

Why Catalan-Spanish Neural Machine Translation? Analysis, comparison and combination with standard Rule and Phrase-based technologies

Marta R. Costa-jussà

TALP Research Center

Universitat Politècnica de Catalunya, 08034 Barcelona

marta.ruiz@upc.edu

Abstract

Catalan and Spanish are two related languages given that both derive from Latin. They share similarities in several linguistic levels including morphology, syntax and semantics. This makes them particularly interesting for the MT task.

Given the recent appearance and popularity of neural MT, this paper analyzes the performance of this new approach compared to the well-established rule-based and phrase-based MT systems.

Experiments are reported on a large database of 180 million words. Results, in terms of standard automatic measures, show that neural MT clearly outperforms the rule-based and phrase-based MT system on in-domain test set, but it is worst in the out-of-domain test set. A naive system combination specially works for the latter.

In-domain manual analysis shows that neural MT tends to improve both adequacy and fluency, for example, by being able to generate more natural translations instead of literal ones, choosing to the adequate target word when the source word has several translations and improving gender agreement. However, out-of-domain manual analysis shows how neural MT is more affected by unknown words or contexts.

1 Introduction

Machine Translation (MT) is the application that allows to translate automatically from one source language to a target language. Approaches vary from rule-based to corpus-based. Rule-based MT systems have been the first largely commercialized MT systems (Douglas Arnold and Lorna

Balkan and R. Lee Humphreys and Siety Meijer and Louisa Sadler, 1994). Years later, corpus-based approaches have reached both the interest in the scientific and industrial community (Hutchins, 1986). Recently, neural MT approach has been proposed. This corpus-based approach uses deep learning techniques (Kalchbrenner and Blunsom, 2013; Cho et al., 2014; Sutskever et al., 2014) and it may be taking over previous popular corpus-based approaches such as statistical phrase or hierarchical-based (Koehn et al., 2003; Chiang, 2007). As a result, large companies, such as Google, have been using rule-based MT, then statistical MT and just very recently, they are replacing some of their statistical MT engines by neural MT engines (Wu et al., 2016).

This paper analyzes how standard neural MT techniques, which are briefly described in section 4.3, perform on the Catalan-Spanish task compared to popular rule-based and phrase-based MT. Additionally, we perform a naive system combination using the standard Minimum Bayes Risk (MBR) technique (Ehling et al., 2007) which reports slight improvements, in terms of standard automatic measures, in in-domain test set but large improvements in out-of-domain test set.

Catalan and Spanish are closely-related languages, which make them particularly interesting for MT and translation performance is quite high for rule-based and statistical-based systems.

Given these similarities, we want to test how neural MT behaves on such related language pairs. This leads us to the main question that this paper tries to solve:

Is neural MT competitive with current well-performing rule-based and phrase-based MT systems?

The answer to this question will be specially useful to industry, since they may decide over results shown if it is worth it to change their current

paradigm which may be either rule or statistical or a combination of both. The aim of this study is to offer a comparison over these systems in terms of translation quality, terms of efficiency or computational cost are out-of-the-scope of this paper.

In this sense, the main contribution of this paper is the analysis and discussion on how the new neural MT approach addresses Catalan-Spanish MT compared to state-of-the-art systems and what are the remaining challenges for this particular language pair.

The rest of this paper is structured as follows. The next section briefly reports on the related work. Section 3 analyses details of this language pair. Section 4 briefly describes each MT approach: rule, phrase and neural-based, respectively. Section 5 details the experimental framework both in data description and in system parameters. Section 6 compares systems based on both automatic and manual analysis and discusses results. Finally, Section 7 reports the main conclusions of this paper.

2 Related work

Previous related publications on the Catalan-Spanish language pair are in rule-based MT (Canals-Marote et al., 2001; Alonso, 2005) and statistical MT (Poch et al., 2009; Costa-jussà et al., 2012). It is worth noting that given the similarity among Catalan and Spanish, Vilar et al (2007) proposed to build a statistical MT system that translated letters, whose underlying idea is similar to recent approaches in neural MT that are character-based (Costa-jussà and Fonollosa, 2016). As far as we are concerned, there are no previous works in neural MT covering Catalan-Spanish language pair.

3 Catalan and Spanish languages

This section reviews several aspects of the language pair we are addressing as a motivation of our study. We point out several social aspects covering language speakers and countries as well as commenting on situations of bilingualism. We also report linguistic aspects of both languages.

