Using Example-Based Machine Translation to translate DVD Subtitles

Marian Flanagan
Centre for Translation and Textual Studies
School of Applied Language and Intercultural Studies
Dublin City University, Ireland
marian.flanagan@dcu.ie

Abstract

Audiovisual Translation (AVT), and in particular subtitling, has been recognised as an area that could potentially benefit from the introduction of machine translation (followed by post-editing) [1],[2],[10],[15],[27],[28]. In recent years the demands on subtitlers have increased, while the payment to subtitlers and time allotted to produce the subtitles have both decreased. Therefore this market is recognised as a potential real-world application of MT. Recent publications have introduced Corpus-Based MT approaches to translate subtitles [15],[27],[28]. An SMT system has been implemented in a Swedish subtitling company to translate between Swedish and Danish and Swedish and Norwegian subtitles, with the company already reporting a successful return on their investment [27]. The hybrid EBMT/SMT system used in the current research, on the other hand, remains within the confines of academic research, and the real potential of the system is not fully known. In this paper we report on the success to date of using this hybrid system to translate English-German DVD subtitles, based on a human evaluation of the system output using real end-users of the subtitles. The subtitles are evaluated in terms of intelligibility and acceptability. We then comment on the developments that are required to make the EBMT/SMT system more accessible for use within the AVT industry, and to put the system in a position where it could compete commercially with the SMT approach.

1 Introduction

The digital challenge is a major challenge Gambier [11] believes the domain of Audiovisual Translation (AVT) is now facing. Digitalisation has meant significant changes for the subtitling industry. These changes include the introduction of the DVD (Digital Versatile Disk), increased time pressures on subtitlers, and a reduction in payment to subtitlers. In addition to digitalisation, subtitlers are also faced with a recently emerging approach to subtitling known as the template-based approach. Using this approach, subtitlers are provided with templates[1] which include intralingual subtitles[2] in the source language and time-codes. The subtitlers are required to fill in the target language subtitle translation. This approach is adopted in order for human subtitlers to efficiently produce subtitles simultaneously in multilingual versions and within short time frames. The creation of subtitles using templates is usually conducted while the subtitler watches the movie being subtitled [6]. Carroll [6] points out that such a template could make sense if it was thoroughly researched and well-timed. She adds that if subtitlers are free to use the template as an aid and are not compelled to force their translation into the template provided, this approach offers clear advantages. However, she also points out the possible disadvantages of this approach adding that the rigidity of such files can result in poor subtitling with little adherence to now common standards of good subtitling practice [18]. Another issue is that sometimes subtitlers are required to fill in subtitle translations in these templates without actually watching the movie. The approach can be seen as the beginning of the profession moving towards (semi)-automating translation as it has a considerable standardising impact on subtitles across different languages because it provides the source input in a fixed manner. The increasing use of technology motivates in part the use of machine translation in the production of subtitles.

[1] These templates are created by either the DVD distributor or the subtitling company. The task lies with whoever creates the intralingual subtitles from the original language soundtrack.
[2] Intralingual subtitles are written in the same language as the original spoken dialogue. They are most often used by hard-of-hearing viewers and language learners.

Mikel L. Forcada, Andy Way (eds.)
Proceedings of the 3rd Workshop on Example-Based Machine Translation, p. 85–92
Dublin, Ireland, November 2009

2 Characteristics of Subtitles

The growing interest in automating the subtitling process invites us to examine the characteristics of subtitles that are deemed to make their translation amenable to automation. Hardmeier and Volk [15] mention that according to Becquemont [3], "the characteristics of subtitles are governed by the interplay of two conflicting principles: *unobtrusiveness* and *readability.*" This means that subtitles should allow viewers to understand the meaning of the dialogue without detracting from enjoyment of the movie. There are spatial and temporal constraints applied to subtitling, meaning subtitles tend to be short and written in a more simple form than sentences in printed sources. The norm for subtitles is that they span one or two lines, and contain up to a maximum of 80 characters (which results in an average reading speed of 175 words per minute) [18]. These reading speeds were established for television audiences. With the advent of the DVD, norms and conventions have evolved once again and reading speeds are being increased. For example 180 words per minute is taken to be the norm with some companies applying even higher rates [7]. Some subtitles can, of course be shorter than 80 characters. Hardmeier and Volk [15], for example, maintain that many subtitles contain only two or three words to signify affirmation, negation or abuse. Brief subtitles can also be expected to be syntactically simple, exhibiting few if any instances of long-range dependencies. Such characteristics make automatic translation of subtitles easier. If some subtitles are ill-formed according to standard grammars, this is not particularly problematic given contemporary Corpus-Based MT techniques, which do not in any case conduct linguistic analysis of input.

