

Latin-Spanish Neural Machine Translation: from the Bible to Saint Augustine

Eva Martínez García, Álvaro García Tejedor

CEIEC-Universidad Francisco de Vitoria

Edificio H, Campus de Pozuelo, Universidad Francisco de Vitoria.

Carretera Pozuelo-Majadahonda km. 1,800. 28223 Pozuelo de Alarcón (Madrid).

{eva.martinez, a.gtejedor}@ceiec.es

Abstract

Although there are several sources where to find historical texts, they usually are available in the original language that makes them generally inaccessible. This paper presents the development of state-of-the-art Neural Machine Systems for the low-resourced Latin-Spanish language pair. First, we build a Transformer-based Machine Translation system on the Bible parallel corpus. Then, we build a comparable corpus from Saint Augustine texts and their translations. We use this corpus to study the domain adaptation case from the Bible texts to Saint Augustine’s works.

Results show the difficulties of handling a low-resourced language as Latin. First, we noticed the importance of having enough data, since the systems do not achieve high BLEU scores. Regarding domain adaptation, results show how using in-domain data helps systems to achieve a better quality translation. Also, we observed that it is needed a higher amount of data to perform an effective vocabulary extension that includes in-domain vocabulary.

Keywords: Machine Translation, domain adaptation, low resourced

1. Introduction

There exist several digital libraries that store large collection of digitalized historical documents. However, most of these documents are usually written in Latin, Greek or other ancient languages, resulting in them being inaccessible to general public. Natural Language Processing (NLP) offers different tools that can help to save this language barrier to bring the content of these historical documents to people. In particular, Machine Translation (MT) approaches can reproduce these historical documents in modern languages.

We present a set of experiments in machine translation for the Latin-Spanish language pair. We build a baseline Transformer-based (Vaswani et al., 2017) system trained on the Bible parallel corpus (Christodoulopoulos and Steedman, 2015) to study the associated difficulties of handling morphologically rich low-resourced languages like Latin. Latin is a low-resourced language, with few publicly available parallel data (González-Rubio et al., 2010a; Resnik et al., 1999). This is a challenge for data-driven approaches in general, and state-of-the-art Neural Machine Translation (NMT) approaches in particular since these systems usually require a high amount of data (Zoph et al., 2016). We create a comparable corpus from Saint Augustine’s works and we study the impact of adapting the baseline Bible translation system towards the Saint Augustine writings.

The paper is organized as follows. In Section 2., we revisit the state-of-the-art MT approaches and their application to Latin. Then, in Section 3. we describe both the parallel and the comparable data that we use in our experiments, explaining how we compiled the comparable corpus. Section 4. gives details on the set of experiments that we carried out to evaluate a baseline NMT trained on the Bible and its adaptation towards the Saint Augustine work. Finally, Section 5. discusses the conclusions and future work.

2. Related Work

There is a growing interest in the computational linguistic analysis of historical texts (Bouma and Adesam, 2017; Tjong Kim Sang et al., 2017). However, there are only a few works related to MT for ancient or historical languages. In (Schneider et al., 2017), the authors treat the spelling normalization as a translation task and use a Statistical Machine Translation (SMT) system trained on sequences of characters instead of word sequences. There exist shared tasks like the CLIN27 (Tjong Kim Sang et al., 2017), a translation shared task for medieval Dutch.

In the particular case of Latin, there exist several NLP tools, for instance, the LEMLAT morphological analyzer for Latin (Passarotti et al., 2017). However, there are only a few works involving MT for Latin. In particular, (González-Rubio et al., 2010b) describe the development of a Latin-Catalan Statistical Machine Translation System and the collection of a Latin-Catalan parallel corpus. However, to the best of our knowledge, the present work describes the first experiments in neural machine translation for the Latin-Spanish language pair.

Neural Machine Translation systems represent the current state-of-the-art for machine translation technologies and even some evaluations claim that they have reached human performance (Hassan et al., 2018). The first successful NMT systems were attentional encoder-decoder approaches based on recurrent neural networks (Bahdanau et al., 2015), but the current NMT state-of-the-art architecture is the Transformer (Vaswani et al., 2017). This sequence-to-sequence neural model is based solely on attention mechanisms, without any recurrence nor convolution. Although RNN-based architectures can be more robust in low-resourced scenarios, Transformer-based models usually perform better according to automatic evaluation metrics (Rikters et al., 2018). All the NMT systems built for our experiments follow the Transformer architecture.

