Leveraging Rule-Based Machine Translation Knowledge for Under-Resourced Neural Machine Translation Models

Daniel Torregrosa and
Nivranshu Pasricha and
Bharathi Raja Chakravarthi and
Maraim Masoud and Mihael Arcan
Insight Centre for Data Analytics
Data Science Institute
National University of Ireland Galway
`name.surname@insight-centre.org`

Juan Alonso and Noe Casas
United Language Group
`name.surname@ulgroup.com`

Abstract

Rule-based machine translation is a machine translation paradigm where linguistic knowledge is encoded by an expert in the form of rules that translate from source to target language. While this approach grants total control over the output of the system, the cost of formalising the needed linguistic knowledge is much higher than training a corpus-based system, where a machine learning approach is used to automatically learn to translate from examples. In this paper, we describe different approaches to leverage the information contained in rule-based machine translation systems to improve a corpus-based one, namely, a neural machine translation model, with a focus on a low-resource scenario. Our results suggest that adding morphological information to the source language is as effective as using subword units in this particular setting.

1 Introduction

In rule-based machine translation (RBMT), a linguist formalises linguistic knowledge into lexicons and grammar rules. This knowledge is used by the system to analyse sentences in the source language and translate them. While this approach does not require any training corpora and grants control over the translations created by the system, the process of encoding linguistic knowledge requires great amounts of expert time. Notable examples of RBMT systems are the original, rule-based Systran (Toma, 1977), Lucy LT (Alonso and Thurmair, 2003) and Apertium (Forcada et al., 2011).

Instead, corpus-based machine translation systems learn to translate from examples, usually in

the form of sentence-level aligned corpora. On the one hand, this approach is generally more computationally expensive and offers limited control over the generated translations. Furthermore, it is not feasible for language pairs that have little to no available parallel resources. On the other hand, it boasts a much higher coverage of the targeted language pair, depending on the availability of parallel corpora. Examples of corpus-based machine translation paradigms are statistical phrase-based translation (Koehn et al., 2003) and neural machine translation (NMT) models (Bahdanau et al., 2015).

In this work, we focused on leveraging RBMT knowledge for improving the performance of NMT systems in an under-resourced scenario. Namely, we used information contained in Lucy LT, an RBMT system where the linguistic knowledge is formalised by human linguists as computational grammars, monolingual and bilingual lexicons. Monolingual lexicons are collections of lexical entries; each lexical entry is a set of feature-value pairs containing morphological, syntactic and semantic information. Bilingual lexicon entries include source-target lexical correspondences and, optionally, contextual conditions and actions. Grammars are collections of transformations to annotated trees. The Lucy LT system divides the translation process into three sequential phases: analysis, transfer, and generation. During the analysis phase, the source sentence is tokenised and morphologically analysed by means of a lexicon that identifies each surface form and all its plausible morphological readings. Next, the Lucy LT chart parser together with a analysis grammar consisting of augmented syntactic rules extracts the underlying syntax tree structure and annotates it. The transfer and generation grammars are then applied in succession on that tree, which undergoes multiple annotations and transformations that add information about the equivalences in the target language and adapt the

original source language structures to the appropriate ones in the target language. Finally, the terminal nodes of the generation tree are assembled into the translated sentence. We focused on the analysis phase, with special interest for two of the features used: the morphological category (CAT) and the inflection class (CL) or classes of the lexical entries.

In order to test this approach, we focused on English-Spanish (both generic and medical domain), English-Basque, English-Irish and English-Simplified Chinese in an under-resourced scenario, using corpora with around one million parallel entries. Results suggested that adding morphological information to the source language is as effective as using subword units in this particular setting.