3 Example-Based Machine Translation

In this paper the system used to translate the subtitles is called MaTrEx (Machine Translation by Example), and it is essentially a hybrid 'example-based SMT' system [14]. The system uses the Marker Hypothesis method [13] to extract the sub-sentential chunks used to create the example database. The idea behind the Marker Hypothesis is that languages are marked for syntactic structure at the surface level and can therefore be broken up into smaller, yet still useful segments or chunks, which can then be recombined to form new correct sentences or segments. The system is a modular data-driven MT engine, built following established design patterns [12] and consisting of a number of extendible and re-implementable modules. The four most important of these modules are outlined in [24]:

Word alignment module: This module takes as its input a segment-aligned corpus and outputs a set of word alignments.

Chunking module: This module takes as its input a segment-aligned corpus and produces source and target chunks.

Chunk alignment module: This module takes in the source and target chunks and aligns them on a segment-by-segment level.

Decoder: This module searches for a translation to a new input using the original aligned corpus and derived chunk and word alignments.[3]

An important thing to observe here is the modular design implemented in the system. This means that all the main modules mentioned above can be easily re-implemented, extended or adapted to allow the integration of existing software, such as the use of wrapper technologies, a technique whereby an interface is created around an existing piece of software [14].

[3] An in-depth analysis of the MaTrEx system is outside of the scope of this paper, and therefore the reader is advised to consult previously published literature on the MaTrEx system including [24] and [17].

3.1 Subtitling Corpora

The data-driven approaches to MT rely on the availability of a sententially-aligned bilingual corpus on which to train the system to extract and store source-target sub-sentential alignments at a later stage [1]. It has been mentioned in the literature [23],[27] that aligned subtitle corpora are generally unavailable. Given this lack of availability, we compile three training corpora (A, B and C) consisting of DVD subtitles for use with the MaTrEx system in order to investigate the relationship that might exist between the profile of the training corpus and the quality of the MT output.

Corpus A is a bilingually-aligned corpus of subtitles taken from the first four *Harry Potter* movies in the series that is available to view on DVD.[4] This corpus belongs to the fantasy genre[5] and it contains approximately 7,000 subtitles. Corpus B is made up of a combination of Corpus A and subtitles taken from the *Lord of the Rings* trilogy. Like Corpus A, Corpus B belongs to the fantasy genre. However, the genre has been expanded slightly to include two different kinds of fantasy movies. Corpus B contains approximately 11,400 subtitles. And lastly, Corpus C is made up of a combination of Corpus B and subtitles taken from twenty-five movies on DVD from various genres ranging from action/adventure to romance and period dramas. Corpus C contains approximately 42,300 subtitles. The three corpora thus differ in size and in homogeneity, with Corpus A being the most homogeneous. They also differ in the number of repeated source language (SL) segments they contain, with Corpus C, the largest corpus, containing the highest number of repeated SL segments.

To create the corpora, English and German subtitles are extracted from DVDs, saved in text format and aligned. The average subtitle length for Corpus A subtitles is 4.92 words, for Corpus B subtitles it is 5.89 words, and for Corpus C 5.16 words. The training corpora are significantly smaller than those used in MT evaluation campaigns, e.g. CESTA and IWSLT, but this is understandable when the subtitle corpora are created in an academic setting. In contrast to this, very large training subtitle corpora (approximately 4 million aligned subtitles) were created for use with the SMT system, given their collaborative work with a large subtitling company [15],[27],[27].