Latin and Spanish can be considered closely-related languages. There are several works that study the benefits of using NMT systems in contrast to using Phrase-Based Statistical MT (PBSMT) systems (Costa-jussà, 2017), observing how NMT systems are better for in-domain translations. (Alvarez et al., 2019) pursue a similar study from the post-editing point of view, showing how NMT systems solve typical problems of PBSMT systems achieving better results.

3. Corpora

In this section, we describe the parallel and comparable data we use to train our NMT models.

3.1. Parallel Data

Latin is a low-resourced language in general, and parallel data for Latin-Spanish are scarce in particular. In the

Corpus	Description	sent. align.
Tatoeba	A collection of translated sentences from Tatoeba ¹	3.9k
Bible	A multilingual parallel corpus created from translations of the Bible	30.3k
wikimedia	Wikipedia translations published by the wikimedia foundation and their article translation system.	0.1k
GNOME	A parallel corpus of GNOME localization files.	0.9k
QED	Open multilingual collection of subtitles for educational videos and lectures collaboratively transcribed and translated over the AMARA ² web-based platform.	6.1k
Ubuntu	A parallel corpus of Ubuntu localization files.	0.6k
	Total:	41.8k

Table 1: Description of Latin-Spanish corpora available in the OPUS repository. The *sent. align.* column shows the number of aligned sentences available per corpus.

OPUS (Tiedemann, 2012) repository there are only 6 Latin-Spanish parallel corpora of different domains. Table 1 shows the statistics of these corpora, with a total of only 41.8k aligned sentences available. For our work, we choose the Bible corpus (Christodouloupoulos and Steedman, 2015) since it is the largest corpus and the only one containing historical texts which are closer to the Saint Augustine texts domain.

3.2. Comparable Data

NMT systems usually need a considerable amount of data to achieve good quality translations (Zoph et al., 2016). We built a comparable Latin-Spanish corpus by collecting several texts from Saint Augustine of Hippo, one of the

most prolific Latin authors. The Federación Agustiniana Española (FAE) promoted the translation into Spanish of the Saint Augustine works and make them available online. We used most of the texts from the Biblioteca de Autores Cristianos (BAC), published under the auspices of the FAE, one of the most complete collections of the Augustinian works in Spanish^{3 4}.

After gathering the texts in Spanish and Latin, we processed the corpus. First, we split the text into sentences using the Moses (Koehn et al., 2007) sentence splitter and we tokenize the text using the Moses tokenizer. Then, we use Hunalign (Varga et al., 2007) to automatically align the data sentence by sentence. We filter out those sentence alignments that have assigned an alignment score below 0. Notice that since we are using automatically aligned data, the resulting corpus is comparable and not a parallel one.

Corpus	#sents	#tokens la	#tokens es
Train	91,044	2,197,422	2,834,749
Development	1,000	22,914	28,812
Test	1,500	31,682	40,587
Total:	93,544	2,252,018	2,904,148

Table 2: Figures for the comparable corpus on Saint Augustine works, showing the number of aligned sentences (*#sents*) and the number of tokens in Latin (*#tokens la*) and in Spanish (*#tokens es*). *Train*, *Development* and *Test* represent the slices used for building the MT systems. *Total* shows the total amount of data.

4. Experiments

We want to study, first, the applicability of the state-of-the-art NMT systems to the Latin-Spanish language pair. Once we have created the comparable corpus on the Saint Augustine writings, we analyze the impact of applying several domain-adaptation techniques to adapt our models from the Bible domain to the Saint Augustine domain.

4.1. Settings

Our NMT systems follow the Transformer architecture (Vaswani et al., 2017) and they are built using the OpenNMT-tf toolkit (Klein et al., 2018; Klein et al., 2017). In particular, we use the Transformer *small* configuration described in (Vaswani et al., 2017), mostly using the available OpenNMT-tf default settings: 6 layers of 2,048 inner-units with 8 attention heads. Word embeddings are set to 512 dimensions both for source and target vocabularies. Adam (Kingma and Ba, 2015) optimizer was used for training, using Noam learning rate decay and 4,000 warmup steps. We followed an early-stopping strategy to stop the training process when the BLEU (Papineni et al., 2002) on the development set did not improve more than 0.01 in the last 10 evaluations, evaluating the model each 500 steps.

³Saint Augustine texts are available in <https://www.augustinus.it>

⁴We use all the texts except the *Tractates on the Gospel of John* and Sermons from *Sermon 100th* onward.

Training data was distributed on batches of 3,072 tokens and we used a 0.1 dropout probability. Finally, a maximum sentence length of 100 tokens is used for both source and target sides and the vocabulary size is 30,000 for both target and source languages. Vocabularies are set at the subword level to overcome the vocabulary limitation. We segmented the data using *Sentencepiece* (Kudo and Richardson, 2018) trained jointly on the source and target training data used for building each model, following the unigram language model (Kud, 2018). The *Sentencepiece* models were trained to produce a final vocabulary size of 30,000 subword units.