2 Related work

Sennrich and Haddow (2016) demonstrated the inclusion of various linguistic knowledge, such as morphological features, part of speech (POS) tags and syntactic dependency labels, as input features for the English-German and English-Romanian NMT systems. Baniata et al. (2018) proposed a multitask-based NMT system with POS information for translation between English, modern standard Arabic and Arabic dialects, i.e. Levantine Arabic and Maghrebi Arabic. The work demonstrated that the POS information for the low resourced Arabic dialects was beneficial for the translation quality, specifically if pre-trained FastText models were injected during the NMT training step. Niehues and Cho (2017) jointly trained several English-German natural language processing tasks in one system with shared encoder and one attention model and decoder per task. By integrating additional linguistic resources via multitask learning, the performance of each individual task was improved. Bastings et al. (2017) showed that incorporating syntactic structure such as dependency tree using graph convolutional encoders was beneficial for neural machine translation. Their work focused on exploiting structural information on the source side by adding a second encoder. The goal of their work was to provide the encoder with access to rich syntactic information without placing rigid constraints on the interaction between syntax and the translation task

Etchegoyhen et al. (2018) studied NMT, RBMT, and phrase-based statistical machine translation approaches for Basque-Spanish. The authors focus on different subword unit representations, i.e. linguistically-motivated or frequency-based word segmentation method. Shi et al. (2016) investigated whether an encoder-decoder translation system learns syntactic information on the source side as a side affect of training the neural models. Several syntactic labels of the source sentence were created and logistic regression models using the learned sentence encoding vectors or learned word by word hidden vectors were used to predict these syntactic labels. Aharoni and Goldberg (2017) presented a method to incorporate syntactic information of the target language in an NMT system, showing improved word reordering compared to their baseline system. Eriguchi et al. (2016) proposed an NMT model leveraging syntactic information to improve the accuracy for English→Japanese translation. The phrase structure of the source sentence was recursively encoded in a bottom-up fashion to first produce a vector representation of the sentence, then decode it while aligning the input phrases and words with the output. Bastings et al. (2017) relied on graph-convolutional networks primarily developed for modelling graph-structured data. These networks used predicted syntactic dependency trees of source sentences to produce representations of words that are sensitive to their syntactic neighbourhoods. Nadejde et al. (2017) introduced CCG supertags within the target word sequence as syntactic information, processed by the decoder of their NMT system. Their evaluation showed translation quality improvements for the German→English and Romanian→English translation directions. Similarly, their approach outperformed multi-tasking approach for the same language pairs. García-Martínez et al. (2016) trained their NMT model to simultaneously generate the lemma and its corresponding factors, i.e. POS, gender, and number, demonstrating that factored architecture increases the vocabulary coverage while decreasing the number of unknown words. Ataman and Federico (2018) described the addition of a recurrent neural network to generate compositional representations of the input words, obtaining better results than systems using byte-pair encoding when translating from morphologically rich languages to English. Banerjee and Bhattacharyya (2018) compared two different approaches for subword units when translating from English to Hindi and Bengali, byte pair encoding and morpheme-based segmentation, showing that the latter approach improves the translations, and further improvements can be achieved by combining both.

```
("snake" NST ALO "snake" CL (P-S S-01) KN CNT ON CO SX (N) TYN (ANI))
("snake" VST ALO "snak" ARGS ((($SUBJ N1 (TYN CNC LOC C-POT)) ($ADV DIR)))
    CL (G-ING I-E P-ED PA-ED PR-ES1) ON CO PLC (NF))
```

Figure 1: The word *snake* as a noun (NST) and a verb (VST) in Lucy LT dictionaries. Each entry is composed of a canonical form, the category (POS), and a list of key-value features, such as the inflection class (CL), the vocalic onset (ON), etc.
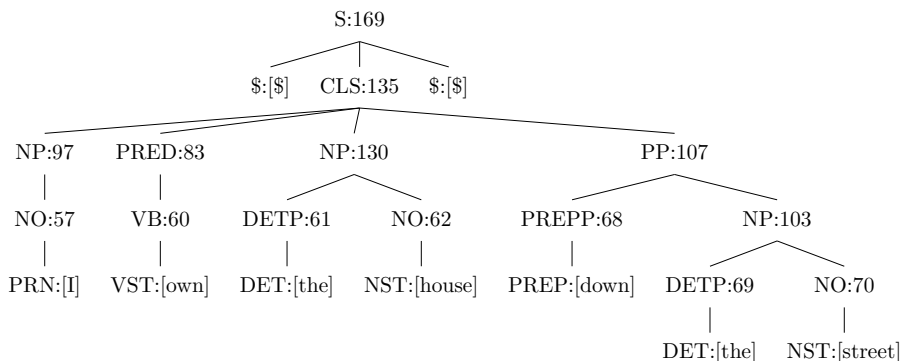
Figure 2: Example of the parse tree for the English sentence *I own the house down the street*.