4 Evaluation

In the next few sections we describe the methods employed to evaluate the EBMT output and the results gathered. Multimodal texts, such as subtitles, differ from 'general' texts, given that there are three semiotic channels of communication involved in the understanding of the subtitles: image, sound and text, and subtitles are not created to be read in isolation. Based on this, it might be assumed that some of the usual human-based evaluation methods used to evaluate the quality of the MT output would not be suitable in this instance.

In recent times the use of automatic metrics to gauge the quality of MT output has become normal practice, with BLEU scores being described as the "*de facto* standard in machine translation evaluation" [5]. But even when automatic metrics are used in the MT research community, they still come second to human judgements of MT output, with automatic scores described as an imperfect substitute for human assessment of translation quality [5]. Given the lack of published literature on the relationship between automatic metrics and the quality of automatically-generated subtitles, we develop a human evaluation methodology, whereby the evaluators consist of end-users of the subtitles. In addition, BLEU scores are generated so that MT researchers involved in automatic evaluation can rank the results in relation to other subtitle evaluation studies that use such scores, and so that the relationship between human evaluations of subtitles and BLEU scores can be ascertained.

The test sets used in the study consist of six 2-minute movie clips taken from the first *Harry Potter* movie [16]. Each movie clip contains between 23 and 36 subtitles. The English subtitles are translated from English into German using the MaTrEx system, and are then added to the clips using Subtitle

[4] These were the only Harry Potter DVDs available at the time of compiling the corpora (January 2007).
[5] The category of genre for each of the corpora is determined by the genre of the DVDs contained in the corpora. This category is assigned by the movie distributors.

Workshop.[6] Of the six clips, three have an English language soundtrack (a language known to the human evaluators) and three have a Dutch language soundtrack (unknown language). 44 German native-speakers participated in the evaluation sessions, which were conducted in 2007.[7] Given that there were three training corpora, this resulted in three sets of MT-generated subtitles. 15 subjects viewed clips with subtitles generated using training Corpus A, 15 subjects viewed clips with subtitles generated using training Corpus B, and 14 subjects viewed clips with subtitles generated using training Corpus C. After each clip was shown, the subjects were interviewed and asked questions from a structured questionnaire, containing open and closed questions. The collated data were analysed quantitatively (statistical analysis) and qualitatively (coding process).

5 Results

The data collected during the evaluation sessions are used to answer three questions regarding the EBMT-generated subtitles from an end-user's perspective. We look at whether increasing levels of SL repetitions between the test and training data, increasing the size of the corpus, and decreasing the homogeneity of the corpus have a significant impact on the intelligibility and acceptability of EBMT-generated subtitles. The intelligibility of the subtitles is operationalised through comprehensibility and readability, and the acceptability of the subtitles is operationalised through style and well-formedness of the subtitles. We argue that intelligibility is a necessary but not a sufficient condition for acceptability of subtitles.

Firstly the findings are evaluated from a quantitative perspective. The quantitative data for comprehensibility was not deemed statistically significant. That said subtitles produced by Corpus B are ranked the highest of the three in terms of comprehensibility (1 being the least comprehensible, 6 being the most comprehensible), as outlined in Table 1. In terms of readability, the proportion of Corpus B subjects who believe the perceived speed of the subtitles is unsuitable is significantly different from the proportion of Corpus C subjects who believe it is unsuitable. The data for Corpus C show that the lowest percentage of subjects deem the subtitle speed as too fast. This statistically significant result deems the subtitles generated by Corpus C as more readable than those generated by either of the other two corpora. There are two statistically significant results in terms of the style[8] of the subtitles. A statistically significant number of subject viewing Corpus A subtitles notice something amusing in the subtitles. The majority of amusing items observed by subjects were negative, including the unusual use of adjectives and some subjects stating that they were not able to understand *any* of the subtitles. The second result shows that subjects viewing Corpus A subtitles deem the style of the subtitles as more inappropriate than subjects viewing subtitles generated by Corpus B, as outlined in Table 2.

