Elhuyar submission to the Biomedical Translation Task 2020 on terminology and abstracts translation

Ander Corral
Elhuyar Foundation
a.corral@elhuyar.eus

Xabier Saralegi
Elhuyar Foundation
x.saralegi@elhuyar.eus

Abstract

This article describes the systems submitted by Elhuyar to the 2020 Biomedical Translation Shared Task, specifically the systems presented in the subtasks of terminology translation for English-Basque and abstract translation for English-Basque and English-Spanish. In all cases a Transformer architecture was chosen and we studied different strategies to combine open domain data with biomedical domain data for building the training corpora. For the English-Basque pair, given the scarcity of parallel corpora in the biomedical domain, we set out to create domain training data in a synthetic way. The systems presented in the terminology and abstract translation subtasks for the English-Basque language pair ranked first in their respective tasks among four participants, achieving 0.78 accuracy for terminology translation and a BLEU of 0.1279 for the translation of abstracts. In the abstract translation task for the English-Spanish pair our team ranked second (BLEU=0.4498) in the case of OK sentences.

1 Introduction

General purpose translation systems usually perform poorly where domain specific knowledge is required (Koehn and Knowles, 2017). Therefore, it is essential to develop translation systems for specific domains. However, high quality parallel corpora for domain specific tasks are only available for a few major languages and obtaining in domain translated data for the majority of language pairs becomes a challenge. With such scarcity of resources, various domain adaptation techniques have shown promising results (Currey et al. 2017, Saunders et al. 2019, Sennrich et al. 2017).

The biomedical domain is of great interest for the application of machine translation. It is a sector of great importance in society, even more so after the advent of COVID-19, where the handling of documentary information plays a major role. Therefore, it is a scenario where machine translation can be of great help in order to facilitate the flow of information between different languages.

We can find different works in the literature that address the task of developing NMT systems adapted to the biomedical domain (Yepes et al., 2017; Sennrich et al., 2017; Khan et al., 2018). In contrast, few papers focus on language pairs where training data is scarce. This is precisely the case of the task of translation from English to Basque. The work of Soto et al. (2019) is specially relevant in this case, which presents a clinical domain oriented system for Basque-Spanish based on RNN and Transformer architectures and which does not require bilingual domain texts but only bilingual clinical terminology (SNOMED CT).

Our participation in the Biomedical Translation Shared Task addresses the translation of biomedical terminology from English to Basque and the translation of biomedical abstracts from English to Basque and from English to Spanish. Given the scarcity of real parallel biomedical domain corpora, our approach focused on different strategies to combine open domain data with in-domain biomedical data. In the case of the English to Basque translations, back translation technique has been applied to generate synthetic examples from a real biomedical Spanish-Basque corpus and several monoligual in-domain Basque corpora have been merged to the training data by copying target side examples to source. Furthermore, we have considered finetuning previously available open domain systems in order to take advantage of the learnt general, open domain patterns.

This article is structured as follows: the following section reviews the most relevant related works. Section 3 describes the systems presented to the sub-tasks that include the English-Basque and English-Spanish pairs, as well as the results

obtained in the experimentation. Section 4 presents the results obtained in the official evaluation, and finally, we end the paper outlining the most relevant conclusions drawn from this work.

2   Related work

Most of the related work on domain adaptation focuses on using either in-domain monolingual corpora, synthetically generated corpora or small parallel corpora.

Regarding in-domain monoligual corpora, Currey et al. (2017) analyze the benefits of augmenting available data by copying the target side monolingual data to source and training the system. Results show significant gains in accuracy on named entities and words remaining identical in source and target languages.

Several studies show the effectiveness of generating synthetic parallel corpora for domain adaptation. Sennrich et al. (2015) use the back translation technique with target monolingual data to strengthen the decoder, Zhang and Zong (2016) make use of source side monolingual data and Park et al. (2017) use both, source and target monolingual data to improve in-domain translations.

Finally, when in-domain parallel corpora is available, previous work focuses on mixed domain NMT systems by using both in-domain and out of domain data. Chu et al. (2017) propose to use control tags to mark in-domain sentences prior to concatenating multiple domain corpora. Sajjad et al. (2017) compare different methods for training a multi domain system, such as, concatenation, interactively training on different domains, selecting out of domain data close to the in domain data and ensembling different domain models. Wang et al. (2017) exploit the internal sentence embeddings to find sentences that are close to in-domain data from out of domain data.