3 Methodology

In this section, we describe the methodology to leverage rule-based machine translation (RBMT) information in neural machine translation (NMT).

3.1 Information acquisition from RBMT

Lucy LT monolingual lexicons are language-pair independent (i.e. the same English knowledge is used for all translation pairs including English as a source or target language) and mainly encode morphological and contextual information. Each entry has a word or multi-word expression (MWE) along with several features, such as the part of speech (POS) and morphological features. The bilingual lexicons mainly encode word-to-word or MWE-to-MWE translations and describe which target language word should replace each source language word. Still, the direct usage of the lexicon entries as a source of information presented a challenge, as there is no means to determine ambiguous surface words. For example, in English, most nouns will also be classified as verbs, as they share the same surface form; e.g. the word *snake* can be both a noun and a verb (Figure 1). For addressing this problem, we took two different approaches: using **ambiguity classes** that describe all the possible analysis for a given surface word; and using **external information** (in the form of a monolingual POS tagger) to disambiguate. For the former approach, we used a unique tag for each possible CAT and CL values concatenation; e.g. the categories NST and VST and all the inflection classes (CL) for *snake* (Figure 1). For the latter, we used the Stanford POS tagger (Toutanova et al., 2003), that uses the Penn Treebank (Marcus et al., 1994) tag

set for English, and the AnCora (Civit and Martí, 2004) tag set for Spanish, and the IXA pipeline POS tagger (Agerri et al., 2014) with the Universal Dependencies POS tag set (Nivre et al., 2018) for Basque. All POS tag sets were mapped to the tag set used by Lucy LT. If the tagger provided POS tag was equivalent to one or more Lucy LT tags, then the non matching Lucy LT tags were removed. Otherwise, we kept the set of tags; e.g. if the POS tag emits noun as the most likely tag, then only NST and the concatenation of all the inflection classes for the corresponding entry would be used as additional information. As a comparison, we also evaluated NMT models trained with Stanford or IXA POS tags as additional information.

3.2 Leveraging Syntactic Tree Information

In addition to the direct use of the linguistic knowledge in lexicon entries, the grammars (monolingual and bilingual lexicons) were indirectly used by exploring the results of each internal intermediate stage of the translation process, which Lucy LT expresses as annotated trees. For example, the sentence parsed in Figure 2,

I own the house down the street

is encoded as

⦅*I own* ⦅*the house*⦆ ⦅*down* ⦅*the street*⦆⦆⦆.[1]

We use this representation as source text when training the NMT models, as sequence-to-sequence deep neural network models do not generally accept hierarchical information. We also used an additional feature: the linguistic phrase the word belongs to. This information is present in the grandparent of

---

[1]To avoid collisions with parenthesis in the text, we used the left (⦅, U+2985) and right (⦆, U+2986) white parenthesis.

		Source (English)		Target		
		# of Subwords	# of Uniq. Subwords	# of Subwords	# of Uniq. Subwords	# of Lines
English––Spanish (generic)	train	17,919,926	33,212	18,408,749	33,076	1,000,000
	validation	180,290	15,714	185,662	18,804	10,000
	evaluation	178,841	15,031	181,188	18,810	10,000
English––Spanish (EMEA)	train	14,440,740	27,112	15,872,405	29,290	1,036,058
	validation	186,685	11,599	204,174	14,306	10,000
	evaluation	219,752	9,412	242,137	10,979	10,000
English––Basque	train	11,760,808	30,946	10,309,229	32,369	1,357,475
	validation	85,919	9,150	76,532	13,593	10,000
	evaluation	85,163	9,283	75,309	13,546	10,000
English––Irish	train	15,234,432	31,834	16,983,046	32,183	1,090,418
	validation	135,986	12,648	152,224	16,113	10,000
	evaluation	140,696	11,613	152,064	16,174	10,000
English––Simplified Chinese	train	27,878,268	31,471	25,199,106	41,458	995,000
	validation	138,640	12,451	126,191	14,490	5,000
	evaluation	129,440	12,175	119,577	14,431	4,500

