

# Developing a Neural Machine Translation System for Irish

Arne Defauw, Sara Szoc, Tom Vanallemeersch, Anna Bardadym, Joris Brabers, Frederic Everaert, Kim Scholte, Koen Van Winckel, Joachim Van den Bogaert

CrossLang  
Kerkstraat 106  
9050 Gentbrugge  
Belgium  
{firstname.lastname}@crosslang.com

## Abstract

In this paper we develop a neural machine translation (NMT) system for translating from English into Irish and vice versa. We evaluate the performance of NMT on the resource-poor English-Irish (EN-GA) language pair, show that we can achieve good translation quality in comparison to previously reported systems, and outperform Google Translate™ with several BLEU points on a domain-specific test set related to the legal domain. We show that back-translation of monolingual data closely related to the domain of the test set can further increase the model’s performance. Finally, we present a lightweight method for filtering synthetic sentence pairs obtained via back-translation using a tool for misalignment detection. We show that our approach results in a slightly higher BLEU score while requiring less training data.

## 1 Introduction

In recent years the performance of machine translation systems has been improving significantly thanks to the shift from statistical machine translation (SMT) to NMT. Replacing the recurrent neural network (RNN) architecture with a Transformer architecture that relies entirely on self-attention to compute representations of its input and output has set a new state of the art in the field of machine translation (Vaswani et al. 2017).

However, for low-resource languages, the performance of (neural) machine translation systems can still be disappointing, as pointed out for instance by Koehn et al. (2017). Many approaches have been suggested to improve the quality of NMT in such a low-resource setting, among which multilingual models (Johnson et al. 2016; Thanh-Le et al. 2016), unsupervised approaches (Lample et al. 2019) and systems relying on back-translation (Sennrich et al. 2016) have been the most successful.

In this paper we focus on the translation of the English-Irish language pair using NMT. The Irish language has been categorized as a ‘weak or not supported language’ by the META-NET report (Judge et al. 2012) due to the lack of good translation resources. Despite this relatively low availability of resources, both in terms of monolingual and bilingual content, it has been shown that an SMT system can achieve promising translation quality both in a domain-specific setting (Dowling et al. 2015) and in a more broad-domain context (Arcan et al. 2016).

First steps were taken by Dowling et al. (2018) to apply NMT methods to EN-GA MT, although the resulting NMT system performed significantly worse than SMT, scoring more than 6 BLEU lower on an in-domain test set.<sup>1</sup>

In this work we will further explore the potential of NMT for the EN-GA language pair. We add web-crawled parallel data to the publicly available resources used in previous studies and show relatively good translation quality both on a domain-specific test set and on a more generic test set. Next, our experiments confirm that NMT translation quality for GA→EN can be significantly improved using back-translation.

---

© 2019 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC BY-ND.

<sup>1</sup> Reported BLEU score of 40.1, compared to a BLEU score of 46.4 for the SMT system.

Due to a lack of Irish monolingual data, back-translation was less useful for EN→GA NMT.

Finally, a set of experiments was performed in which the synthetic parallel corpus, obtained via back-translation, was filtered with Bicleaner<sup>2</sup> (Sánchez-Cartagena et al. 2018), a tool for misalignment detection. We show that applying misalignment detection on a synthetic corpus before adding it to the parallel training data results in small increases in BLEU score and could be a useful strategy in terms of data selection.

Filtering of parallel data has been the subject of various studies (Axelrod et al. 2011; van der Wees et al. 2017), but such data selection methods have only been scarcely investigated in the context of back-translation. Fadaee et al. (2018) suggest several sampling strategies for synthetic data obtained via back-translation, targeting difficult to predict words. More closely related to our filtering technique is the method proposed by Imankulova et al. (2017). They present a method in which a synthetic corpus was filtered by calculating the BLEU score between the target monolingual sentence and the translation of the synthetic source sentence in the target language and report small increases in translation quality in a low-resource setting.

## 2 Materials and methods

### 2.1 Data

In this section, we give an overview of the data used for training our NMT systems. Both bilingual and monolingual data are used.

In Table 1 an overview of the parallel data is shown. Three types of data were collected: 1) Baseline data, i.e. a collection of publicly available resources; 2) Web-crawled data, i.e. data scraped from two bilingual websites, and 3) ParaCrawl data.

