SMT versus NMT: Preliminary comparisons for Irish

Meghan Dowling meghan.dowling@adaptcentre.ie
Teresa Lynn teresa.lynn@adaptcentre.ie
Alberto Poncelas alberto.poncelas@adaptcentre.ie
Andy Way andy.way@adaptcentre.ie
ADAPT Centre, Dublin City University, Glasnevin, Dublin 9, Ireland

Abstract

In this paper, we provide a preliminary comparison of statistical machine translation (SMT) and neural machine translation (NMT) for English→Irish in the fixed domain of public administration. We discuss the challenges for SMT and NMT of a less-resourced language such as Irish, and show that while an out-of-the-box NMT system may not fare quite as well as our tailor-made domain-specific SMT system, the future may still be promising for EN→GA NMT.

1 Introduction

In recent times, NMT has been widely hailed as a significant development in the improvement in quality of machine translation (MT). However, as a technique that is data-hungry, there is a concern that languages with fewer resources may not benefit to the same degree that well-resourced major languages do. In order to prevent a low-resource language such as Irish being left behind in the context of these advancements, we take the first steps towards applying NMT methods to English→Irish (EN→GA) translation.

Irish is the national and official language of the Republic of Ireland, and an official EU language. While EN→GA MT is rarely used for comprehension purposes,[1] MT is invaluable in meeting the language rights needs of native Irish speakers. MT has already been proven useful in the post-editing environment of an official Irish government department, where the translation of EN→GA documents has been facilitated by a Moses-based statistical machine translation (SMT) system (Dowling et al., 2015). The success of this domain-specific SMT system is due in part to the availability of high quality parallel data in this particular domain (see Table 1). The quality of MT is currently unreliable for official translation in an EU setting, however. This is partly due to a derogation imposed on the production of official Irish language texts in the EU.[2] While the European Commission is moving towards using NMT engines in the new eTranslation platform,[3] Irish is not yet sufficiently supported.

Despite a relatively low availability of resources – in terms of both bilingual and monolingual digital content – we have previously shown that a domain-tailored SMT system can achieve promising translation quality (Dowling et al., 2015).[4] The question remains whether NMT can

[1]Most (if not all) Irish speakers have fluency in English.
[2]`http://publications.europa.eu/code/en/en-370204.htm`
[3]`https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Machine+Translation`
[4]Results: BLEU .43/ TER .46

achieve a similar level of usability for Irish in this setting. While the introduction of deep learning methods to the field of MT has witnessed a breakthrough in recent years, the positive impact of NMT is not felt across the board. As Koehn and Knowles (2017) highlight, current NMT systems can face a number of challenges when dealing with specific tasks. These challenges include low-resourced languages, low-frequency words arising from inflection, long sentences, and out-of-domain texts. The latter may not apply to our test case, as the success of our earlier SMT system lies in the closed domain nature of the use case (public administration data), yet the other factors are very real for the Irish language in general. In this study, we report on recent scores from the training of an updated Irish SMT engine, based on our latest data sets. We then present a preliminary NMT baseline, based on the same training and test data as previous SMT experiments, in order to investigate its strengths and weaknesses with respect to Irish.

The paper is divided as follows: Section 2 provides the context within which our work is relevant, both in terms of low-resourced MT and the use of MT in professional translation environments. In Section 3 we outline the datasets we use in training and testing, and give some background on the types and sources of this data. Section 4 details how the SMT and NMT experiments were implemented. Section 5 provides updated results for EN-GA SMT in this domain and establishes preliminary results for EN-GA NMT. Finally, in Section 6 we provide some conclusions and indicate possible options for future work in this area.