Based on these data Corpus B subtitles are considered to be written in the most appropriate style, even if this finding seems slightly counterintuitive given the greater homogeneity of Corpus A. In relation to the well-formedness of the subtitles, subjects viewing subtitles generated by Corpus C noticed a significantly higher number of class 1 errors. Errors noted by subjects were divided up according to Flanagan's [9] MT error classification, with class 1 errors having the least effect on intelligibility and class 3 errors having the most effect on intelligibility. Therefore the high number of class 1 errors does not necessarily render subtitles from Corpus C less 'intelligible'. This point is supported by the mean score results for errors, which did not seem to be affected by the significant difference in the number of class 1 errors noted by subjects viewing subtitles generated by Corpus C, shown in Table 3. However, the results do suggest that the subtitles generated by Corpus C are less well-formed than the subtitles generated by either Corpus A or Corpus B, thus affecting the acceptability of Corpus C subtitles.

[6] http://www.urusoft.net/products.php?cat=sw

[7] The MT output used for the evaluation sessions was generated in June 2007. The MaTrEx system has since been modified, and the reader is advised to consult recent publications on the system including [8] and [22].

[8] The style characteristic used in this study is synonymous with register, as defined in the Framework for Evaluation of Machine Translation in ISLE (FEMTI) model. The style of the MT output is evaluated in terms of appropriateness and naturalness in the given context. We acknowledge, however, that style in the context of AVT studies concerns the manner of speech, expressions used and characterisation, and human translators usually use "compensation" strategies to address stylistic issues [7].

Corpus	Mean Scores
A	3.10
B	**3.35**
C	3.17

Table 1: Mean scores for comprehensibility for the three corpora

Corpus	Mean Scores
A	**3.35**
B	**3.87**
C	**3.61**

Table 2: Mean scores for style for the three corpora

Corpus	Mean Scores
A	2.63
B	**2.88**
C	2.82

Table 3: Mean scores for the severity of errors noted in the subtitles

Secondly we evaluate the findings from a qualitative perspective. The qualitative results show that in relation to the comprehensibility of the subtitles, all corpora received negative comments from the subjects. Some subjects, for example, said that they could understand more from using the (unknown) soundtrack rather than the subtitles. Therefore, the qualitative results alone do not identify the corpus that generated the most comprehensible subtitles. In order to validate the comprehensibility results, we investigated the inter-subject agreement for the three rating scales used in this study. The inter-subject agreement for Corpus A (.32) and Corpus C (.35) was *fair* and it was *substantial* for Corpus B (.69). This result, coupled with the fact that the Corpus B mean score for comprehensibility was the highest of the three, allows us to say that the subtitles generated by Corpus B are deemed more comprehensible than those generated by either Corpus A or Corpus C.

The qualitative findings for readability show that subjects judged Corpus C as having the lowest number of subtitles out of context. Combining this finding with the quantitative result, the readability of the subtitles generated using Corpus C is ranked higher than the other two corpora.

The qualitative findings for style supported the quantitative findings with the style of the subtitles generated by Corpus A deemed to be the least appropriate of the three corpora. Subjects viewing subtitles generated by Corpus A gave the most negative comments in relation to style, and recalled the most examples of inappropriate style. Subtitles generated by Corpus B and Corpus C were judged to be more appropriate in relation to style. In order to differentiate between the style of Corpus B and Corpus C subtitles, the mean scores were examined (see Table 2). Subjects ranked Corpus B subtitles higher than Corpus C subtitles. This result is further supported by the *substantial* inter-subject agreement for Corpus B.

When generating the qualitative data, subjects were asked two questions in relation to the well-formedness of the subtitles. Firstly they were asked whether the EBMT-generated subtitles were well-formed for use on DVD if the viewer did not understand the soundtrack, and secondly whether they could recall any well-translated subtitles. The responses were individually coded, and common themes emerged for each corpus, many of which overlapped between the corpora. These included comprehension issues and linguistics issues; and subjects commonly deemed subtitles acceptable or acceptable with improvements. The findings show that Corpus C received the highest number of negative responses in relation to well-formedness. The quantitative and qualitative findings for the well-formedness of the subtitles suggest that Corpus C subtitles are the least well-formed. Focusing on the findings from Corpus A and Corpus B, subjects' judgements on Corpus B subtitles were more positive overall; they ranked the observed errors in the Corpus B subtitles as less annoying than the errors in Corpus A subtitles, and subjects were able to recall a higher number of well-translated subtitles generated by Corpus B. Therefore Corpus B subtitles are deemed to be the most well-formed of the three corpora.