The baseline data has been described in detail in previous publications (Dowling et al. 2015; Arcan et al. 2016). We note that there are some other parallel corpora available for the EN–GA language pair, the largest of which are the *KDE*<sup>3</sup>

and *GNOME*<sup>4</sup> corpora. However, due to the very specific nature of these corpora, they were not included in the training data.

The web-crawled dataset consists of sentence pairs we scraped and aligned ourselves from two bilingual websites. This data was scraped using Scrapy<sup>5</sup> and then document-aligned using Malign<sup>6</sup>, a tool for document alignment that makes use of MT. Sentence alignment of these document pairs was subsequently performed using Hunalign<sup>7</sup> (Varga et al. 2005). Finally, the misalignment detection tool Bicleaner (Sánchez-Cartagena et al. 2018) was applied to these aligned sentences (the Bicleaner threshold was set to 0.5<sup>8</sup>).

Parallel corpus	#unique sentence pairs	#EN tokens
DGT <sup>9</sup>	38,672	948,037
+ DCEP <sup>10</sup>	7,303	158,035
+ EU Bookshop <sup>11</sup>	95,705	2,182,873
+ Irish legislation <sup>12</sup>	172,272	4,285,570
+ EU constitution <sup>13</sup>	6,702	140,101
<b>= Baseline data</b>	<b>315,748</b>	<b>7,634,954</b>
www.education.ie	128,016	3,408,864
+ www.courts.ie	2,791	66,260
<b>= Web-crawled data</b>	<b>130,807</b>	<b>3,475,124</b>
<b>ParaCrawl data<sup>14</sup> (0.5&lt;Bicleaner score)</b>	<b>784,606</b>	<b>17,646,315</b>
<b>Total</b>	<b>1,195,067</b>	<b>27,860,572</b>

Table 1 Parallel NMT training data (EN–GA)

We also used the ParaCrawl corpus as a bilingual resource. We used the Raw EN–GA ParaCrawl corpus v4.0<sup>15</sup> consisting of 156M sentence pairs. ParaCrawl is known to contain a diversity of noise such as misalignments, untranslated sentences, non-linguistic characters, wrong encoding, language errors, short segments etc. that may harm NMT performance (Khayrallah et al. 2018). Therefore, only pairs with a Bicleaner score

<sup>2</sup> <https://github.com/bitextor/bicleaner>

<sup>3</sup> <http://opus.nlpl.eu/KDE4.php>

<sup>4</sup> <http://opus.nlpl.eu/GNOME.php>

<sup>5</sup> <https://scrapy.org>

<sup>6</sup> Now part of the ParaCrawl pipeline:

<https://github.com/bitextor/bitextor>

<sup>7</sup> <http://mokk.bme.hu/en/resources/hunalign>

<sup>8</sup> Threshold based on manual inspection.

<sup>9</sup> <http://opus.nlpl.eu/DGT.php>

<sup>10</sup> <https://wt-public.emm4u.eu/Resources/DCEP-2013/DCEP-Download-Page.html>

<sup>11</sup> <http://opus.nlpl.eu/EUbookshop-v2.php>

<sup>12</sup> <http://www.gaois.ie/en>

<sup>13</sup> <http://opus.nlpl.eu/EUconst.php>

<sup>14</sup> <https://paracrawl.eu>

<sup>15</sup> <https://s3.amazonaws.com/web-language-models/paracrawl/release3/en-ga.classify.gz>

greater than 0.5 were considered. After deduplication, the size of the ParaCrawl corpus diminishes to 784k sentence pairs.

We extracted two test sets from this parallel data: a domain-specific test set (legal domain) and a more generic test set, both consisting of 3k sentence pairs, and held out from the *Irish Legislation* and *EU Constitution* corpora (legal) and the *DGT* and *DCEP* corpora (generic), respectively. We note that the *DGT* corpus is derived from the translation memories of the European Commission's Directorate-General for Translation, while the *DCEP* corpus originates from the European Parliament. While both are linked to the administrative text type, the *DCEP* corpus includes a wider variety of text types compared to the former (Hajlaoui et al. 2014).

In comparison to previous publications, two relatively large EN–GA corpora could not be used in this work, due to their not being publicly available: 1) a set of translation memory files from the Department of Arts, Heritage and the Gaeltacht (DAHG), containing approximately 40k parallel sentences (Dowling et al. 2015); 2) translations of second level textbooks (Cuimhne na dTéacsleabhar) in the domain of economics and geography, holding around 350k sentence pairs (Arcan et al. 2016).