2 Related Work

As discussed above, currently the primary focus of the application of Irish MT is within the context of a professional translation workflow (involving post-editing by human translators), and as such, progress in this area in terms of advances in state-of-the-art approaches is of interest to us. For many years, there have been extensive studies to show how the integration of MT within such a workflow (often complementary to the use of translation memory tools) improves productivity, both in industry-based and in academic-based research (e.g. Etchegoyhen et al. (2014); Arenas (2008)). With the introduction of NMT methods, there have been subsequent studies examining the differences between the impact that SMT and NMT have within such a setting. For example, Bentivogli et al. (2016) carried out a small scale study on post-editing of English→German translated TED talks, and concluded that NMT had made significantly positive changes in the field. Bojar et al. (2016) report a significant step forward using NMT instead of SMT in the automatic post-editing tasks at the Conference on Statistical Machine Translation (WMT16). More recently, Castilho et al. (2017) carried out a more extensive quantitative and qualitative comparative evaluation of PBSMT and NMT using automatic metrics and professional translators. Results were mixed overall. They varied from showing positive results for NMT in terms of improved (perceived) fluency and errors, to achieving no particular gains over SMT at document level for post-editing. While these studies were carried out on better resourced language pairs (English→German, Portuguese, Russian and Greek), they are still highly relevant in indicating the potential impact that the change in MT approaches can have in real-life translation scenarios.

Aside from examining the impact on translator productivity, there has also been increased focus in addressing the shortcomings of NMT, such as those outlined by Koehn and Knowles (2017). As such, a number of innovative approaches have emerged to this end. The application of various transfer learning methods has proven successful for certain low-resourced language (Zoph et al., 2016; Passban et al., 2017), as has the inclusion of linguistic features when addressing data sparsity that faces morphologically rich languages (Sennrich and Haddow, 2016). Luong et al. (2015) show that the use of attention-based NMT can have positive results in many aspects of MT, including the handling of long sentences.

In the case of Irish language, the lack of sufficient data, along with a lack of skilled resources has resulted in limited progress in the area of English-Irish (EN-GA) MT to date: As discussed in Section 1, a domain specific (public administration) SMT system is currently in use by in-house translators in the Department of Culture, Heritage and the Gaeltacht (DCHG) (Dowling et al., 2015). DCHG is the Irish government department responsible for ensuring that the Irish language needs of the Irish public are being met by the government. In addition, some steps have been taken to develop a more broad domain system (Arcan et al., 2016). This current study is, to our knowledge, the first attempt to apply NMT methods to EN-GA MT.

3 Data

In order to provide an accurate comparison in our SMT vs NMT experiments, we use the same data sets for each approach (apart from the absence of monolingual data in the NMT set-up). This data is almost identical to the datasets that we have used in training earlier SMT systems Dowling et al. (2015). We indicate an extended version of a dataset with ± and our additional datasets with † in Tables 1 and 2.

Bilingual corpora – translation model

Our data sets are based on that of our earlier SMT systems, with some additional corpora. The domain in question is public administration. As Table 1 shows, the majority of the data used to train the translation model was provided by DCHG. These sources include staff notices, annual reports, website content, press releases and official correspondence. We supplement the existing corpus with additional recently translated in-domain data provided by the DCHG. Parallel texts from two EU bodies: the Digital Corpus of the European Parliament (DCEP) and Directorate General for Translation, Translation Memories (DGT-TM) are included in the training data (referred to collectively as 'EU' in Table 1). In addition, we include data crawled from websites[5] that were deemed to contain text from a domain similar to public administration (using the ILSP Focused Crawler (Papavassiliou et al., 2013)). Finally, we contribute a new parallel corpus, which was collected from Conradh na Gaeilge (CnaG), an Irish language organisation which promotes the Irish language in Ireland.

Monolingual data – language model

SMT engines require additional monolingual data in order to train a language model that helps to improve the fluency of the SMT output. This monolingual data does not necessarily need to be in-domain, and thus our language model is trained not only on the GA data used for the translation model, but also on a combination of two additional out-of-domain data sets: 'Paradocs', a corpus of national and European legal texts from `www.gaois.ie` and digital GA content we recently sourced from The University Times (UT)[6].

Data–set	# of words (EN)	# of words (GA)	# of sentences	% proportion
DCHG±	995,419	1,094,707	66,214	60.86%
EU	439,262	483,149	29,445	27.06%
Crawled	213,054	234,028	11,770	10.81%
CnaG†	20,692	21,365	1,367	1.25%
TOTAL	**1,668,427**	**1,833,249**	**108,796**	**100%**

Table 1: Size and distribution of translation model training data.