Table 1: Statistics on the used training, validation and evaluation datasets.

each node; e.g. in Figure 2 the noun *house* appears in a noun phrase (NP).

4 Experimental Setting

In this section, we describe the resources we used to train and evaluate the systems, along with the neural machine translation framework used.

4.1 Training and Evaluation Datasets

In this work, we focused on NMT for under-resourced scenarios. On the one hand, we consider languages, such as Basque or Irish, which do not have a significant amount of parallel data necessary to train a neural model. On the other hand, an under-resourced scenario can be a specific domain, e.g. medical, where a significant amount of data exists, but does not cover the targeted domain. The Table 1 shows the statistics on the used datasets.

For Basque and Irish, we used the available corpora stored on the OPUS webpage.[2] We used OpenSubtitles2018 (Lison and Tiedemann, 2016),[3] Gnome and KDE4 datasets (Tiedemann, 2012). Additionally, the English-Irish parallel corpus is augmented with second level education textbooks (*Cuimhne na dTéacsleabhar*) in the domain of economics and geography (Arcan et al., 2016).

In addition to that, we also focused on well resourced languages (Spanish and Simplified Chinese), but limited the training datasets to around one million aligned sentences. To ensure a broad lexical and domain coverage of our NMT system, we merged the existing English-Spanish parallel corpora from the OPUS web page into one parallel data set and randomly extracted the sentences. In addition to the previous corpora, we added Europarl (Koehn, 2005), DGT (Steinberger et al., 2014), MultiUN corpus (Eisele and Chen, 2010), EMEA and OpenOffice (Tiedemann, 2009). To evaluate the targeted under-resourced scenario within medical domain, we exclusively used the EMEA corpus. For Simplified Chinese, we used a parallel corpus provided by the industry partner, which was collected from bilingual English-Simplified Chinese news portals.

The corpora were tokenised using the OpenNMT toolkit, with the exception of Simplified Chinese, that was tokenized using Jieba,[4] and lowercased.

4.2 NMT framework

We used OpenNMT (Klein et al., 2017), a generic deep learning framework mainly specialised in sequence-to-sequence models covering a variety of tasks such as machine translation, summarisation, speech processing and question answering as NMT framework. Due to computational complexity, the vocabulary in NMT models had to be limited. In order to overcome this limitation, we used byte pair encoding (BPE) to generate subword units (Sennrich et al., 2016). BPE is a form of data compression that iteratively replaces the most frequent pair of bytes in a sequence with a single, unused byte. We also added the different morphological and syntactic information as word features.

We used the following default neural network training parameters: two hidden layers, 500 hidden

---

[2] opus.nlpl.eu
[3] www.opensubtitles.org

[4] github.com/fxsjy/jieba

LSTM (long short term memory) units per layer, input feeding enabled, 13 epochs, batch size of 64, 0.3 dropout probability, dynamic learning rate decay, 500 dimension embeddings, maximum vocabulary size of 50,000 subwords, maximum of 32,000 unique BPE merge operations, unlimited different values for the word features and between 11 and 23 dimension embeddings for word features.[5]

4.3 Evaluation

In order to evaluate the performance of the different systems, we used BLEU (Papineni et al., 2002), an automatic evaluation that boasts high correlation with human judgements, and translation error rate (TER) (Snover et al., 2006), a metric that represents the cost of editing the output of the MT systems to match the reference. Additionally, we used bootstrap resampling (Koehn, 2004) with a sample size of 1,000 and 1,000 iterations, and reported statistical significance with $p < 0.05$. We also presented a box-and-whisker plot with the first, second and third quartiles as a box, and the first (<0.025) and last (≥ 0.975) 40-quantiles as whiskers, corresponding to $p < 0.05$. In addition, we compared the performance of our NMT systems with the NMT-based Google Translate,[6] and the translations performed using Lucy LT RBMT; for the latter, only English-Spanish and English-Basque models are available.