We evaluate the quality of the outputs by calculating BLEU, TER (Snover et al., 2006) and METEOR (Denkowski and Lavie, 2011) metrics. We used *multeval* (Clark et al., 2011) to compute these scores on the truecased and tokenized evaluation sets.

4.2. Results

First, we trained a baseline model on the Bible parallel corpus. Table 3 shows the results of the automatic evaluation of this system in its in-domain development and test sets. The *checkpoint-30000* is the model that achieved the best BLEU score on the development data. Following a usual technique to improve the translation quality, we averaged the 8 checkpoints with the best BLEU on the development set resulting in the *avg-8* model. In this particular case, the average model is able to improve +0.47 on the development set and +0.78 on the test set with respect to the *ckpt-30000* model. Also, the *avg-8* system improves the TER metric both on the development and the test set by 1.4 and 1.5 points respectively.

Bible models	dev		test	
	BLEU \uparrow	TER \downarrow	BLEU \uparrow	TER \downarrow
ckpt-30000	11.6	76.8	9.7	82.3
avg-8	12.2	75.4	10.5	80.8

Table 3: Automatic evaluation of the Bible NMT models on the development (*dev*) and test sets extracted from the Bible corpus. *ckpt-30000* is the model resulting from the training step 30000, and the *avg-8* is the average of 8 checkpoints.

We selected the *avg-8* for adapting it to the Saint Augustine text via fine-tuning (Crego et al., 2016; Freitag and Al-Onaizan, 2016), that is, by further training the *avg-8* on the in-domain data (hereafter the *Bible* model). We created two systems adapted by fine-tuning, the first one uses the Bible vocabulary (*Bible-ft*), and the second one updates the Bible vocabulary by adding those missing elements from the Saint Augustine texts vocabulary (*Bible-ft-vocabExt.*). Furthermore, we also built a model trained only using the comparable corpus (*SAugustine*) and a model trained on the concatenation of the data from the Bible and the Saint Augustine comparable data (*Bible+SAugustine*)⁵. For all the systems, we selected those models that achieved the best BLEU scores on the development sets, considering also the models resulting from averaging 8 checkpoints with higher

BLEU scores on the development set like we did for the Bible model.

System	BLEU \uparrow	METEOR \uparrow	TER \downarrow
Bible	0.9	6.9	106.1
Bible-ft	9.4	25.3	79.2
Bible-ft-vocabExt.	7.1	21.9	84.4
SAugustine	9.1	25.2	79.7
Bible+SAugustine	10.1	26.6	78.5

Table 4: Automatic evaluation of the different MT systems on the in-domain manually validated Saint Augustine test set.

Table 4 shows the results of the automatic evaluation of the different systems on the *ValTest* from the Saint Augustine texts.

The best system is *Bible+SAugustine*, the one trained on the concatenated data, improving +0.7 points on BLEU regarding the best-adapted model *Bible-ft*. Also, it outperforms the model trained only on the in-domain data. These results show the importance of having enough data to train an NMT system as well as having an important percentage of data from the working domain.

The impact of using in-domain data to tune or train the translation models is remarkable. All the fine-tuned models outperform significantly the *Bible* model performance, gaining up to 8.5 points of BLEU. Notice that the fine-tuned model (*Bible-ft*) uses the same vocabulary as the *Bible* model. These numbers support the importance of having in-domain data for developing MT systems. Since many of the Saint Augustine writings discuss texts from the Bible, these results also evidence the sensitivity of MT systems to capture characteristics from different writing styles. These features can come from different authors or different time periods, which can be very important when studying historical texts, giving a wider sense to the domain definition.

Extending the vocabulary when fine-tuning the *Bible* model does not result in improvements regarding any of the automatic metrics. In fact, the *Bible-ft-vocabExt.* model is 2.3 BLEU points below the *Bible-ft* model. Although the model with the extended vocabulary can have wider coverage, it does not have enough data to learn a good representation for the new elements in the vocabulary.

We observe also that the *SAugustine* model obtains better scores than the *Bible* model since its training data is larger and belongs to the test domain, although it was trained on comparable data. However, the results of the adapted model *Bible-ft* are slightly better than the *SAugustine*. This evidences the importance of having data of quality to model the translation from Latin to Spanish.