Combining the two sets of findings, we can say that the readability of machine-generated subtitles is improved if the size of the corpus is increased and with that the number of SL repetitions, and the corpus heterogeneity (Corpus C). However, the same cannot be said for comprehensibility and acceptability. Corpus B generated subtitles that are the most comprehensible, exhibit the most appropriate style and are the most well-formed. Therefore Corpus B subtitles are the most comprehensible and acceptable.

In addition to a quantitative and qualitative analysis of end-user judgements, we generated BLEU scores for each of the corpora and these are given in Table 4.

All 6 clips (178 subtitles)	Corpus A	Corpus B	Corpus C
	25.97	26.29	26.11

Table 4: BLEU scores for the EBMT output for each of the training corpora

Crime Series	Comedy Series	Car Documentary
63.9	54.4	53.6

Table 5: BLEU scores given by Volk [27] using an SMT system to generate subtitles

These scores correlate with our main human findings, albeit by a very small margin. The BLEU scores generated for the three corpora would normally be thought of as low scores (mid-twenties) and the EBMT output would possibly be discarded as 'bad quality' output. This is especially true if we compare the level of scores attained in this study with the kind of BLEU scores achieved by Volk [27], given in Table 5. It is clear, however, that these scores cannot be compared directly, given the differing subject matter and size of the training corpora. It might also be argued that the high BLEU scores obtained in [27] were partly due to the language combinations used. Swedish, Danish and Norwegian are all close members of the Indo-European Germanic language family, but some linguists even consider these three languages to be three dialects of the same language. Other research [19],[20] has reported high BLEU scores between languages within the same family. Koehn et al. [20] also comment on the fact that closely-related languages could share morphological forms, and that this is captured reasonably well in translation models. Despite the low scores presented in Table 4, an analysis of the qualitative data showed that even low scoring subtitles can be considered 'acceptable' in particular contexts by end-users. It should be noted, however, that the majority of the 'acceptable' subtitles were shorter subtitles (on average 2.23 words compared to the average subtitle length of 4.92 words for the entire movie). Another reason we cannot compare the quality of the EBMT subtitle output and the SMT subtitle output is that the 'human' element of the evaluations differed. The current study used end-users of the subtitles, while Volk used subtitlers working in the industry. These two groups have different priorities with regard to the subtitles, and therefore further studies would need to be carried out using similar groups of human evaluators.

6 Subtitling Corpora and Industry Collaboration

Both the current study and that described in [1] showed that in particular contexts, EBMT-generated subtitles are considered intelligible and acceptable for use on DVD. We should note that the subtitles which were evaluated in this study were raw EBMT output. This means that data based on raw EBMT output represents a baseline for human evaluations, and any subsequent studies using post-edited output should achieve better results [27],[28].

The reported success of an SMT system to translate subtitles is helped significantly by collaboration between researchers in academia and a professional subtitling company [15],[27],[28]. One obvious advantage of this collaboration is the availability of SL subtitles and their corresponding TL translations within the subtitling company, which can then be aligned to create large training corpora by the SMT developers. In addition, alignment strategies for subtitles have been a topic of investigation in recent times [21],[25],[26]. Results have shown that the accuracy of the alignment is improving and that subtitles prove to be suitable for segment alignment if the time-codes are taken into consideration.

7 Making EBMT Accessible

During our study we comment on the need for human evaluation of subtitles, given the text type and the differences between automated subtitle evaluation and the more 'standard' evaluation research. This coincides with the arguments from within the SMT community regarding the need for human evaluation combined with the use of automatic evaluation metrics [5]. However, we also discuss the difficulties associated with human evaluation, including subject recruitment, time and costs. As a result we suggest the use of online evaluation modules. These modules would use a refined version of the questionnaire used in our study, and each module could be customised to the particular research question. As a matter of fact, the use of Amazon's Mechanical Turk [4] could be explored as a method for end-user evaluation. This approach could tackle many of the issues associated with human MT evaluation. In addition to these suggestions a user interface for the MaTrEx system should be developed. Investigating real-world applications of MT systems has become an interdisciplinary research area, with humanities-based