Monolingual EN corpus	#unique EN sentences	#EN tokens
DCEP	2,004,062	52,163,146
+ DGT	1,644,325	39,468,227
+ EAC <sup>16</sup>	1,341	21,828
+ ECDC <sup>17</sup>	2,027	35,935
+ JRC-Acquis <sup>18</sup>	463,073	13,316,245
<b>= Total</b>	<b>3,989,791</b>	<b>102,178,555</b>
Monolingual GA corpus	#unique GA sentences	#GA tokens
Wikipedia	217,695	6,540,334
+ ParaCrawl corpus, GA side (0.0<Bicl. score<0.5)	301,141	5,661,168
<b>= Total</b>	<b>518,836</b>	<b>12,201,497</b>

Table 2 Data (EN|GA) for back-translation

<sup>16</sup> <https://ec.europa.eu/jrc/en/language-technologies/eac-translation-memory>

<sup>17</sup> <https://ec.europa.eu/jrc/en/language-technologies/ecdc-translation-memory>

<sup>18</sup> <http://opus.nlpl.eu/JRC-Acquis.php>

In Table 2 we give an overview of the monolingual data used for back-translation (see Section 2.4). The English monolingual corpus consists of the English side of the EN–FR *DCEP*, *DGT*, *EAC*, *ECDC* and *JRC-Acquis* corpora. The last corpus is related to the legal domain, while the other corpora serve as generic data for this test case.

The Irish monolingual corpus consists of data extracted from Irish Wikipedia articles and the Irish side of the ParaCrawl corpus (sentences with a Bicleaner score greater than 0 and smaller than 0.5). Imposing a Bicleaner threshold equal to 0 ensures that non-Irish sentences and noisy sentences in general are excluded.

Sentences from the monolingual corpora overlapping with the English or Irish side of the test sets were excluded.

## 2.2 Machine translation

Neural MT engines were trained with OpenNMT-tensorflow<sup>19</sup> using the Transformer architecture during 20 epochs and default training settings.<sup>20</sup> Preprocessing was done with aggressive tokenization, and joint subword (BPE) and vocabulary sizes set to 32k. NMT systems were trained in both translation directions, EN→GA and GA→EN. The translation quality of the MT models is measured by calculating BLEU scores on two held out test sets (see Section 2.1). The reported BLEU score is the maximal BLEU reached in the last 10 epochs of training.

## 2.3 Bicleaner

Bicleaner detects noisy sentence pairs in a parallel corpus by estimating the likelihood of a pair of sentences being mutual translations (value near 1) or not (value near 0). Very noisy sentences are given the score 0 and detected by means of hand-crafted hard rules. This set of hand-crafted rules tries to detect evident flaws such as language errors, encoding errors, short segments and very different lengths in pairs of parallel sentences. In a second step, misalignments are detected by means of an automatic classifier. Finally, sentences are scored based on fluency and diversity. More details are provided by Sánchez-Cartagena et al. (2018).

<sup>19</sup> <https://github.com/OpenNMT/OpenNMT-tf>

<sup>20</sup> <https://github.com/OpenNMT/OpenNMT-tf/blob/master/opennmt/models/catalog.py>

In order to clean the EN–GA web-crawled corpus and the synthetic data obtained via back-translation we used a pre-trained classifier provided by the authors.<sup>21</sup>

## 2.4 Back-translation

Following the methodology described by Sennrich et al. (2016), we paired EN|GA monolingual data (see Table 2) with EN→GA and GA→EN back-translated data, respectively, and used it as additional synthetic parallel training data.

Our MT engines for back-translation were trained using the RNN (Recurrent Neural Network) architecture in OpenNMT (Klein et al. 2017) on the parallel data described in Table 1. The RNN architecture was chosen because of higher inference speed compared to the Transformer architecture (Vaswani et al. 2017; Zhang et al. 2018), which speeds up the process of back-translation.

We applied Bicleaner to the resulting synthetic parallel corpus in an effort to filter out data that may harm the performance of an NMT engine.

## 3 Results

Neural MT engines (see Section 2.2) were trained on the different types of training data described in Section 2.1. We evaluated our MT systems on two test sets: a generic test set and a domain-specific test set related to the legal domain. In Table 3 we give an overview of the results. It shows the results of the generic test sets in the third and fifth

column and the results for the domain-specific test sets in the fourth and sixth column. In the left column, the types of data are indicated, such as synthetic data obtained after back-translation.