3.1 Social aspects

There are around 470 million native speakers for Spanish compared to 4 million for Catalan (as claimed in the Wikipedia). As a consequence, resources for Spanish are much larger than resources



Figure 1: Map showing countries/regions where Spanish (blue) is official and Catalan is spoken (yellow).

for Catalan. Catalan is mainly spoken in Catalonia, Valencia and Balearic Islands, all regions of Spain. There are also some remaining speakers in the south of France and in the island of Sardinia. It is official language of the small country of Andorra. Spanish is official language in 20 countries including Mexico, Colombia and Spain. See Figure 1.

Catalan-Spanish bilingualism only occurs in the regions of Spain and in Andorra. The tendency is that all Catalan native speakers, in practice, also speak Spanish. However, it is not the same for Spanish native speakers. This leads us to a first example of use case for an MT system for this language pair: Spanish (native) speakers that do not understand Catalan. Other use cases include professional translations or web page translations.

3.2 Linguistic aspects

Catalan and Spanish belong to the romance languages which are the modern languages that evolved from Latin. Since both languages are from the same linguistic family, both share similar linguistic features such as morphological inflections or word reordering. Translation between both languages is quite straightforward since there are very few word reorderings and both vocabulary sizes and morphology inflection are quite similar.

4 MT Approaches

This section briefly reports standard baseline architectures for rule-based, phrase-based and neural-based MT. Description for all systems is done in a generic way, particular details from each one used in this work are described later in section 5.2. It is worth mentioning that the rule-based system significantly differs from the other two systems because it is not corpus-based. And

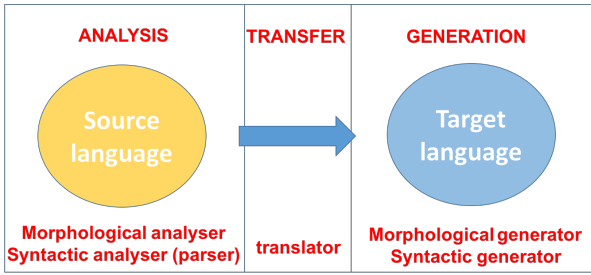


Figure 2: Standard architecture schema of a Rule-based MT system.

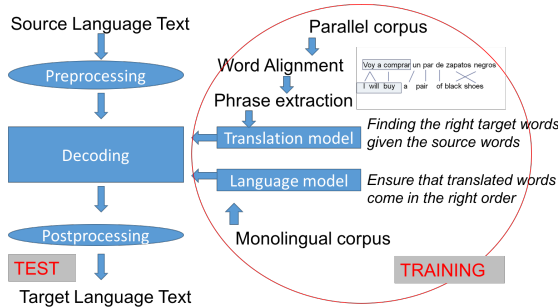


Figure 3: Standard architecture schema of a Phrase-based MT system.

phrase-based and neural-based, although both being corpus-based, they manage data very differently. The phrase-based system uses frequency counts and the neural-based system uses non-linear transformations. The main advantage from corpus-based approaches over the rule-based is that they learn from data. While the main advantage of neural-based over phrase-based is that the architecture allows for an end-to-end optimization.

4.1 Rule-based MT

Rule-based MT combines dictionaries and hand-made rules to generate the target output given the source input. Generally, a morphological and syntactic analysis of the source input is needed before doing the transfer into a simplified target. The final target is generated adding the appropriate morphology and/or syntax. See Figure 2 for an schematic representation of this approach.

4.2 Phrase-based Statistical MT

Standard phrase-based statistical MT (Koehn et al., 2003) focuses on finding the most probable target text given the source text by means of probabilistic techniques. Given a parallel corpus at the level of sentences, statistical co-occurrences are studied to extract a bilingual dictionary of sequences of

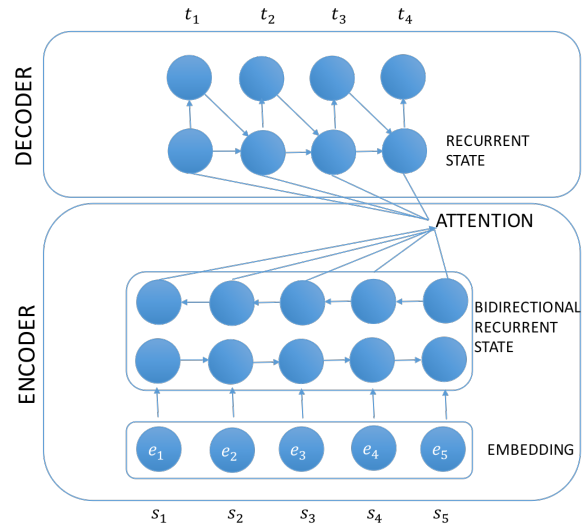


Figure 4: Standard architecture schema of a neural MT system.