5 Results

In this section we describe the quantitative and qualitative evaluation of the different models: the NMT baseline (Baseline), baseline enhanced with ambiguous `CAT` and `CL` (CAT-CL), baseline with disambiguated `CAT` and `CL` (CAT-CL D), baseline with external POS tags (POS), baseline with indirect `CAT`, `CL` and syntactic information (CAT-CL L), the hierarchical model (Tree), Lucy LT (RBMT) and Google Translate (Google).

5.1 Quantitative results

The quantitative results of the evaluation are presented in Figure 3. All the models tested significantly outperformed the RBMT system Lucy LT both when using BLEU and TER as evaluation metrics. Even when trained with only around a million sentences, the NMT baseline model for English-Basque and English-Irish performed better

than Google Translate with generic domain corpora, and were not statistically significantly different for English→Simplified Chinese. Unsurprisingly, the in-domain medical domain English-Spanish models outperformed Google Translate. Conversely, Google Translate was significantly better than the NMT baselines only for the English-Spanish generic domain, excluding English→Spanish TER. While some of the feature-enriched models obtained slightly better results in terms of BLEU and TER compared to the baseline, no model obtains scores that are statistically significantly different than the baseline subword model. We also observed that the kind of information we added to the system in the form of `CAT` and `CL` features can also be learned by NMT models based on subword units, that may split the root from the rest of the word. In case of the tree model, the results were consistently lower than the rest. Finally, we learned that the system could not cope with this complex representation with the amount of data available.

5.2 Qualitative results

Table 2 analyses a sentence translated using all different models from Spanish to English. The analysis showed that, even when RBMT makes some grammatical mistakes, the sentence still conveyed the correct message. Nevertheless, it was the only hypothesis with a BLEU of 0, as it shared no four-gram with the reference, and was the hypothesis with the highest TER. The baseline model hypothesis was tied for the best TER score and the second best BLEU score, but it failed to convey the proper message, as it lacked translation for *easing of price increases*.

6 Conclusions and future work

In this work we explored the use of rule-based machine translation (RBMT) knowledge to improve the performance of neural machine translation (NMT) models in an under-resourced scenario, showing that adding morphological information to the source language is as effective as using subword units in this particular setting. We also found that RBMT translations were often adequate but both BLEU and TER poorly reflected this, often scoring worse than incorrect NMT-generated translations.