Our NMT engines already perform reasonably well in both language directions using the baseline data only. An increase in BLEU score is observed when adding the web-crawled data and the ParaCrawl data: on the generic test set our results become on par with Google Translate™ in both language directions, while on the domain-specific test set our results are clearly better in terms of BLEU score.

We note that Google Translate™ uses the Google Neural Machine Translation system (Wu et al. 2016) for translating from English into Irish and vice versa. We use Google Translate™, being an open-domain translator, merely as a benchmark.

Adding monolingual data paired with back-translated data (see Section 2.4) to the parallel training corpus resulted in mixed outcomes depending on the translation direction: for EN→GA a small decrease in performance was observed on both test sets, while for GA→EN an increase of almost 4 BLEU was observed on the generic test set. A possible explanation for this may be found in the different nature of the EN and GA monolingual data. The GA monolingual data, consisting of *Wikipedia* and *ParaCrawl*, is less relevant for the domain of our test sets, compared to the EN monolingual data consisting of data closely related to the generic and domain-specific test set.

Type of data	#unique sentence pairs	EN→GA generic	EN→GA domain-sp.	GA→EN generic	GA→EN domain-sp.
Baseline	316k	36.2	52.1	45.4	62.3
+ web-crawled	447k	42.5	59.4	52.6	68.1
+ web-crawled + ParaCrawl	1,189k	44.9	63.5	55.2	71.9
+ web-crawled + ParaCrawl + GA→EN back-translation, Bicleaner score > 0.7	1,414k	44.3	63.0	/	/
+ web-crawled + ParaCrawl + EN→GA back-translation, Bicleaner score > 0.7	3,111k	/	/	59.0	71.1
Google Translate™	/	45.3	49.3	55.3	65.3

Table 3 BLEU scores of our NMT systems for different test sets and types of training data

<sup>21</sup> <https://github.com/bitextor/bitextor-data/releases/download/bicleaner-v1.0/en-ga.tar.gz>

As mentioned in Section 2.4, we applied Bicleaner to the resulting synthetic parallel corpus obtained after back-translation. In order to investigate the effect of this filtering strategy, another set of experiments was performed, for two translation directions.

Table 4 shows the results of two experiments for EN→GA. In the first experiment (second row), an engine was trained on the concatenation of the baseline data, the web-crawled data, the ParaCrawl data and the synthetic parallel corpus; no Bicleaner filtering was applied. In the second experiment, the Bicleaner threshold was set at 0.7. We observe that adding filtered synthetic data results in slightly higher BLEU scores on the domain-specific test set, compared to the scenario in which no filtering was applied. On the generic test set, filtering of the synthetic data did not impact the translation quality in terms of BLEU score.

Table 5 shows the results of a similar set of experiments for GA→EN. Various amounts of synthetic data, filtered with various Bicleaner thresholds, were added to the parallel data. In the second and third row of the table, we show results for the case where only domain-specific data (legal, i.e. back-translation of the *JRC-Acquis* corpus) was used for back-translation. The other experiments used the domain-specific monolingual data and a sample of the other EN

monolingual data (i.e. *DCEP, DGT, EAC, ECDC*) for back-translation. In all our experiments, we observed an increase in BLEU score for the generic test set when adding synthetic data to the parallel training corpus. The performance on the domain-specific test set only slightly increases, but only when domain-specific data is used for back-translation, in all other cases a slight decrease is observed. On the generic test set, we found that adding a larger amount of synthetic data results in better performance. However, doubling the amount of training data through back-translation seems sufficient: we only notice small improvements in terms of BLEU score when synthetic data is added beyond the 1:1 ratio between synthetic and real data. This is in line with Poncelas et al. (2018) and Fadaee et al. (2018), who show that a ratio around 1:1 between synthetic and real data is optimal.

Filtering the synthetic data using a misalignment detection tool seems to be a useful strategy in terms of data selection, as slightly higher BLEU scores could be obtained with less data. We refer to the last three rows of Table 5: when using approximately 500k less synthetic sentence pairs, we observe an increase in BLEU of 0.4 on the generic test set (59.0 vs. 58.6). However, we note that one must be careful when setting the Bicleaner threshold: we observe a decrease in BLEU score when increasing the threshold to 0.8.