words (*phrases*) which are ranked using several features (i.e. conditional and posterior probabilities). Additionally to this bilingual dictionary, which is considered the translation model, other models such as reordering or language models are trained. Note that language modeling is trained on monolingual corpus and it gives information about the fluency of a sentence in the target language. All models are combined in the decoder which uses a beam search to extract the most probable target output given a source input. Note that the system is optimized in several steps since the word alignment is determined before building the translation model. See Figure 3 for an schematic representation of this approach.

4.3 Neural MT

Neural MT computes the conditional probability of the target sentence given the source sentence by means of an autoencoder architecture (Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014; Cho et al., 2014). First, the encoder reads the source sentence (s_1, s_2, \dots, s_N) of N words, the encoder does a word embedding (e_1, e_2, \dots, e_n) and encodes it into an intermediate representation (also referred to as context vector) by means of a recurrent neural network, which uses the gated recurrent unit (GRU) as activation function. The GRU function allows for a better performance with long sentences. Then, the decoder, which is also a recurrent neural network, generates a cor-

responding translation $(t_1, t_2 \dots t_M)$ of M words based on this intermediate representation. Both encoder and decoder are jointly trained using the common statistical technique of Maximum (log-)likelihood Estimation (MLE).

This baseline autoencoder architecture is improved with an attention-based mechanism (Bahdanau et al., 2014), in which the encoder uses a bi-directional recurrent neural network. Now, the decoder predicts each target word with the intermediate representation plus the information of context given by the attention. See Figure 4.

5 Experimental Framework

This section reports details on the data used for training, optimizing and testing as well as a description of the parameters for each system in the comparison.

5.1 Data

We use a large corpus extracted from ten years of the paper edition of a bilingual Catalan newspaper, *El Periódico* (Costa-jussà et al., 2014). The Spanish-Catalan corpus is partially available via ELDA (Evaluations and Language Resources Distribution Agency) in catalog number ELRA-W0053. Development and test sets are extracted from the same corpus, but additionally, to test system performance in out-of-domain, we use a test corpus within the medicine domain. This medical corpus was kindly provided by the Universal-Doctor project¹. Preprocessing was limited to tokenization. Corpus statistics are shown in Table 1.

5.2 System details

Rule-based We use the Apertium rule-based system (Forcada et al., 2011). Apertium is open-source shallow-transfer MT system which was initially designed for the translation between related language pairs. In particular, this rule-based system does not do full syntactic parsing in contrast to the general rule-based architecture described in section 4.1. The system is available from Sourceforge², and we use its last version 1.2.1.

Phrase-based We use Moses (Koehn et al., 2007) which is an open-source phrase-based MT system and it has a large community of developers behind. To build the system, we use stan-

dard/default parameters which include: grow-diagonal-final-and word alignment symmetrization, lexicalized reordering, relative frequencies (conditional and posterior probabilities) with phrase discounting, lexical weights, phrase bonus, accepting phrases up to length 10, 5-gram language model with Kneser-Ney smoothing, word bonus and MERT (Minimum Error Rate Training) optimisation.

Neural-based The neural MT system was built using the open-source software available in github³. This code implements the auto-encoder with attention that we presented in section 4.3. We use the parameters defined in Table 2. Regarding vocabulary limitation, we use a vocabulary size of 90,000 both in Spanish and in Catalan. We replace out-of-vocabulary words (UNKs) using the standard methodology (Jean et al., 2015): we use the word-to-word translation model learned with 'fast-align' (Dyer et al., 2013) or, if not available, the aligned source word is used. We use an embedding of 512 and a dimension of 1024, a batch size of 32, and no dropout, learning-rate of 0.001 and adadelta optimization.

6 Results

This section evaluates the three systems in terms of standard automatic measures. Then, we show some examples of translation outputs and we do a manual comparison.