researchers working closely with system developers. Introducing non-technical evaluation approaches would allow future interdisciplinary researchers (i.e. AVT and MT) to replicate evaluations such as those described above, and the accessibility to the EBMT system will support future empirical evaluations. It would also mean that researchers have control over their training and test data, have more flexibility to conduct the human evaluations, and they would develop a broader understanding of the EBMT process. In addition, providing increased accessibility to the EBMT system gives researchers with differing backgrounds (linguistic, computational, language learning) the opportunity to offer feedback on the system, and this makes collaborative improvements possible. Increased accessibility would establish EBMT research as a driving force behind real-world applications that incorporate MT technology.

Given the demonstrated success of integrating an SMT system into the subtitling process (from the perspective of the subtitler), and the end-user data gathered while evaluating the MaTrEx system (from the perspective of the DVD viewer), we should investigate the possibilities of introducing EBMT into the subtitling domain. To facilitate this integration, researchers in academia need to work in partnership with industry, in a similar fashion to the SMT system development. Firstly, this would give developers access to large amounts of data required to compile corpora, and the system could be built to suit the needs of the company; secondly, the system can be integrated into the dedicated subtitling software applications used by professional subtitlers and tested on site. The already reported success of SMT technology in this domain strengthens the motivation for further EBMT research in this area.

8 Conclusions

We describe interdisciplinary research that employs a hybrid EBMT/SMT system to translate English-German subtitles for use on DVD. The findings presented in this paper show that, on the one hand, the readability of machine-generated subtitles is improved if the size of the corpus is increased and with that the number of SL repetitions, and the corpus heterogeneity, in relation to the movie clips being subtitled (Corpus C). On the other hand, the comprehensibility and acceptability of the subtitles is not improved given the same conditions (with Corpus B proving most successful given these metrics). We suggest possible ways of reproducing these findings, and hence strengthening the validity of the empirical evaluation. We also discuss the apparent success of an SMT system to date, and argue that improving accessibility to EBMT technology could only have positive consequences, given the potential growth of using this kind of technology as a solution to the increased pressures within the AVT industry.

Acknowledgements

We would like to thank the anonymous reviewers whose valuable comments helped improve the quality of the paper. This research was funded by an IRCHSS[9] PhD scholarship.

References

[1] S. Armstrong, C. Caffrey, M. Flanagan, D. Kenny, M. O'Hagan and A. Way: Leading by Example: Automatic Translation of Subtitles via EBMT, *Perspectives*, 14(3):163-184, 2006.

[2] S. Armstrong: *Using EMBT to produce foreign language subtitles*, MSc Thesis, Dublin City University, 2007.

[3] D. Becquemont: Le sous-titrage cinématographique: contraintes, sens, servitudes, in Y. Gambier, ed., *Les transferts linguistiques dans les medias audiovisuals*, Villeneuve d'Ascq: Presses universitaires du Septentrion, p. 145–155, 1996.

[4] C. Callison-Burch: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk, in *Proceedings of the EMNLP* (Singapore, August 6-7), 2009.

[5] C. Callison-Burch, C. Fordyce, P. Koehn, C. Monz and J. Schroeder: (Meta-) evaluation of Machine Translation, in *Proceedings of the 45th Association for Computational Linguistics Workshop on Statistical Machine Translation,* (Prague, Czech Republic, 25-27, 2007), p. 136-158, 2007.

[9] www.irchss.ie

[6] M. Carroll: Subtitling: Changing standards for new media? LISA Newsletter Global Insider, XIII, 3.5, (2004), URL: http://www.lisa.org/globalizationinsider/2004/09/subtitling_chan.htm

[7] J. Díaz Cintas and A. Remael: *Audiovisual Translation: Subtitling*, Manchester, UK/Kinderhook, New York: St. Jerome Publishing, 2007.

[8] J. Du, Y. He, S. Penkale and A. Way: MaTrEx: the DCU MT System for WMT 2009, in *Proceedings of the 4th Workshop on Statistical Machine Translation,* (Athens, Greece, 30-31 March), (in press), 2009.

[9] M. Flanagan: Error classification for MT evaluation, in *Proceedings of the 1st Conference of the Association for Machine Translation in the Americas,* (Maryland, USA, 5-8 October, 1994), p. 65-72, 1994.