Several works in the literature address the task of developing NMT systems adapted to the biomedical domain. Saunders et al. (2019) apply transfer learning technique by training on a large, general domain corpus and finetuning a series of systems on different biomedical domains. They perform multi domain ensembling to further improve the results. Khan et al. (2018) iteratively apply transfer learning on various biomedical domains.

Regarding works dealing with Basque the work of Soto et al. (2019) presents a clinical domain oriented system for Basque-Spanish based on RNN

and Transformer architectures, which does not require bilingual domain texts but bilingual clinical terminology (SNOMED CT). They also analyse different back translation techniques. Aimed at translating clinical terminology into Basque, we find the work of Perez-de Viñaspre (2017) which proposes a system for translating SNOMED CT into Basque by combining lexical resources, transliteration of neoclassical terms, generation of nested terms and a domain-adapted RBMT system.

3   Experiments

In this section we describe the experiments carried out when training the translation systems. We compare the results obtained by the systems on different open domain and in-domain test sets prior to selecting the best runs to submit for the biomedical translation task. We also detail the Transformer architecture used and its parameters.

3.1   Datasets

For the experiments, we considered open domain general data and in-domain task specific data for fine-tuning purposes. Open domain data comprises the publicly available Paracrawl v5 (Esplà et al., 2019) corpus for English-Spanish and a Elhuyar's internal synthetic corpus for Basque-Spanish.

Due to the scarcity of in-domain data in the case of English-Basque, back translation has also been applied to generate synthetic examples. Furthermore, we augmented training data by using monolingual Basque data gathered from artificially generated hospital notes, SNOMED-CT terminological content and Wikipedia biomedical articles. We have created parallel data by copying the Basque sentences to source so that each source sentence is identical to the target sentence. Table 1 offers a summary of the corpora used.

3.2   Architecture

For training the models the Transformer architecture (Vaswani et al., 2017) has been chosen. Specifically, the Python implementation of the OpenNMT (Klein et al., 2017) library has been used. Transformers are based on an encoder-decoder system with an attention mechanism. Both the encoder and the decoder are composed of 6 layers composed in turn by a recurrent neural network and an attention mechanism. Default values of the architecture without any optimization of the parameters have been applied. The size of the recurrent neu-

Corpus	Pair	Domain	Description	Sentences
Elhuyar synthetic (ELH Syn)	EN-EU	Open	Elhuyar's internal synthetic corpus obtained by translating an internal ES-EU corpus with our best performing ES-EN out of domain system	6.9M
EHU books	EN-EU	Biomed.	Collection of biomedical books translated from English to Basque	22.6k
ICD-10	EN-EU	Biomed.	ICD-10 codes translated from English to Basque and publicly available for the shared task	25.9k
EHU Synthetic (EHU Syn)	EN-EU	Biomed.	Synthetically generated corpus by back translating an in domain EHU book collection dataset (ES-EU) from Basque to English with a previously trained in domain EU-EN system	303k
Hospital notes	EU	Biomed.	Artificially generated hospital notes to use as guide for practitioners	2.2k
SNOMED	EU	Biomed.	Automatic translation of the terminological content of SNOMED-CT (2020), no manually revised	105k
Wikipedia	EU	Biomed.	Medical domain articles from Wikipedia	1.3k
Paracrawl v5 (PCv5)	EN-ES	Open	Publicly available Paracrawl v5 corpus, which comprises parallel segments crawled from the web	33.3M
Biomedical (Mix)	EN-ES	Biomed.	Comprises subsets of the Scielo dataset (Soares et al., 2018), and previous years' WMT shared task datasets	560k

Table 1: Size of corpora used for training the systems.

ral network of each layer is 512. Thus, 512 size embeddings have been used for both source and target sentences. Adam optimizer has been used during the training, and a learning-rate of 2 with a warm-up phase of 8000 steps. The dropout ratio is 0.1 and the batch size is 4096 sentences. All models have been trained until the results on the development set stopped improving.