6.1 Automatic measures

Table 3 shows results in terms of METEOR (Lavie and Agarwal, 2007) and BLEU (Papineni et al., 2002). The best results for the in-domain test set are achieved when using the neural MT system for both translation directions. Best results for the out-of-domain corpus vary depending on the translation direction and measure: for Catalan-to-Spanish, best results are obtained with the phrase-based system; and for Spanish-to-Catalan, best results are obtained with the rule-based system in terms of BLEU, but with the phrase-based system in terms of METEOR. In all cases, results are statistically significant (99%) following the "pair bootstrap resampling" (Koehn, 2004).

To summarise, neural MT is significantly better in the in-domain translation, but it is left behind in out-of-domain. In this out-of-domain task, rule-

¹<http://www.universaldactor.com>

²<https://sourceforge.net/projects/apertium/>

³<http://github.com/nyu-dl/dl4mt-tutorial/>

	Set	Sentences	Words	Vocabulary	Singletons
Catalan	Training	6,5	179,9	713	336
	Development	2.2	60	11	8
	Test	2.2	60	12	7
	Test (Out-of-domain)	0.6	4	1	0.5
Spanish	Training	6,5	165,2	737	343
	Development	2.2	55	12	8
	Test	2.2	56	12	8
	Test (Out-of-domain)	0.6	4	1	0.5

Table 1: Corpus details (in thousands) for Catalan-Spanish.

Vocabulary	90,000
Embeddings	512
Dimension	1024
Batch	32
Dropout	none
Learning rate	0.001
Optimization	Adadelta

Table 2: Neural MT main parameters.

based becomes competitive with corpus-based approaches.

As expected, a simple naive system combination like MBR provides the best final translation results. This means that systems can complement each other, specially for the out-of-domain test set.

6.2 Manual analysis

Manual analysis in this section is intended to complement information provided by the automatic measures in previous section.

Table 4 shows several translation examples from the three systems for the in-domain test set. Examples show the advantages of the neural MT system compared to rule and/or phrase-based systems. Coherently with previous automatic results, neural MT shows best results. Each example in Table 4 specifically shows how neural MT is able to improve translation in the following terms:

1. *Better gender agreement* (compared to phrase-based MT), which clearly affects fluency of the final translation.
2. *No missing content words* (compared to phrase-based MT) and using the *right verb tense* (compared to the rule-based), which has an impact in adequacy of the translation.

3. *Avoiding redundant words* like “botar” produces a better translation since this would not sound fluent in this context in Catalan.
4. *Choosing the right translation from a polysemic word* improves adequacy and fluency at the same time, the verb “ser” in Catalan has mainly two different translations in Spanish which are “ser” o “estar”, in this case, the correct one is the latter.
5. *Avoiding using literate translation*, if possible, improves translation, in particular, the obligation “s’ha de” in Catalan has to be translated to “hay que” or “han tenido que” in Spanish.
6. *Right preposition translation*.
7. *Adding words to make translation more fluent*. The use of “cuyas” which improves translation.

Finally, example 8 shows the main mistake that neural MT does systematically for this pair of languages: missing initial determiners.

Table 5 shows examples in the out-of-domain text. In this case, example 1 shows how the neural MT system correctly uses the pronoun but it does not coincide with the reference. Example 2, neural MT uses the wrong translation of “pedir” which would correspond to a correct translation in some contexts of the training material. Examples 3 shows how a new unnecessary (but also correct) word is added to the translation in the case of the neural MT. Finally, example 4 shows a missing translation of a word, which is an out-of-vocabulary.

Most of neural MT errors could be addressed by using already existing techniques. The example of missing determiners could be solved using

System	CAES				ESCA			
	In-domain		Out-domain		In-domain		Out-domain	
	METEOR	BLEU	METEOR	BLEU	METEOR	BLEU	METEOR	BLEU
Rule	87.11	75.20	67.22	50.53	83.92	70.15	63.77	50.41
Phrase	90.59	81.80	72.31	57.20	90.88	84.24	64.31	49.67
Neural	90.65	83.01	66.87	52.10	92.15	86.31	60.73	47.63
MBR	91.51	83.35	73.15	58.07	92.25	86.33	66.73	53.20

Table 3: METEOR and BLEU results. In bold, best results among individual systems and system combination.