[5] `www.citizensinfo.ie` (An Irish government website that provides information on public services) and `www.teagasc.ie` (Website for the state agency providing research, advisory and education in agriculture, horticulture, food and rural development in Ireland)

[6] The University Times is a university newspaper in Trinity College Dublin

Data–set	# of words	# of sentences
Paradocs	1,596,569	98,758
UT†	15,377	598

Table 2: Additional monolingual (GA) text used for training the SMT language model

4 Experiment Set–Up

4.1 SMT

To attain the most up-to-date results for this use-case, we train a phrase-based SMT system using Moses (Koehn et al., 2007) with the training data described in Section 3. Earlier findings showed that a 6-gram language model helps address divergent word order in EN-GA (Dowling et al., 2015). We therefore use KenLM (Heafield, 2011) to train a 6-gram language model with the monolingual data outlined in table 1. In addition, we implement hierarchical re-ordering tables to address issues surrounding word order. Our earlier system was tailored to address some consistent errors that arose from data sparsity, which resulted from inflectional variations. We took steps to reduce the repetitive task of the translator in correcting these slight orthographic changes at the token level. Our approach involved the introduction of an automated post-editing (APE) module in the pipeline, which consists of hand-coded grammar rules (Dowling et al., 2016). In order to maximise consistency with our previous work, we chose to include this APE module in our MT experiments.

4.2 NMT

Baseline

In order to provide a preliminary NMT baseline for EN-GA in this domain, we implement a 'vanilla' NMT system, i.e. using default parameters where possible (this system is referred to as NMT-base in Figure 1). We use OpenNMT (Klein et al., 2017), which is an implementation of the popular NMT approach that uses an attentional encoder-decoder network (Bahdanau et al., 2014). We train a 2-layer LSTM with 500 hidden layers for 13 epochs. For the sake of comparison we use the same training data as used in the SMT system (see Table 1). The resulting vocabulary size is 50,002 (English) and 50,004 (Irish). Note that we also apply the APE module to the output of the NMT system.

Further NMT experiments

To add to this baseline system, we also perform a few preliminary experiments to investigate the affect that altering parameters or using other methods would have on an EN-GA NMT system.

- **NMT-250** One such experiment involves experimenting with the number of hidden layers in our NMT system. We implement a smaller model i.e. reduced the number of hidden states from 500 to 250. The results for this system are presented in Table 3 wherein this system is referred to as 'NMT-250'.

- **NMT+ADAM** We also experiment with implementing the stochastic gradient descent with 'Adam', a method for stochastic optimisation (Kinga and Adam, 2015). This method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. We implement this method using the recommended learning rate for Adam (0.001) and denote this system in Table 3 as NMT+ADAM.

- **NMT+BPE** In order to address the inflectional nature of the Irish language, we experiment with the use of byte-pair encoding (BPE). BPE is a technique presented by Gage (1994) and adapted for NMT by Sennrich et al. (2016b). In terms of MT, it aims to increase vocabulary coverage by encoding rare and unknown words as sequences of subword units. As data sparsity is an issue especially relevant to a low-resourced inflectional language such as Irish, reducing out of vocabulary (OOV) words is a promising technique. This system is referred to as NMT+BPE in Table 3 and Figure 1.