To avoid the open vocabulary issue and for a better translation of unknown words, BPE tokenization (Sennrich et al., 2016) has been applied to source and target sequences. Rare or unseen words are represented as a sequence of subword units. In the case of Basque, this encoding is particularly useful as declensions generate a larger vocabulary.

3.3 EN-EU experiments

In this section we provide a detailed description of the experiments carried out for the English-Basque pair. We describe the systems built and the results obtained on different test sets. Looking for a robust experimental setup, we conducted both open domain and in-domain evaluation. A brief description of the test sets can be found on Table 2.

3.3.1 Systems

For the English-Basque pair we trained the following systems:

Baseline1. A strong baseline by pivoting Elhuyar's best out of domain EN-ES and ES-EU models. These models are trained on the PCv5 corpus and a Elhuyar's internal Spanish-Basque corpus respectively.

Baseline2. A further improvement of Baseline1, by fine-tuning the EN-ES pivoting system with Medline and SCIELO in-domain biomedical data.

Baseline3. A previously available out of domain EN-EU system trained with back translated synthetic data.

Baseline4. A simple baseline trained with the task's official in-domain data (ICD-10 corpus).

SystemA1, SystemA2 and SystemA3. These systems are the result of fine-tuning Baseline3 with in-domain data. SystemA1 uses a subset (250k) of ELH Syn open domain corpus as well as shared task ICD-10 in domain data. A small portion of the ICD-10 corpus (1k) is used for validation. SystemA2 is a variant of SystemA1 by adding more in domain data from the EHU books corpus. A small subset of the EHU books corpus (1k) is also added to the validation set. Finally, SystemA3 includes a subset (5k) of the out of domain ELH Syn corpus in the validation set in order to prevent the system from forgetting about prior out of domain knowledge.

SystemA4. A variant of SystemA3 using all the available ELH Syn open domain data.

SystemB1. This system was trained by adding synthetically generated EHU Syn in-domain data to the data used in SystemA4. To create synthetic data, a fine-tuned EU-EN model has been used to back translate an internal ES-EU biomedical corpus gathered from a collection of EHU books.

SystemB2. This system was trained by further augmenting data from SystemB1 with copied monolingual Basque target data from the shared task (Hospital notes, SNOMED terminology and Wikipedia).

SystemC. A variant of SystemB2. Synthetic Medline data from the WMT19 EN-ES biomedical shared task was added to the validation set to improve the performance on the Medline domain.

Test	Pair	Domain	Description	Sentences
Synthetic (Syn)	EN-EU	Open	Subset of the Elhuyar's internal synthetic corpus	5k
EHU books (EHU)	EN-EU	Biomed.	Subset of the collection of biomedical books translated from English to Basque	1k
Medline pro (PRO)	EN-EU	Biomed.	Professional translation from Spanish to Basque of the WMT19 ES-EN shared task data	200
Terminology (ICD-10)	EN-EU	Biomed.	Subset of the shared task in-domain ICD-10 terminology set	368
Paracrawl v5 (PCv5)	EN-ES	Open	Subset of the Paracrawl v5 corpus	5k
Elhuyar TMs (ELH)	EU-ES	Open	Data collected from Elhuyar's internal translation memories	1k
WMT18	EU-ES	Biomed.	WMT18 biomedical task test set	277
WMT19	EU-ES	Biomed.	WMT19 biomedical task test set	368

Table 2: Description of the test sets used to evaluate the models.

System	Train data	Dev data	Open Syn	In-domain ICD-10	EHU	PRO
Baseline1	-	-	15.05	10.02	15.58	13.31
Baseline2	-	-	15.06	10.03	**16.62**	**13.37**
Baseline3	ELH Syn	ELH Syn	**15.28**	9.28	15.08	11.80
Baseline4	ICD-10	ICD-10	0.00	**89.18**	0.00	0.0
SystemA1	ELH Syn (250k); ICD-10	ICD-10	9.33	**90.46**	9.86	6.29
SystemA2	ELH Syn (250k); ICD-10; EHU books	ICD-10; EHU books	10.42	90.26	32.20	8.19
SystemA3	ELH Syn (250k); ICD-10; EHU books	ELH Syn; ICD-10; EHU books	12.72	87.60	13.76	9.38
SystemA4	ELH Syn (all); ICD-10; EHU books	ELH Syn; ICD-10; EHU books	15.47	80.36	24.43	12.95
SystemB1	ELH Syn (all); ICD-10; EHU books; EHU Syn	ELH Syn; ICD-10; EHU books	15.81	82.05	26.45	12.85
SystemB2	ELH Syn (all); ICD-10; EHU books, EHU Syn ; Monoligual	ELH Syn; ICD-10; EHU books	15.69	81.01	26.51	**13.61**
SystemC	ELH Syn (all); ICD-10; EHU books; EHU Syn; Monoligual	ELH Syn; ICD-10; EHU books; Medline Syn 19	**15.91**	83.79	**27.73**	13.50