[10] M. R. Flanagan: *Recycling Texts: Human evaluation of Example-Based Machine Translation Subtitles for DVD*, PhD Thesis, Dublin City University, 2009.

[11] Y. Gambier: Recent developments and challenges in audiovisual translation research, in D. Chiaro, C. Heiss and C. Bucaria, eds., *Between Text and Image,* Amsterdam/Philadelphia: John Benjamins Publishing, p. 11-33, 2008.

[12] E. Gamma, R. Helm, R. Johnson and J. M. Vlissides: *Design Patterns: Elements of Reusable Object-Oriented Software*, Reading, MA: Addison-Wesley, 1995.

[13] T. Green: The necessity of syntax markers. Two experiments with artificial languages, *Journal of Verbal Learning and Behavior*, 18(4):481-496, 1979.

[14] D. Groves: *Hybrid Data-Driven Models of Machine Translation*, PhD Thesis, Dublin City University, 2007.

[15] C. Hardmeier and M. Volk: *Using Linguistic Annotations in Statistical Machine Translation of Film Subtitles*, in *Proceedings of Nodalida,* (Odense, Denmark, 14-16 May, 2009), no pagination, 2009.

[16] *Harry Potter and the Philosopher's Stone*: (2001). [DVD]. UK: 1492 Pictures.

[17] H. Hassan, Y. Ma and A. Way: MaTrEx: the DCU Machine Translation System for IWSLT 2007, in *Proceedings of the International Workshop on Spoken Language Translation*, (Trento, Italy, 15-16 October, 2007), p. 69-75, 2007.

[18] J. Ivarsson and M. Carroll: *Subtitling*. Simrishamn: TransEdit, 1998.

[19] P. Koehn: Europarl: A Parallel Corpus for Statistical Machine Translation, in *Proceedings of the 10th MT Summit*, (Phuket, Thailand, 13-15 September, 2005), p. 79-86, 2005.

[20] P. Koehn, A. Birch and R. Steinberger: 462 Machine Translation Systems for Europe, in *Proceedings of the 12th MT Summit*, (Ottawa, Canada, 26-30 August, 2009), p. 65-72.

[21] C. Lavecchia, K. Smaili and D. Langlois: Building parallel corpora from movies, in *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science,* (Funchal, Madeira, 12-13 June 2007), no pagination, 2007.

[22] Y. Ma, O. Tsuyoshi, Ç. Özlem, J. Du and A. Way: Low-Resource Machine Translation Using MaTrEx: the DCU MT System for IWSLT 2009, in *Proceedings of the IWSLT 2009 Workshop*, (Tokyo, Japan, December 1-2) (forthcoming), 2009.

[23] M. O'Hagan: Can Language Technology Respond to the Subtitler's Dilemma?, in *Proceedings of the 25th International Conference on Translating and the Computer*, (20-21 November, 2003). London: ASLIB. No pagination, 2003.

[24] N. Stroppa and A. Way: MaTrEx: DCU Machine Translation System for IWSLT 2006, in *Proceedings of the International Workshop on Spoken Language Translation*, (Kyoto, Japan, 27-28 November 2006), p. 31-36, 2006.

[25] J. Tiedemann: Improved sentence alignment for movie subtitles, in *Proceedings of Recent Advances in Natural Language Processing,* (Borovets, Bulgaria, 27-29 September, 2007), p. 582–588, 2007.

[26] J. Tiedemann: Synchronizing Translated Movie Subtitles, in *Proceedings of Language Resources and Evaluation Conference,* (Marrakesh, Morocco, 28-30 May, 2008). No pagination, 2008.

[27] M. Volk: The Automatic Translation of Film Subtitles. A Machine Translation Success Story?, in J. Nivre, M. Dahllöf, and B. Megyesi, eds., *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein.* Sweden: Uppsala University, p. 202-214, 2008.

[28] M. Volk and S. Harder: Evaluating MT with Translations or Translators. What is the Difference?, in *Proceedings of the 11th MT Summit*, (Copenhagen, Denmark, 10-14 September, 2007), p. 499-506, 2007.