One of the paths of our future work will further focus on the extraction of RBMT knowledge and the inclusion of transfer rules to improve the performance of the NMT model. A second improvement
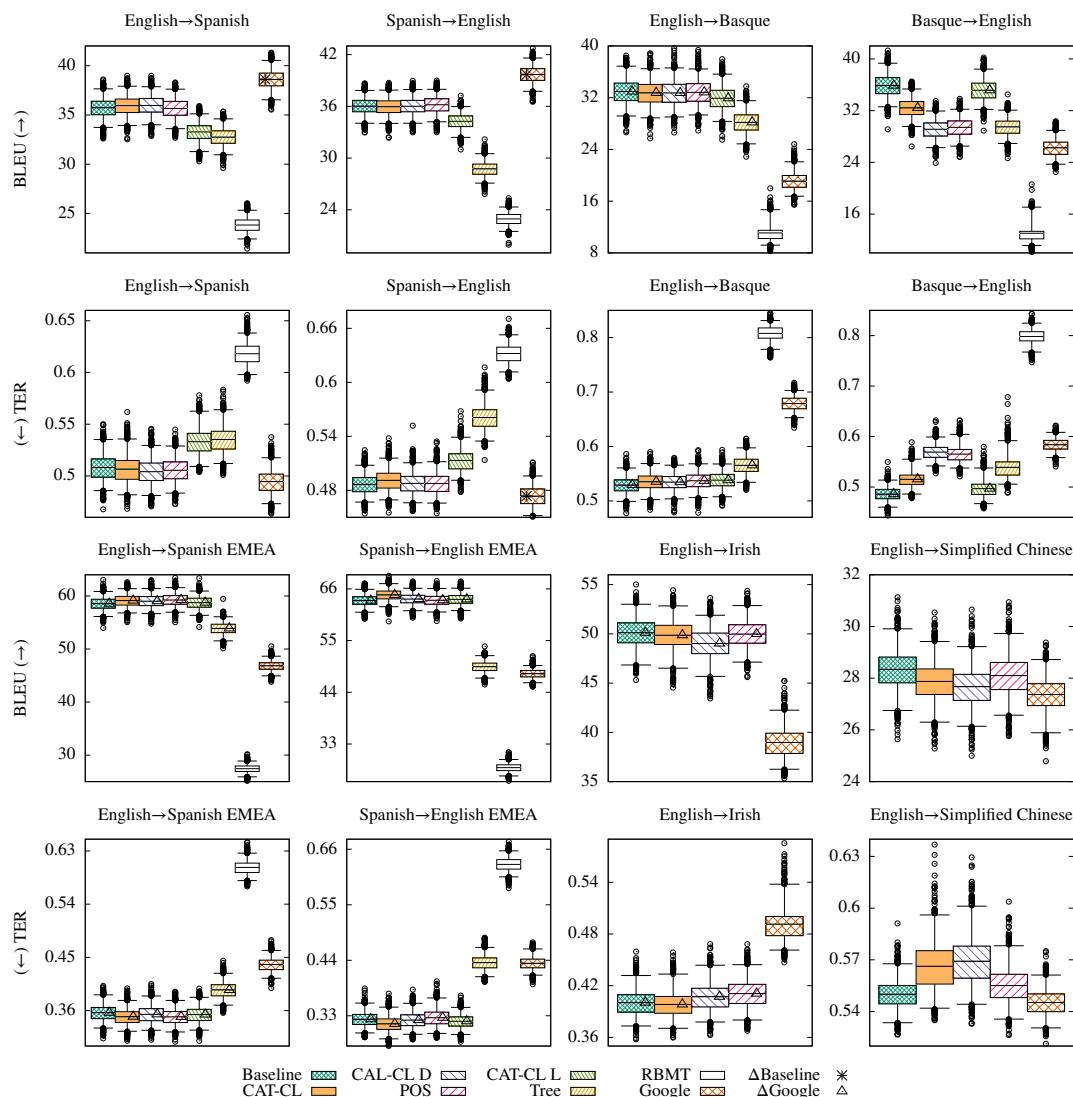
---

[5]The size of the embedding for word features depend on the number of unique values for the feature.
[6]`translate.google.com` retrieved March 2019.

Figure 3: Results for the evaluation for English-Spanish, both for generic and medical (EMEA) domains, English-Basque, English→Irish and English→Simplified Chinese. No RMBT models are available for Irish and Simplified Chinese in Lucy LT. Models marked with ∗ are significantly better than the NMT baseline, and models marked with △ are significantly better than Google Translate. All models are statistically significantly better than RBMT.

Source	Pese a que los incrementos de los precios fueron menores en el segundo semestre de 2008 , los precios siguen siendo muy elevados .	BLEU	TER
Reference	Despite an easing of price increases in the second half of 2008, prices remain at very high levels.		
Baseline	Despite the increases in prices in the second half of 2008, prices remain very high.	47.48	0.35
CAT-CL	*Although* price increases were **minor** in the second half of 2008, prices remain very high.	47.48	0.35
CAT-CL D	*Although* increases in prices were lower in the second half of 2008, prices remain high.	44.50	0.45
POS	*Despite the fact that* price increases were lower in the second half of 2008, prices remain very high.	48.25	0.35
CAT-CL L	*Although* price increases were lower in the second half of 2008, prices remain very high.	47.48	0.35
Tree	*Although* **prices of** prices were lower in the second half of 2008 prices remain very high.	45.51	0.40
RBMT	*Even though* the increases of the prices **were smaller** in the second *semester* of 2008, the prices **keep being** *sky-high*.	0.00	0.70
Google	*Although* the price increases *were lower* in the second half of 2008, prices *are still* very high.	41.81	0.40