Table 3: BLEU scores for the English to Basque experiments on out of domain and in-domain test sets.

3.3.2 Results

Table 3 shows BLEU scores for the English to Basque experiments on out of domain and in-domain test sets (Table 2).

As for abstract translation, SystemA4, SystemB1, SystemB2 and SystemC showed a significant improvement on those test sets when compared to the other trained systems. In particular, SystemB2 obtained the best results on PRO test and SystemC the second best result.

The gap between SystemA4 and the other three SystemA's showed that using all the available out of domain data helps avoiding the "catastrophic forgetting" phenomena, where all previous knowledge fades when learning new in-domain examples.

SystemB1 introduces in-domain synthetic data in the training process, which significantly improves the results on the EHU test. This is due to the fact that synthetic data and the EHU test set share the same domain (EHU biomedical books). However, the drop on the PRO test is almost insignificant and it also improves all the baselines on the Synthetic test.

SystemB2 further improves the results on the PRO test set by adding monolingual corpora to the training data and SystemC shows the effect of the validation set by adding more biomedical domain data to the validation set. Both systems improve all the baselines on every test set.

SystemB2 (run1), SystemC (run2) and Baseline2 (run3) were submitted as the best runs for the English to Basque abstract translation task.

Terminology translation task greatly differs from the abstract domain which can be clearly seen in the results. In this case, the best results are obtained by SystemA1 which only includes a small part of the available out of domain data. Furthermore, results are not distant from Baseline4 which was trained with ICD-10 training data. This behaviour shows the specificity of the task where previous complete sentence translation knowledge is not essential.

SystemA2 (run1), SystemA1 (run2) and Baseline4 (run3) were submitted as the best runs for the English to Basque terminology translation task.

3.4 EN-ES experiments

Below we present the different systems trained for the English-Spanish pair and the results obtained

for each of them.

3.4.1 Systems

For the English-Spanish pair we trained the following systems:

Baseline1. We have considered a strong baseline by using Elhuyar's best out of domain EN-ES model. This model was trained on the PCv5 corpus.

SystemA. This model has been trained from scratch mixing in-domain and out domain data. In-domain data comprises previous years' shared tasks Medline data and a subset of the SCIELO dataset. Out of domain data was obtained from the PCv5 corpus.

SystemA'. A variant of SystemA by averaging the three best performing checkpoints.

3.4.2 Results

Table 4 shows BLEU scores for the English to Spanish experiments on out of domain and in-domain test sets (Table 2).

SystemA improves the baseline on all the test sets and SystemA' further improves those results obtaining the best results for the task. Similar to the English to Basque abstract translation task, for the English to Spanish pair adding in-domain data and fine-tuning a previously trained out of domain system improved the results. Furthermore, averaging the first three best checkpoints helped improving the results of the best checkpoint.

SystemA' (run1), SystemA (run2) and Baseline1 (run3) were submitted as the best runs for the English to Spanish abstract translation task.

4 Official results

For the English to Basque abstract translation task we selected the PRO test as the most representative one when choosing the best runs for submission. We assumed that this test set would be the closest to the official task test. The EHU test set was also a great indicator of how robust our system was for the biomedical domain.

BLEU scores were calculated using the multi-eval tool and tokenization as provided in Moses. Table 5 shows the performance of all the submitted runs for the official abstract translation test set. Our submitted run2 has obtained the best score on the official task test, achieving a 0.1279 BLEU score. When compared to other teams, all our submitted runs significantly outperform all the runs. It is worth mentioning that our Baseline2 (run3) which

is based on pivoting between two systems (EN-ES and ES-EU) has obtained really close results which indicates that some biomedical knowledge was present in the Spanish to Basque system.