5 Results and Preliminary Analysis

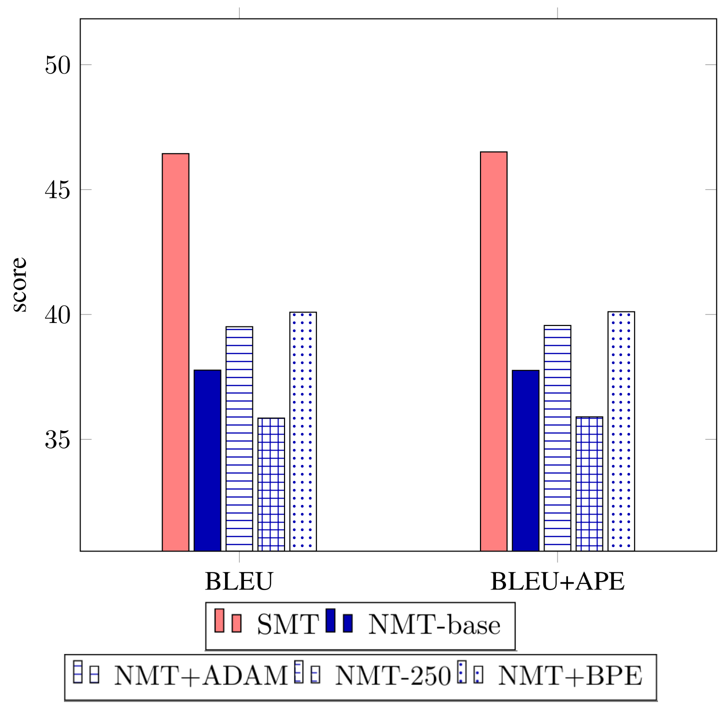

Figure 1: Bar graph displaying the BLEU scores of the SMT and NMT systems, with and without the APE module applied.

Both the SMT and NMT systems were tested on the same test set that were used in earlier experiments (Dowling et al., 2015, 2016), consisting of 1,500 in–domain sentences randomly selected and set aside from the bilingual corpus.

	BLEU	+APE	TER	+APE
SMT	46.44	**46.51**	**43.31**	43.32
NMT	37.77	37.76	47.94	47.79
NMT+ADAM	39.51	39.56	46.98	46.81
NMT-250	35.85	35.9	50.18	50.02
NMT+BPE	40.09	40.11	46.73	46.72

Table 3: BLEU scores for SMT and NMT EN-GA systems before and after applying the automated post-editing module. The highest BLEU score and lowest TER score are highlighted in bold.

We present our results in Table 3 and Figure 1. The results show that for our EN→GA use case, an out-of-the-box NMT system can establish a respectable baseline of BLEU 38.04 and TER 47.94. However, it does not achieve the same level of quality of our tailored SMT system (showing a decrease of between 8.4 and 8.75 BLEU – see Figure 1). Some alterations proved beneficial - the use of Adam as a stochastic optimisation method sees the NMT output increase in BLEU score, and the use of BPE shows an even more marked improvement. Despite these advancements, the scores are still not reaching the same quality as the SMT system.

With respect to the NMT-250 experiment, the use of 250 hidden states in lieu of 500 sees a decrease in BLEU score. More testing will be necessary to identify the optimal number of hidden states for EN-GA NMT.

We note that when the APE module is applied to the NMT output, we see very little change in BLEU score, which is in line with the trends for SMT. However, it should be noted that sentence level analysis carried out in earlier work revealed that the BLEU score increase did not always represent better quality translation from a post-editing perspective (Dowling et al., 2016). This prompts us to carry out some investigation in this regard.

5.1 Sentence-level BLEU

In order to gain a preliminary insight into specific differences between EN-GA SMT and NMT, we chose to perform a sentence-level BLEU on our SMT output and NMT-base output. In Examples 1–4, we highlight some instances where SMT out-performs NMT, and vice-versa.

(1) ***Source:*** Islands[7]
Irish reference: na hOileáin .
SMT: na hOileáin .
NMT: Oileáin .

(NMT decrease: -69.67 BLEU)

(2) ***Source:*** when a requester agrees to amend a request that s / he has submitted , the date of receipt of the refined request is deemed to be the date of receipt of the FOI request .
Irish reference: nuair a chomhaontaíonn iarrthóir leasú a dhéanamh ar iarratas a chuir sé / sí isteach , glacfar leis gurb ionann dáta faighte an iarratais leasaithe agus dáta faighte an iarratais ar SF .
SMT: nuair a chomhaontaíonn iarrthóir leasú a dhéanamh ar iarratas a chuir sé / sí isteach , an dáta faighte an iarratais leasaithe a bheidh an dáta faighte an iarratais SF .
NMT: nuair a aontaíonn iarrthóir <u>iarratas</u> ar <u>iarratas</u> a leasú , meastar go bhfuil an t-iarratas <u>faighte faighte</u> ag an iarrthóir a bheith faighte .