Table 2: Qualitative analysis of a sentence translated by all models for Spanish to English translation. Fragments in bold face are translation mistakes, and fragments in italics are translation alternatives that, while being penalised by TER and BLEU, can be considered correct.

path would be using multiple encoders. This approach has been used to improve the performance NMT (Zoph and Knight, 2016), but, in our scenario, one of the inputs would be the output of the RBMT system. As previously mentioned, corpus-based machine translation gives limited control over the output to the user, specially when dealing with homographs and terminology; instead, RBMT gives total control. Combining the source sentence with the RBMT output that contains the user-selected translations might lead to improvements in domain-specific or low resource scenarios.

Finally, we also plan to leverage information contained in other freely available RBMT systems, such as Apertium. While Apertium is a shallow-transfer system, meaning that there is less syntactic information, features similar to the ones used in this work are available in Apertium.

References

Agerri, Rodrigo, Josu Bermudez, and German Rigau. 2014. IXA pipeline: Efficient and Ready to Use Multilingual NLP tools. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Aharoni, Roee and Yoav Goldberg. 2017. Towards String-To-Tree Neural Machine Translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, volume 2: Short Papers, pages 132–140, Vancouver, Canada, July. Association for Computational Linguistics.

Alonso, Juan A and Gregor Thurmair. 2003. The Comprendium Translator system. In *Proceedings of the Ninth Machine Translation Summit*.

Arcan, Mihael, Caoilfhionn Lane, Eoin Ó Droighneáin, and Paul Buitelaar. 2016. Iris: English-irish machine translation system. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia.

Ataman, Duygu and Marcello Federico. 2018. Compositional Representation of Morphologically-Rich Input for Neural Machine Translation. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 305–311.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. *Proceedings of the Third International Conference on Learning Representations*.

Banerjee, Tamali and Pushpak Bhattacharyya. 2018. Meaningless yet meaningful: Morphology grounded subword-level NMT. In *Proceedings of the Second Workshop on Subword/Character LEvel Models*, pages 55–60, New Orleans, June. Association for Computational Linguistics.

Baniata, Laith H., Seyoung Park, and Seong-Bae Park. 2018. A Multitask-Based Neural Machine Translation Model with Part-of-Speech Tags Integration for Arabic Dialects. *Applied Sciences*, 8(12).

Bastings, Joost, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Simaan. 2017. Graph Convolutional encoders for syntax-aware neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1957–1967, Copenhagen, Denmark, September. Association for Computational Linguistics.

Civit, Montserrat and Ma Antònia Martí. 2004. Building cast3lb: A spanish treebank. *Research on Language and Computation*, 2(4):549–574.

Eisele, Andreas and Yu Chen. 2010. MultiUN: A Multilingual Corpus from United Nation Documents. In Tapias, Daniel, Mike Rosner, Stelios Piperidis, Jan Odjik, Joseph Mariani, Bente Maegaard, Khalid Choukri, and Nicoletta Calzolari (Conference Chair), editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation*, pages 2868–2872. European Language Resources Association (ELRA), 5.

Eriguchi, Akiko, Kazuma Hashimoto, and Yoshimasa Tsuruoka. 2016. Tree-to-Sequence Attentional Neural Machine Translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 1: Long Papers, pages 823–833. Association for Computational Linguistics.

Etchegoyhen, Thierry, Eva Martínez Garcia, Andoni Azpeitia, Gorka Labaka, Iñaki Alegria, Itziar Cortes Etxabe, Amaia Jauregi Carrera, Igor Ellakuria Santos, Maite Martin, and Eusebi Calonge. 2018. Neural Machine Translation of Basque.