5. Conclusions and Future Work

We built NMT systems for translating from Latin to Spanish. We identified the typical issues for low-resourced languages for the particular case of Latin-Spanish. Since we only found few parallel corpora available for this particular language pair, we collected the work of Saint Augustine

⁵The concatenated corpus resulted in 119,330 sentence pairs.

of Hippo in Spanish and Latin and built a comparable corpus of 93,544 aligned sentences. Furthermore, we created a manually validated test set to better evaluate the translation quality of our systems.

We built 5 NMT models trained on different data. First, we built a baseline system trained on the Bible parallel corpus. Then, we adapted the *Bible* model towards the Saint Augustine domain by fine-tuning it in two ways: maintaining the Bible vocabulary and extending this vocabulary by including new elements from the Saint Augustine data. Finally, we trained two models using directly the in-domain data. We built a model trained only on the comparable Saint Augustine corpus and, finally, we trained an NMT on the concatenation of the Bible and the Saint Augustine writings corpora. The automatic evaluation results show significant differences among the *Bible* model and the rest of the models that somehow include information from the in-domain data when translating the manually validated Saint Augustine test set, showing the importance of the in-domain data. The best system was the one trained on the concatenated data *Bible+SAugustine*, showing the importance of having enough data to train an NMT model.

As future work, we want to study the behavior of training NMT systems in the other direction: from Spanish to Latin. We find interesting to analyze if the issues observed when trying to translate into other morphologically rich languages like Basque (Etchegoyhen et al., 2018) or Turkish (Ataman et al., 2020) can be observed when dealing with Latin. In this line, we want to study the impact of using morphologically motivated subword tokenization like the ones proposed by (Alegria et al., 1996) for Basque and by (Ataman et al., 2020; Ataman et al., 2017) for Turkish. Also, we want to include a more in depth analysis of the linguistic related issues that can appear for these closely-related languages (Popović et al., 2016).

In order to deal with the low resource feature of the Latin-Spanish language pair, we want to continue with our work by applying data augmentation techniques like back-translation (Sennrich et al., 2016) to artificially extend the training data. The Latin-Spanish scenario seems to apply the unsupervised NMT approaches (Artetxe et al., 2018; Artetxe et al., 2019; Lample et al., 2018), since there are available resources in both languages but only a few parallel data. Also, we want to explore how a Latin-Spanish MT system can benefit from other languages in a multilingual scenario (Johnson et al., 2017; Lakew et al., 2018), i.e. romance languages, to improve the final translation quality.

6. Acknowledgements

We would like to thank Beatriz Magán and Miguel Pajares for their assistance during the development of this research work. We would also like to thank the Federación Agustiniiana Española (FAE) and the Biblioteca de Autores Cristianos (BAC) for making available online the Spanish translations of the Saint Augustine’s works.

7. Bibliographical References

Alegria, I., Artola, X., Sarasola, K., and Urkia, M. (1996). Automatic morphological analysis of basque. *Literary and Linguistic Computing*, 11(4):193–203.

- Alvarez, S., Oliver, A., and Badia, T. (2019). Does NMT make a difference when post-editing closely related languages? the case of Spanish-Catalan. In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, pages 49–56, Dublin, Ireland, August. European Association for Machine Translation.
- Artetxe, M., Labaka, G., Agirre, E., and Cho, K. (2018). Unsupervised neural machine translation. *Proceedings of the ICLR2018*.
- Artetxe, M., Labaka, G., and Agirre, E. (2019). An effective approach to unsupervised machine translation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.
- Ataman, D., Negri, M., Turchi, M., and Federico, M. (2017). Linguistically motivated vocabulary reduction for neural machine translation from turkish to english. *The Prague Bulletin of Mathematical Linguistics*, 108(1):331–342.
- Ataman, D., Aziz, W., and Birch, A. (2020). A latent morphology model for open-vocabulary neural machine translation. *Proceedings of the ICLR2020*.
- Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Gerlof Bouma et al., editors. (2017). *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, Gothenburg, May. Linköping University Electronic Press.
- Christodoulopoulos, C. and Steedman, M. (2015). A massively parallel corpus: the bible in 100 languages. In *Language Resources and Evaluation*.
- Clark, J. H., Dyer, C., Lavie, A., and Smith, N. A. (2011). Better hypothesis testing for statistical machine translation: Controlling for optimizer instability. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 176–181. Association for Computational Linguistics.
- Costa-jussà, M. R. (2017). Why Catalan-Spanish neural machine translation? analysis, comparison and combination with standard rule and phrase-based technologies. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 55–62, Valencia, Spain, April. Association for Computational Linguistics.
- Crego, J., Kim, J., Klein, G., Rebollo, A., Yang, K., Senelart, J., Akhanov, E., Brunelle, P., Coquard, A., Deng, Y., et al. (2016). Systran’s pure neural machine translation systems. *arXiv preprint arXiv:1610.05540*.
- Denkowski, M. and Lavie, A. (2011). Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems. In *Proceedings of the sixth workshop on statistical machine translation*, pages 85–91. Association for Computational Linguistics.
- Etchegoyhen, T., Martínez Garcia, E., Azpeitia, A., Labaka, G., Alegria, I., Cortes Etxabe, I., Jauregi Carrera, A., Ellakuria Santos, I., Martin, M., and Calonge, E. (2018). Neural machine translation of basque.