In the case of the terminology task, ICD-10 test set is the official validation task and therefore the most representative one for choosing the best runs.

For the evaluation of terminology we provide two metrics: (i) accuracy, by relying on strict matches (case insensitive) between ground truth ans predictions; and (ii) BLEU score, as measured by the NLTK module sentence_bleu. Table 6 shows the performance of all the submitted runs for the official terminology test set.

Our best submitted run has obtained the best score on the official task test, achieving 0.78 accuracy and a 0.7373 BLEU score. When compared to other teams, all our submitted runs outperform all the runs, except for DCU_MT's run2. It is worth mentioning that out Baseline4 (run3) which was trained on task's English-Basque data has outperformed almost all of the others teams' results. This highlights the improvements obtained by our systems over the baseline.

Finally, for the English to Spanish abstract translation task WMT19 and WMT18 test sets were selected as reference for selecting the best runs.

Table 7 shows the performance of all the submitted runs for the official English to Spanish abstract translation test set. Our best submitted run has obtained the fifth best score on all sentences achieving a 0.4364 BLEU score and the second best team with OK sentences (BLEU=0.4498). In this case, our submitted baseline (run3) also shows a great robustness, indicating some prior biomedical domain knowledge.

5 Conclusions

For the Biomedical Translation Task 2020, we considered several strategies combining open domain and in-domain biomedical data. We have successfully applied transfer learning by fine-tuning a previously available open domain system with in-domain specific data. To tackle the scarcity of English-Basque domain data, we have performed data augmentation by back translating real data.

The systems submitted for the terminology and abstract translation tasks for the English-Basque pair have ranked first on the official task test, achieving 0.78 accuracy for terminology translation and a BLEU of 0.1279 for the translation of

System	Train data	Dev data	Open		In-domain	
			PCv5	ELH	wmt19	wmt18
Baseline1	PCv5	PCv5	**44.69**	35.68	44.27	30.73
SystemA	PCv5 and Biomedical Mix	PCv5 and Biomedical Mix	46.37	36.25	45.72	30.99
SystemA'	PCv5 and Biomedical Mix	PCv5 and Biomedical Mix	**46.54**	**36.29**	**45.74**	**31.22**

Table 4: BLEU scores for the English to Spanish experiments on out of domain and in-domain test sets.

Team	Runs	BLEU
Elhuyar_NLP	run1	0.1271
Elhuyar_NLP	run2	(1) **0.1279**
Elhuyar_NLP	run3	0.1268
DCU_MT	run1	0.0867
DCU_MT	run2	0.0825
DCU_MT	run3	0.0808
UTS_NLP	run1	0.0530
UTS_NLP	run2	0.0549
UTS_NLP	run3	0.0528
Ixamed	run1	0.0815
Ixamed	run2	0.0782
Ixamed	run3	0.0884
Baseline	-	0.0596

Table 5: Performance scores on the official English to Basque abstract translation test set.

Team	Runs	Accuracy	BLEU
Elhuyar_NLP	run1	(1) **0.78**	(1) **0.7373**
Elhuyar_NLP	run2	0.77	0.7356
Elhuyar_NLP	run3	0.75	0.7229
DCU_MT	run1	0.73	0.7083
DCU_MT	run2	0.76	0.7239
DCU_MT	run3	0.75	0.7179
UTS_NLP	run1	0.73	0.7115
UTS_NLP	run2	0.73	0.7122
UTS_NLP	run3	0.73	0.7085
Ixamed	run1	0.12	0.1314
Ixamed	run2	0.08	0.0721
Ixamed	run3	0.13	0.1481

Table 6: Performance scores on the official terminology test set.

Team	Runs	BLEU	BLEU OK
Elhuyar_NLP	run1	(5) 0.4364	(4) 0.4498
Elhuyar_NLP	run2	0.4359	0.4493
Elhuyar_NLP	run3	0.4263	0.4394
Ixamed	run1	0.4052	0.4171
Ixamed	run2	0.3729	0.3836
Ixamed	run3	0.3755	0.3858
Sheffield	run1	**0.4493**	0.4493
TRAMECAT	run1	0.4238	0.4361
UNICAM	run1	0.4434	0.4572
UNICAM	run2	0.4464	**0.4672**
UNICAM	run3	0.4453	0.4662
Baseline	-	0.3709	0.3813

Table 7: Performance scores on the official English to Spanish test set.

of the pivoting pairs (EN-ES) we have created a robust baseline for the biomedical domain.