Forcada, Mikel L, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim O'Regan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, and Francis M Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine translation*, 25(2):127–144.

Garcıa-Martınez, Mercedes, Loıc Barrault, and Fethi Bougares. 2016. Factored Neural Machine Translation Architectures. In *International Workshop on Spoken Language Translation (IWSLT'16)*, Seattle, US.

Klein, Guillaume, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017. OpenNMT: Open-Source Toolkit for Neural Machine Translation. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, System Demonstrations:67–72.

Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational*

Linguistics on Human Language Technology, volume 1, pages 48–54. Association for Computational Linguistics.

Koehn, Philipp. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Koehn, Philipp. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Conference Proceedings: the tenth Machine Translation Summit*. AAMT.

Lison, Pierre and Jörg Tiedemann. 2016. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Marcus, Mitchell, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn Treebank: Annotating Predicate Argument Structure. In *Proceedings of the Workshop on Human Language Technology*, HLT '94, pages 114–119, Stroudsburg, PA, USA. Association for Computational Linguistics.

Nadejde, Maria, Siva Reddy, Rico Sennrich, Tomasz Dwojak, Marcin Junczys-Dowmunt, Philipp Koehn, and Alexandra Birch. 2017. Predicting target language ccg supertags improves neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pages 68–79, Copenhagen, Denmark, September. Association for Computational Linguistics.

Niehues, Jan and Eunah Cho. 2017. Exploiting Linguistic Resources for Neural Machine Translation Using Multi-task Learning. In *Proceedings of the Second Conference on Machine Translation*, pages 80–89, Copenhagen, Denmark, September. Association for Computational Linguistics.

Nivre, Joakim, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Rogier Blokland, Victoria Bobicev, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomaž Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta Gonzáles Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Radu Ion, Elena Irimia, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Kamil Kopacewicz, Natalia Kotsyba, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédayọ̀ Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Slav Petrov, Jussi Piitulainen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roșca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang

Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Dima Taji, Yuta Takahashi, Takaaki Tanaka, Isabelle Tellier, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Seyi Williams, Mats Wirén, Tsegay Woldemariam, Tak-sum Wong, Chunxiao Yan, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Daniel Zeman, Manying Zhang, and Hanzhi Zhu. 2018. Universal Dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Sennrich, Rico and Barry Haddow. 2016. Linguistic Input Features Improve Neural Machine Translation. In *Proceedings of the First Conference on Machine Translation*, volume 1: Research Papers, pages 83–91. Association for Computational Linguistics.

Sennrich, Rico, Barry Haddow, and Alexandra Birch. 2016. Neural Machine Translation of Rare Words with Subword Units. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, abs/1508.07909.

Shi, Xing, Inkit Padhi, and Kevin Knight. 2016. Does String-Based Neural MT Learn Source Syntax? In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1526–1534, Austin, Texas, November. Association for Computational Linguistics.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of association for machine translation in the Americas*, volume 200.

Steinberger, Ralf, Mohamed Ebrahim, Alexandros Poulis, Manuel Carrasco-Benitez, Patrick Schlüter, Marek Przybyszewski, and Signe Gilbro. 2014. An overview of the european union's highly multilingual parallel corpora. *Language Resources and Evaluation*, 48(4):679–707.

Tiedemann, Jorg. 2009. News from OPUS – A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In *Advances in Natural Language Processing*, volume 5, pages 237–248. Borovets, Bulgaria.

Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in opus. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation*, Istanbul, Turkey, may.

Toma, Peter. 1977. Systran as a multilingual machine translation system. In *Proceedings of the Third European Congress on Information Systems and Networks, Overcoming the language barrier*, pages 569–581.

Toutanova, Kristina, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, volume 1, pages 173–180. Association for Computational Linguistics.

Zoph, Barret and Kevin Knight. 2016. Multi-Source Neural Translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 30–34, San Diego, California, June. Association for Computational Linguistics.