coexistence" (Gorjanc, 2013; Požgaj Hadži et al., 2013). In the light of the above mentioned historical events, the constitutions of the Socialist Republic of Montenegro of 1963 and 1974 define Serbo-Croatian as the language in official use in Montenegro. This polycentric language, as its very name suggests, stemmed from two main standards: the eastern one (with its center in Belgrade) and the western (with its center in Zagreb), while other language forms with their center in Podgorica (i.e. Titograd, as the capital of Montenegro was called then) and Sarajevo were marginalized as regional variants, and re-codified as provincialisms. This situation would later be seen as having significant political implications especially regarding the politics of assimilation and hegemony.

With the disintegration of Yugoslavia in 1991, Serbo-Croatian was re-codified into 4 separate standards, starting with Serbian and Croatian, and later followed by Bosnian and Montenegrin. Due to the official state policy of the day, the official language of Montenegro in the constitution of 1992 was designated as Serbian of the Ijekavian standard, and this remained the case until 2007. Shortly after the independence which took place in 2006, and upon the ratification of the new Constitution which took place on 22 October 2007, Montenegrin became the official language in Montenegro. After much controversy arising from two different approaches to the process of standardization, the first Montenegrin grammar and orthography were adopted in 2010 by the Council for General Education. The question of language standardization still remains an ongoing issue and it's highly debatable whether some of solutions which are found in the mentioned orthography and grammar book will fully integrate into language practice.

Census data from 2011 shows an increase in the number of speakers who designate their mother tongue as Montenegrin. One of the most significant events was the international recognition of Montenegrin language and the assignment of the international code. This was approved on

8 December 2017 and the ISO 639-2 and 3 code [cnr] was assigned. Needless to say, much still needs to be done. The first volume of the Dictionary of the Montenegrin Literary and Vernacular Language published by the Montenegrin Academy of Sciences and Arts in 2016 was soon withdrawn due to reactions of part of political public to some entries that seemed to be controversial. Regarding the electronic language resources, up to this date, there haven't been any officially published electronic corpora of Montenegrin language that would be available to researchers for various linguistic and translatological studies. That is why the electronic corpus which we will present in this paper is of high importance as it is the first electronic corpus of Montenegrin language.

4. Corpus compilation and accessibility

The corpus Opus-MontenegrinSubs 1.0 contains parallel English-Montenegrin subtitles. The data and copyrights were obtained from the Radio and Television of Montenegro, the public service broadcaster of Montenegro. The corpus consists of English and Montenegrin subtitles of three series: House of Cards, Damages, and Tudors. The corpus contains 10 seasons, and 110 episodes, which are cca. 5,563 minutes in length. A detailed breakdown is given in Table 1.

Series	No. of seasons	No. of episodes	Length
House of Cards	1	13	686 mins.
Damages	5	59	2878 mins.
Tudors	4	38	1999 mins.

Table 1: Corpus breakdown

4.1. Processing the corpus

Sentence alignment and basic encoding was performed inside the OPUS project3. The original subtitle files were converted to Unicode UTF-8 using *iconv* and the Unix tool *file* for automatic detection of the character set in the original file. After that the OPUS subtitle tools (Lison & Tiedemann, 2018) were applied to convert the files to standalone XML with sentence markup; the remaining XML-well formedness problems were fixed with the program *tidy*. Finally, all translated subtitles were aligned using the time-based alignment method described in Tiedemann (2007) and the standard OPUS import pipeline was used to integrate the data in OPUS with download formats in XML, plain text and TMX.

In the second stage, the source XML data was converted to the latest version of TEI (TEI, 2018), so that the subtitles for each language are stored in a separate <text> element, with the sentence alignments being maintained by cross-links as well as separately, in a <linkGrp> element. An important part of this conversion was also the encoding of the <teiHeader> element, which contains the meta-data of the corpus, explicating its authors, license etc. but also listing all the used XML elements in the corpus, together with a short explanation, and how the MSD annotation prefixes are to be interpreted.

Then, the English and Montenegrin texts were tokenized, sentence segmented and tagged with morphosyntactic descriptions (MSDs) and lemmas. To perform this annotation for Montenegrin, we used the ReLDI tokeniser[4] and tagger[5] (Ljubešić & Erjavec, 2016) with its model for Serbian. The MSD tagset used follows the MULTEXT-East specifications (Erjavec, 2012), in particular, the version 5 specifications for Bosnian[6].