Adding monolingual corpora to the training data, as copied target, seems to improve the decoder by adapting the systems to better perform on domain specific terminology. Even though some noise is introduced by copying the target to the source side, the result are improved.

Terminology task showed promising results when translating biomedical domain terms, which could led to a production ready system.

abstracts. For the English to Spanish abstract translation task our systems have obtained competitive enough results, being the second team for OK sentences (BLEU=0.4498).

In all cases, even developed baselines have achieved outstanding results. For the English-Spanish task, Paracrawl v5 has proven to be a robust baseline for biomedical domain systems, as it seems to contain some biomedical crawled websites. For the English-Basque task, fine-tuning one

References

Chenhui Chu, Raj Dabre, and Sadao Kurohashi. 2017. An empirical comparison of domain adaptation methods for neural machine translation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 385–391.

Anna Currey, Antonio Valerio Miceli-Barone, and Kenneth Heafield. 2017. Copied monolingual data improves low-resource neural machine translation. In Proceedings of the Second Conference on Machine Translation, pages 148–156.

Miquel Esplà, Mikel Forcada, Gema Ramírez-Sánchez, and Hieu Hoang. 2019. ParaCrawl: Web-scale parallel corpora for the languages of the EU. In Proceedings of Machine Translation Summit XVII

Volume 2: Translator, Project and User Tracks, pages 118–119, Dublin, Ireland. European Association for Machine Translation.

Abdul Khan, Subhadarshi Panda, Jia Xu, and Lampros Flokas. 2018. Hunter nmt system for wmt18 biomedical translation task: Transfer learning in neural machine translation. In Proceedings of the Third Conference on Machine Translation: Shared Task Papers, pages 655–661.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. arXiv preprint arXiv:1701.02810.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. arXiv preprint arXiv:1706.03872.

Jaehong Park, Jongyoon Song, and Sungroh Yoon. 2017. Building a neural machine translation system using only synthetic parallel data. arXiv preprint arXiv:1704.00253.

Hassan Sajjad, Nadir Durrani, Fahim Dalvi, Yonatan Belinkov, and Stephan Vogel. 2017. Neural machine translation training in a multi-domain scenario. arXiv preprint arXiv:1708.08712.

Danielle Saunders, Felix Stahlberg, and Bill Byrne. 2019. Ucam biomedical translation at wmt19: Transfer learning multi-domain ensembles. arXiv preprint arXiv:1906.05786.

Rico Sennrich, Alexandra Birch, Anna Currey, Ulrich Germann, Barry Haddow, Kenneth Heafield, Antonio Valerio Miceli Barone, and Philip Williams. 2017. The university of edinburgh's neural mt systems for wmt17. arXiv preprint arXiv:1708.00726.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. Improving neural machine translation models with monolingual data. arXiv preprint arXiv:1511.06709.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Xabier Soto, Olatz Perez-de Viñaspre, Maite Oronoz, and Gorka Labaka. 2019. Leveraging snomed ct terms and relations for machine translation of clinical texts from basque to spanish. In Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation, pages 8–18.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008.

Olatz Perez-de Viñaspre. 2017. Automatic medical term generation for a low-resource language: translation of SNOMED CT into Basque. Ph.D. thesis, PhD thesis, University of the Basque Country, Donostia, Euskal Herria.

Rui Wang, Andrew Finch, Masao Utiyama, and Eiichiro Sumita. 2017. Sentence embedding for neural machine translation domain adaptation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 560–566.

Antonio Jimeno Yepes, Aurélie Névéol, Mariana Neves, Karin Verspoor, Ondřej Bojar, Arthur Boyer, Cristian Grozea, Barry Haddow, Madeleine Kittner, Yvonne Lichtblau, et al. 2017. Findings of the wmt 2017 biomedical translation shared task. In Proceedings of the Second Conference on Machine Translation, pages 234–247.

Jiajun Zhang and Chengqing Zong. 2016. Exploiting source-side monolingual data in neural machine translation. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1535–1545.