(NMT decrease: -41.56 BLEU)

(3) ***Source:*** this also assists any possible reviews .
Irish reference: Cabhraíonn sé seo le haon athbhreithniú féideartha <u>chomh maith</u> .
SMT: tacaíonn <u>aon</u> athbhreithnithe féideartha seo <u>freisin</u> .
NMT: cabhraíonn sé seo <u>freisin</u> le <u>haon</u> athbhreithniú féideartha .

(NMT increase: +51.62)

(4) ***Source:*** more about CentenaryMayo.ie :
Irish reference: tuilleadh eolais faoi CentenaryMayo.ie :
SMT: <u>níos mó</u> faoi CentenaryMayo.ie :

[7] This is a single word heading.

NMT: tuilleadh faoi CentenaryMayo.ie :

(NMT increase: +35.0)

In Example 1, the SMT BLEU score is significantly higher than that of the NMT output. Delving into the translations, we can see that grammatically, NMT has correctly translated the source text (*Oileáin* 'Islands'). However, the SMT system correctly translates 'Islands' as *na hOileáin*, which literally translates as '*the* Islands'. In this domain, within the context of public administration, it is standard for 'Islands' to refer to the proper noun string '**The** Islands (of Ireland)'. This example highlights the value of a fixed domain, especially for low-resource MT.

Example 2 shows the translation of a longer sentence. It is clear, even to those unfamiliar with the Irish language, why the SMT output prevails in this case. The first phrase in this example is translated perfectly, when compared to the reference – meaning that it is likely that this exact phrase or very similar phrases are present in the training data, and the SMT system is therefore well-equipped to translate it. Looking at the NMT output we can see that a phenomenon, not uncommon in NMT, has occurred: the translations for 'request' and 'receipt' are repeated unnecessarily (*'iarratas'* and *'faighte'*). This is sometimes referred to as 'over-translation' (Tu et al., 2016) and can pose problems for NMT quality.

Examples 3 and 4 show cases where NMT produces translations with a higher BLEU score than that of the SMT system. In Example 3, NMT outputs a more accurate verb (*cabhraíonn* 'assists') as opposed to the SMT output (*tacaíonn* 'supports'), and in fact achieves an almost perfect translation (*freisin* 'also' being a synonym for *chomh maith* 'as well'). It also chooses the correct inflection for *haon* 'any', which the SMT system fails to do (outputting *aon*). The *h* inflection is required following the vowel ending on the preceding preposition *le* 'with'. In Example 4, we again see NMT achieving an almost perfect translation. The translation generated by the SMT system in this case is not entirely incorrect. However, it could be argued that the NMT output is more fluent. Both of these examples highlight the strength in fluency sometimes observed with NMT.

6 Conclusion and Future Work

Our study reveals that an out-of-the-box NMT system, trained on the same EN–GA data, achieves a much lower translation quality than a tailored SMT system, at least in terms of automatic metrics. These results are not necessarily surprising given that Irish presents many of the known challenges that NMT currently struggles with (data scarcity, long sentences and rich morphology). Despite this, these preliminary experiments cannot suggest that NMT be discounted with respect to the future of EN-GA MT. it should be noted that minimal tuning and additional processing has been carried out to date.

In future experiments, we hope to investigate methods for tailoring NMT to this particular domain and language pair. A possible avenue of research to explore is the inclusion of linguistic features in NMT such as the work carried out by Sennrich and Haddow (2016). We wish to address over-translation issues discussed in Section 5, possibly with the use of coverage vectors (Tu et al., 2016). Another approach worth considering is addressing the divergent word order in the EN-GA language pair with a pre-reordering approach such as the one taken by Du and Way (2017). Methods which address data sparsity will also be investigated – options include the use of back translation (Sennrich et al., 2016a) and/or data augmentation (Fadaee et al., 2017).

In addition, it will be important in the future to include human evaluation in our studies to ensure that the MT systems designed for public administration use will be optimised to enhance the task of a human translator, and will not merely be tuned to automatic metrics.

Finally, the derogation on the production of Irish language documents within the EU is due to lift in 2021. By this point there will be a huge increase in the (already high) EN↔GA translation demands, and national and EU bodies will need look to technological advancements to support professional EN↔GA translators. It is vital, therefore, that MT resources are well-developed, up-to-date and designed accordingly to meet this demand.