- Freitag, M. and Al-Onaizan, Y. (2016). Fast domain adaptation for neural machine translation. *arXiv preprint arXiv:1612.06897*.
- González-Rubio, J., Civera, J., Juan, A., and Casacuberta, F. (2010a). Saturnalia: A latin-catalan parallel corpus for statistical mt. In *LREC*.
- González-Rubio, J., Civera, J., Juan, A., and Casacuberta, F. (2010b). Saturnalia: A latin-catalan parallel corpus for statistical mt. In *LREC*.
- Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M., Liu, S., Liu, T., Luo, R., Menezes, A., Qin, T., Seide, F., Tan, X., Tian, F., Wu, L., Wu, S., Xia, Y., Zhang, D., Zhang, Z., and Zhou, M. (2018). Achieving human parity on automatic Chinese to English news translation. *CoRR*, abs/1803.05567.
- Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J. (2017). Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.
- Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the ICLR2015*.
- Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. (2017). OpenNMT: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada, July. Association for Computational Linguistics.
- Klein, G., Kim, Y., Deng, Y., Nguyen, V., Senellart, J., and Rush, A. (2018). OpenNMT: Neural machine translation toolkit. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*, pages 177–184, Boston, MA, March. Association for Machine Translation in the Americas.
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics on Interactive Poster and Demonstration Sessions (ACL)*, pages 177–180.
- (2018). Subword regularization: Improving neural network translation models with multiple subword candidates. In *Kudo, Taku*.
- Kudo, T. and Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 66–71.
- Lakew, S. M., Federico, M., Negri, M., and Turchi, M. (2018). Multilingual neural machine translation for zero-resource languages. *Italian Journal of Computational Linguistics. Volume 1, Number 1*.
- Lample, G., Ott, M., Conneau, A., Denoyer, L., and Ranzato, M. (2018). Phrase-based & neural unsupervised machine translation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 311–318.
- Passarotti, M., Budassi, M., Litta, E., and Ruffolo, P. (2017). The lemlat 3.0 package for morphological analysis of Latin. In *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, pages 24–31, Gothenburg, May. Linköping University Electronic Press.
- Popović, M., Arčan, M., and Klubička, F. (2016). Language related issues for machine translation between closely related south Slavic languages. In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*, pages 43–52, Osaka, Japan, December. The COLING 2016 Organizing Committee.
- Resnik, P., Olsen, M. B., and Diab, M. (1999). The bible as a parallel corpus: Annotating the ‘book of 2000 tongues’. *Computers and the Humanities*, 33(1-2):129–153.
- Rikters, M., Pinnis, M., and Krišlauks, R. (2018). Training and adapting multilingual nmt for less-resourced and morphologically rich languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Schneider, G., Pettersson, E., and Percillier, M. (2017). Comparing rule-based and SMT-based spelling normalisation for English historical texts. In *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, pages 40–46, Gothenburg, May. Linköping University Electronic Press.
- Sennrich, R., Haddow, B., and Birch, A. (2016). Improving neural machine translation models with monolingual data. In *Proceedings of the ACL2016 (Volume 1: Long Papers)*, pages 86–96.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*, pages 223–231.
- Tiedemann, J. (2012). Parallel data, tools and interfaces in opus. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Tjong Kim Sang, E., Bollman, M., Boschker, R., Casacuberta, F., Dietz, F., Dipper, S., Domingo, M., van der Goot, R., van Koppen, J., Ljubešić, N., et al. (2017). The clin27 shared task: Translating historical text to contemporary language for improving automatic linguistic annotation. *Computational Linguistics in the Netherlands Journal*, 7:53–64.
- Varga, D., Halácsy, P., Kornai, A., Nagy, V., Németh, L.,

- and Trón, V. (2007). Parallel corpora for medium density languages. *Amsterdam Studies In The Theory And History Of Linguistic Science Series 4*, 292:247.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS)*, pages 6000–6010.
- Zoph, B., Yuret, D., May, J., and Knight, K. (2016). Transfer learning for low-resource neural machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1568 – 1575.