Type of data	#synthetic sentence pairs before filtering	#unique sentence pairs, total	%synthetic data in total	EN→GA generic	EN→GA domain-sp.
Baseline + web-crawled + ParaCrawl	0k	1,189k	0%	44.9	63.5
+ GA(mono)→EN, no Bicl. threshold	518k	1,708k	30%	44.4	62.0
+ GA(mono)→EN, Bicl. score > 0.7	518k	1,414k	15%	44.3	63.0

Table 4 BLEU scores for EN→GA, given various Bicleaner thresholds for filtering synthetic data

Type of data	#synthetic before filtering	#unique	%synthetic	GA→EN generic	GA→EN domain-sp.
Baseline + web-crawled + ParaCrawl	0k	1,189k	0%	55.2	71.9
+ EN(mono, domain-sp.)→GA, no Bicl. threshold	463k	1,652k	28%	57.5	<b>72.3</b>
+ EN(mono, domain-sp.)→GA, Bicl. score >0.7	463k	1,564k	24%	57.4	72.0
+ EN(mono, domain-sp.+generic)→GA, no Bicl. thres.	1,463k	2,652k	54%	58.4	71.3
+ EN(mono, domain-sp.+generic)→GA, Bicl. score >0.7	1,463k	2,338k	48%	58.6	71.6
+ EN(mono, domain-sp.+generic)→GA, Bicl. score >0.8	1,463k	1,997k	40%	58.4	71.7
+ EN(mono, domain-sp.+generic)→GA, no Bicl. thres.	2,463k	3,652k	67%	58.6	70.6
+ EN(mono, domain-sp.+generic)→GA, Bicl. score >0.7	2,463k	3,111k	62%	<b>59.0</b>	71.1
+ EN(mono, domain-sp.+generic)→GA, Bicl. score >0.8	2,463k	2,536k	53%	58.3	71.4

Table 5 BLEU scores GA→EN, various amounts of synthetic data and Bicleaner thresholds



## 4 Conclusion and future work

In this paper we present a well-performing NMT system for the EN–GA language pair. While EN→GA NMT systems presented in previous work (Dowling et al. 2018) were still performing sub-par, our NMT system outperforms Google Translate™ by several BLEU points on a domain-specific test set in both translation directions.

By carefully adding web-crawled data, we were able to increase the training corpus from 316k sentence pairs to more than 1M parallel sentences, leading to better translation performance in terms of BLEU score. In previous studies (Dowling et al. 2015; Arcan et al. 2016), EN↔GAMT systems were trained on significantly smaller corpora.

Next, we showed that back-translation can increase the performance of EN↔GA NMT systems. For the GA→EN translation direction, back-translation proved very useful, especially when EN monolingual data closely related to the domain of the test set was used for back-translation, in line with Sennrich et al. (2016). For the EN→GA translation direction, back-translation proved less effective. We think this might be solved by using Irish monolingual data that is more closely related to the domain of interest. Such data is, to the best of our knowledge, not publicly available. The Corpus of Contemporary Irish, a monolingual collection of Irish-language texts in digital format,<sup>22</sup> containing around 24.7M words, may be a possible candidate. However, this corpus is only searchable and we could therefore not use it in the present study.

Finally, we presented a lightweight method for filtering synthetic sentence pairs obtained via back-translation, using a tool for misalignment detection, Bicleaner (Sánchez-Cartagena et al. 2018). We show that our approach results in small increases in BLEU score, while requiring less training data.

In future work we will investigate to what extent our proposed methodology can be applied to other languages with a similar amount of data available. Another interesting research direction would be the development of a multilingual MT system which includes not only Irish but also other Gaelic languages, and which is based on methods such as the one described by Johnson et al. (2016). It should also be investigated whether unsupervised MT approaches like the one of Lample et al.

(2019) can be used to increase the translation quality of EN↔GAMT systems.

### Acknowledgement

This work was performed in the framework of the SMART 2015/1091 project ("Tools and resources for CEF automated translation"), funded by the CEF Telecom programme (Connecting Europe Facility).