For English, we used Tree Tagger (Schmidt 1994, 1995) with its model for English, which uses the Penn Tree Bank tagset. In order to make the English tagset harmonized with the one for Montenegrin, we converted it to the SPOOK tagset for English[7], i.e. performed a 1-1 mapping between the original PTB tagset to MULTEXT-East compatible SPOOK tagset.

Figure 1 illustrates the TEI encoding of the linguistically annotated corpus, giving the first translation unit (annotated as anonymous block, <ab>) for both languages. As can be seen, each language text contains the divisions marking the structure of the corpus, while the translation units are given IDs and the alignment via their @corresp attribute. Each translation unit is then divided into sentences, and these into words, punctuation symbols and whitespace. The tokens are lemmatized and MSD tagged, where the value prefix *mte* resolves to the MULTEXT-East MSD definition (i.e. its decomposition into features), while the *spook* one resolves to the SPOOK decomposition. It should be noted that the original PTB tag is retained as the value of the @*function* attribute.

4.2. Corpus distribution and use

The TEI corpus was converted to the so called vertical format, used by (no)Sketch Engine and mounted on the CLARIN.SI concordancers,[8] namely noSketch Engine and KonText, as well as Sketch Engine, so that it is available on-line for searching and exploration; both concordancers also allow displaying the aligned translation units.

The complete corpus in TEI, as well as vertical format, was also made available for download in the CLARIN.SI repository (Božović et al., 2018) under the Creative Commons - Attribution-ShareAlike license.

5. First corpus studies

The first study based on this corpus is the one conducted for the purpose of the Ph.D project by Petar Božović with the thesis topic *Audiovisual Translation and Elements of Culture: A Comparative Analysis of Transfer with Reception Study in Montenegro*, which is in the field of translation studies and corpus linguistics. Corpus-based translation studies are becoming increasingly relevant for the industry ever since the methodology was first introduced from an allochthonous field of corpus linguistics in the seminal paper by Baker (1993). It wasn't long after this that it became evident that using corpora in translation research was to have a great potential for scholarly empirical research, but also for terminologists and practitioners.

```
<text xmlns="http://www.tei-c.org/ns/1.0" xml:lang="cnr">
  <body>
    <div type="series" xml:id="Damages-cnr">
      <div type="season" xml:id="Damages.S1-cnr">
        <div type="episode" xml:id="Damages.S1.dam0101-cnr">
          <ab n="1" xml:id="Damages.S1.dam0101.SL1-cnr" corresp="#Damages.S1.dam0101.SL1-en">
            <s>
              <w ana="mte:Agpfpny" lemma="opasni">OPASNE</w><c> </c>
              <w ana="mte:Ncfpn" lemma="igra">IGRE</w><c> </c>
              <w ana="mte:Ncmsn" lemma="pilot">Pilot</w><c> </c>
              <w ana="mte:Ncfpg" lemma="epizoda">epizoda</w>
            </s>
          </ab>
  ...

<text xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
  <body>
    <div type="series" xml:id="Damages-en">
      <div type="season" xml:id="Damages.S1-en">
        <div type="episode" xml:id="Damages.S1.dam0101-en">
          <ab n="1" xml:id="Damages.S1.dam0101.SL1-en" corresp="#Damages.S1.dam0101.SL1-cnr">
            <s>
              <w lemma="Season" function="NP" ana="spook:Np-s">Season</w><c> </c>
              <w lemma="1" function="CD" ana="spook:M-c">1</w><c> </c>
              <w lemma="episode" function="NN" ana="spook:Nc-s">Episode</w><c> </c>
              <w lemma="1" function="CD" ana="spook:M-c">1</w><c> </c>
              <w lemma="pilot" function="NN" ana="spook:Nc-s">Pilot</w><c> </c>
              <pc function="(" ana="spook:Z">(</pc>
              <w lemma="Dimension" function="NP" ana="spook:Np-s">Dimension</w>
              <pc function=")" ana="spook:Z">)</pc>
            </s>
          </ab>
  ...
```

Figure 1: TEI encoding of the corpus texts

Figure 2: Searching the corpus in the KonText concordander

Hence, the mentioned research is focused on the highly-influential and fast-growing audiovisual translation field with the aim of mapping the different translation strategies used by translators for rendering the extralinguistic elements of culture in subtitling. This is an issue that is at the core of some of the major challenges in the industry as the transfer of elements of culture proves to be one of the "crisis points" in translation process,

especially in subtitling due to the time and spatial constraints of the modality which were mentioned before. Furthermore, this specific aspect of the translation process can have an important influence on the reception and placement of the audiovisual product on the target market and for the target audience (Pedersen, 2011). The extralinguistic elements in the study were extracted as types, not tokens, by using one of the concordancers. After that, these elements were categorized according to the level of transculturality, and, finally, translation strategy for rendering each element was defined. The goal is to gain a better understanding and map the different ways of rendering culture in subtitling from English into Montenegrin and to supplement this with the reception study. From a practical perspective, it is hoped that this will provide an important empirical feedback for translators and broadcasting companies who could tailor the translation policy better to meet the needs and expectations of the real, not ideal or intuitive, target audience.