Acknowledgments

This work was part-funded by the Department of Culture, Heritage and the Gaeltacht (DCHG) and is also supported by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2016) and is co-funded by the European Regional Development Fund. We would also like to thank the four anonymous reviewers for their useful comments.

References

Arcan, M., Lane, C., Droighneáin, E. O., and Buitelaar, P. (2016). Iris: English-Irish machine translation system. In *The International Conference on Language Resources and Evaluation*.

Arenas, A. G. (2008). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus*, 7(1):11–21.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473.

Bentivogli, L., Bisazza, A., Cettolo, M., and Federico, M. (2016). Neural versus phrase-based machine translation quality: a case study. In *Proceedings of the 2016 Conference on Empirical Methods in NaturalLanguage Processing*, pages 257–267. Association for Computational Linguistics.

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., Jimeno Yepes, A., Koehn, P., Logacheva, V., Monz, C., Negri, M., Neveol, A., Neves, M., Popel, M., Post, M., Rubino, R., Scarton, C., Specia, L., Turchi, M., Verspoor, K., and Zampieri, M. (2016). Findings of the 2016 conference on machine translation. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 131–198. Association for Computational Linguistics.

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y., Lohar, P., Way, A., Miceli Barone, A. V., and Gialama, M. (2017). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In *Proceedings of Machine Translation Summit XVI*, Nagoya, Japan.

Dowling, M., Cassidy, L., Maguire, E., Lynn, T., Srivastava, A., and Judge, J. (2015). Tapadóir: Developing a statistical machine translation engine and associated resources for Irish. In *Proceedings of the The Fourth LRL Workshop: "Language Technologies in support of Less-Resourced Languages"*, Poznan, Poland.

Dowling, M., Lynn, T., Graham, Y., and Judge, J. (2016). English to Irish machine translation with automatic post-editing. *PARIS Inalco du 4 au 8 juillet 2016*, page 42.

Du, J. and Way, A. (2017). Pre-reordering for neural machine translation: Helpful or harmful? *The Prague Bulletin of Mathematical Linguistics*, 108(1):171–182.

Etchegoyhen, T., Bywood, L., Fishel, M., Georgakopoulou, P., Jiang, J., van Loenhout, G., del Pozo, A., Maucec, M. S., Turner, A., and Volk, M. (2014). Machine translation for subtitling: a large-scale evaluation. In *Proceedings of theNinth International Conference on Language Resources and Evaluation*.

Fadaee, M., Bisazza, A., and Monz, C. (2017). Data augmentation for low-resource neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 567–573.

Gage, P. (1994). A new algorithm for data compression. *The C Users Journal*, 12(2):23–38.

Heafield, K. (2011). Kenlm: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197. Association for Computational Linguistics.

Kinga, D. and Adam, J. B. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. (2017). Opennmt: Open-source toolkit for neural machine translation. *Proceedings of the 55th Annual Meeting of the Association of Computational Linguistics 2017, System Demonstrations*, pages 67–72.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180. Association for Computational Linguistics.

Koehn, P. and Knowles, R. (2017). Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39.

Luong, T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421.

Papavassiliou, V., Prokopidis, P., and Thurmair, G. (2013). A modular open-source focused crawler for mining monolingual and bilingual corpora from the web. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora*, pages 43–51, Sofia, Bulgaria. Association for Computational Linguistics.

Passban, P., Liu, Q., and Way, A. (2017). Translating low-resource languages by vocabulary adaptation from close counterparts. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(4):29.

Sennrich, R. and Haddow, B. (2016). Linguistic input features improve neural machine translation. In *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, volume 1, pages 83–91.

Sennrich, R., Haddow, B., and Birch, A. (2016a). Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 86–96.

Sennrich, R., Haddow, B., and Birch, A. (2016b). 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)*, volume 1, pages 1715–1725.

Tu, Z., Lu, Z., Liu, Y., Liu, X., and Li, H. (2016). Modeling coverage for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 76–85.

Zoph, B., Yuret, D., May, J., and Knight, K. (2016). Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575.