### References

- Mihael Arcan, Caoilfhionn Lane, Eoin Ó Droighneáin and Paul Buitelaar. 2016. *Iris: English-Irish Machine Translation System*. Proceedings of LREC 10, pages 566–572.
- Amittai Axelrod, Xiaodong He and Jianfeng Gao. 2011. *Domain Adaptation via Pseudo In-domain Data Selection*. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pages 355–362.
- Meghan Dowling, Lauren Cassidy, Eimear Maguire, Teresa Lynn, Ankit Srivastava and John Judge. 2015. *Tapadóir: Developing a Statistical Machine Translation Engine and Associated Resources for Irish*. Proceedings of The Fourth LRL Workshop: "Language Technologies in support of Less-Resourced Languages", Poznan, Poland.
- Meghan Dowling, Teresa Lynn, Alberto Poncelas and Andy Way. 2018. *SMT versus NMT: Preliminary Comparisons for Irish*. Proceedings of AMTA 2018 Workshop: LoResMT 2018, pages 12–20.
- Marzieh Fadaee and Christof Monz. 2018. *Back-translation Sampling by Targeting Difficult Words in Neural Machine Translation*. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pages 436–446.
- Thanh-Le Ha, Jan Niehues and Alexander Waibel. 2016. *Toward Multilingual Neural Machine Translation with Universal Encoder and Decoder*. Proceedings of the 13th International Workshop on Spoken Language Translation (IWSLT).
- Najeh Hajlaoui, David Kolovratnik, Jaakko Väyrynen, Ralf Steinberger and Daniel Varga. 2014. *DCEP – Digital Corpus of the European Parliament*. Proceedings of LREC 9, pages 3164–3171.
- Aizhan Imankulova, Takayuki Sato and Mamoru Komachi. 2017. *Improving Low-resource Neural Machine Translation with Filtered Pseudo-parallel Corpus*. Proceedings of the 4th Workshop on Asian Translation, pages 70–78.

---

<sup>22</sup> <https://www.gaois.ie/g3m/en>

- Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes and Jeffrey Dean. 2016. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. In *Transactions of the Association for Computational Linguistics 5*, pages 339–351.
- John Judge, Ailbhe Ní Chasaide, Rose Ní Dhubhda, Kevin P. Scannell and Elaine Uí Dhonnchadha. 2012. An Ghaeilge sa Ré Dhigiteach – The Irish Language in the Digital Age. In *META-NET White Paper Series*. Georg Rehm and Hans Uszkoreit (Series Editors). Springer. Available online at <http://www.meta-net.eu/whitepapers>.
- Huda Khayrallah and Philipp Koehn. 2018. *On the Impact of Various Types of Noise on Neural Machine Translation*. Proceedings of the 2nd Workshop on Neural Machine Translation. Association for Computational Linguistics, pages 74–83.
- Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart and Alexander M. Rush. 2017. *OpenNMT: Open-Source Toolkit for Neural Machine Translation*. <https://arxiv.org/abs/1701.02810>.
- Philipp Koehn and Rebecca Knowles. 2017. *Six Challenges for Neural Machine Translation*. Proceedings of the First Workshop on Neural Machine Translation, pages 28–39.
- Guillaume Lample and Alexis Conneau. 2019. *Cross-lingual Language Model Pretraining*. <https://arxiv.org/abs/1901.07291>
- Alberto Poncelas, Dimitar Shterionov, Andy Way, Gideon Maillette de Buy Wenniger and Peyman Passban. 2018. *Investigating Backtranslation in Neural Machine Translation*. <https://arxiv.org/abs/1804.06189>
- Víctor M. Sánchez-Cartagena, Marta Bañón, Sergio Ortiz-Rojas and Gemma Ramírez-Sánchez. 2018. *Prompsit's Submission to WMT 2018 Parallel Corpus Filtering Shared Task*. Proceedings of WMT18, Volume 2: Shared Task Papers, pages 995–962.
- Rico Sennrich, Barry Haddow and Alexandra Birch. 2016. *Improving Neural Machine Translation Models with Monolingual Data*. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany. Association for Computational Linguistics, pages 86–96.
- Marlies van der Wees, Arianna Bisazza and Christof Monz. 2017. *Dynamic Data Selection for Neural Machine Translation*. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark. Association for Computational Linguistics, pages 1400–1410.
- Daniel Varga, László Németh, Péter Halácsy, András Kornai, Viktor Trón and Viktor Nagy. 2005. *Parallel Corpora for Medium Density Languages*. Proceedings of RANLP 2005, pages 590–596.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser and Illia Polosukhin. 2017. *Attention is All You Need*. Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017), pages 5998–6008.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey et al. 2016. *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. [arxiv.org/abs/1609.08144](https://arxiv.org/abs/1609.08144)
- Biao Zhang, Deyi Xiong and Jinsong Su. 2018. *Accelerating Neural Transformer via an Average Attention Network*. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 1789–1798.