6. Conclusions

The Opus-MontenegrinSubs 1.0 is the first publicly available parallel electronic corpus of Montenegrin language the appearance of which is timely considering the recent sociolinguistic developments, especially constitutional and international acknowledgement that this language has received. Needless to say, a lot still remains to be done in order to provide the computational resources and tools necessary for state-of-the-art linguistic approaches and analyses. It is hoped that this corpus will encourage other researches and contribute to the affirmation and development of corpus linguistics and corpus-based translation studies in the region. Moreover, it is hoped that it will encourage the development of other corpora of Montenegrin language, first and foremost of the reference corpus, which would be of a pivotal importance for the process of restandardization and without which a modern linguistic description of Montenegrin will not be possible.

7. References

Jan Pedersen. 2011. *Subtitling Norms for Television: An Exploration Focusing on Extralinguistic Cultural References*. John Benjamin Publishing.

Jörg Tiedemann. 2007. Improved Sentence Alignment for Movie Subtitles. In *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP'07)*.

Jörg Tiedemann. 2007. Building a Multilingual Parallel Subtitle Corpus. In *Proceedings of the 17th Conference on Computational Linguistics in the Netherlands (CLIN 17)*. Leuven, Belgium.

Kaibao Hu. 2016. *Introducing Corpus-based Translation Studies*. Springer.

Lindsay Bywood, Martin Volk, Mark Fishel, and Panayota Georgakopoulou. 2013. Parallel subtitling corpora and their applications in machine translation and translatology. *Perspectives: Studies in Translatology*, 21(4): 595–610.

Mike Maxwell and Baden Hughes. 2006. Frontiers in linguistic annotation for lower-density languages. In *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006*, pages 29–37, Sydney, Australia. Association for Computational Linguistics.

Mona Baker. 1993. Corpus Linguistics and Translation Studies – Implications and Applications. In: Mona Baker et al., ed., *Text and Technology: In Honor of John Sinclair*, pages 233–252, John Benjamins Publishing.

Nikola Ljubešić and Tomaž Erjavec. 2016. Corpus vs. lexicon supervision in morphosyntactic tagging: the case of Slovene. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016*, Portorož. http://www.lrec-conf.org/proceedings/lrec2016/pdf/811_Paper.pdf

Petar Božović, Tomaž Erjavec, Jörg Tiedemann, Nikola Ljubešić, and Vojko Gorjanc. 2018. *English-Montenegrin parallel corpus of subtitles Opus-MontenegrinSubs 1.0*, Slovenian language resource repository CLARIN.SI, http://hdl.handle.net/11356/1176.

Pierr Lison and Jörg Tiedemann. 2018. OpenSubtitles2018: Statistical Rescoring of Sentence Alignments in Large, Noisy Parallel Corpora. *LREC 2018*.

Ranko Bugarski. 2009. *Nova lica jezika*. Biblioteka XX vek.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*. Manchester.

Helmut Schmid. 1995. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin.

Tomaž Erjavec. 2012. MULTEXT-East: morphosyntactic resources for Central and Eastern European languages. *Language resources and evaluation*, 46(1): 131–142, doi: 10.1007/s10579-011-9174-8

Vojko Gorjanc. 2013. Slovenačka jezička politika i odnosi društvene moći. In: Vesna Požgaj Hadži, ed., *Jezik između lingvistike i politike*, pp.12–36. Biblioteka XX vek.

Vesna Požgaj Hadži, Tatjana Balažic Bulc, and Vlado Miheljak. 2013. Srpskohrvatski jezik iz slovenske perspektive. In: Vesna Požgaj Hadži, ed., *Jezik između lingvistike i politike*, pages 12–36. Biblioteka XX vek, Beograd.

TEI Consortium, eds. 2018. *TEI P5: Guidelines for Electronic Text Encoding and Interchange 3.3.0*. TEI Consortium. http://www.tei-c.org/Guidelines/P5/