1	SRC Rule Phrase NN REF	una cosa lúdica y divertida una cosa lúdica i divertida una cosa lúdica i divertit una cosa lúdica i divertida una cosa lúdica i divertida
2	SRC Rule Phrase NN REF	los investigadores creen que el cuerpo pudo ser arrastrado desde otro lugar . els investigadors creuen que el cos va poder ser arrossegat des d'un altre lloc els investigadors creuen que el cos va ser arrossegat des d un altre lloc . els investigadors creuen que el cos podria haver estat arrossegat des d un altre lloc . els investigadors creuen que el cos hauria pogut ser arrossegat des d un altre lloc .
3	SRC Rule Phrase NN REF	cerca de 10.000 personas botaron al ritmo de Ruff down (...) prop de 10.000 persones van botar al ritme de Ruff down (...) prop de 10.000 persones van al ritme de Ruff down (...) unes 10.000 persones van saltar al ritme de Ruff down (...)
4	SRC Rule Phrase NN REF	quan el pilot ja era a l'entrada del parc (...) cuando el piloto ya era a la entrada del parque (...) cuando el piloto era ya a la entrada del parque (...) cuando el piloto estaba ya a la entrada del parque (...) cuando el pelotón ya se hallaba en la entrada del parque (...)
5	SRC Rule Phrase NN REF	i s hi han d afegir dues querelles de particulars . y han tenido que añadir dos querellas de particulares . y se han de añadir dos querellas de particulares . (...) y hay que añadir dos querallas de particulares a estos hay que añadir dos querellas de particulares .
6	SRC Rule Phrase NN REF	(...) ja teníem als nostres magatzems un important estoc de peces (...) ya teníamos a nuestros almacenes un importante stock de piezas (...) ya teníamos nuestros almacenes un importante estoc de piezas (...) ya teníamos en nuestros almacenes un importante estoc de piezas (...) ya teníamos en nuestros almacenes un importante estoc de piezas
7	SRC Rule Phrase NN REF	va anunciar ahir el començament d un cicle de conferències que analitzaran l obra d Elliot . les conclusions es recolliran en un llibre . anunció ayer el comienzo de un ciclo de conferencias que analizarán la obra de Elliot . las conclusiones se recogerán en un libro . (...) anunció ayer el inicio de un ciclo de conferencias que analizarán la obra de Elliot . las conclusiones se recogerán en un libro . (...) anunció ayer el comienzo de un ciclo de conferencias que analizarán la obra de Elliot , cuyas conclusiones se recogerán en un libro . anunció ayer el inicio de un ciclo de conferencias que analizarán la obra de Elliot y cuyas conclusiones se recogerán en un libro .
9	SRC Rule Phrase NN REF	el cas dels professionals és diferente . el caso de los profesionales es diferente . en el caso de los profesionales es diferente caso de los profesionales es diferente el caso de los profesionales es diferente .

Table 4: Translation examples.

coverage neural MT (Tu et al., 2016); wrong translations may be reduced using a language model (Gulcehre et al., 2017); and out-of-vocabulary words may be reduced using existing approaches such as Byte Pair Encoding (BPE) (Sennrich et al., 2016) or character-based (Costa-jussà and Fonollosa, 2016). The integration of these new advances for Catalan-Spanish language pair is left for future work.

7 Discussion and Further Work

This paper shows a comparison between rule, phrase and neural MT systems in the Catalan-

Spanish language pair. Performance is better in the case of the neural MT system when using the in-domain test set, but best performance in the out-of-domain test set is better for the rule-based system (Spanish-to-Catalan, in BLEU) and for the phrase-based system (Catalan-to-Spanish).

Regarding our research question: *Is neural MT competitive with current well-performing rule-based and phrase-based MT systems?* Based on the automatic and manual analysis from this paper, the answer is yes, specially, for in-domain sets. Therefore, it is worth it to use neural MT for Catalan-Spanish when building domain spe-

1	SRC Rule Phrase NN REF	Debe ponerse el collarín . Ha de posar-se el collet . ha de posar el collaret . S ' ha de posar el collaret . Ha de posar-se el collar .
2	SRC Rule Phrase NN REF	Pedir un informe . Demanar un informe . Demanar un informe . <i>resumeixi</i> un informe . Demanar un informe .
3	SRC Rule Phrase NN REF	No hauria de consumir alcohol No tendria que consumir alcohol No deberia consumir alcohol <i>yo</i> no tendria que consumir alcohol No deberia consumir alcohol
4	SRC Rule Phrase NN REF	Ha sagnat per algun lloc del cos Ha sangrado por algún lugar de su cuerpo . Ha sangrado por algún lugar del cuerpo . Ha por algún lugar del cuerpo . Ha sangrado por algún lugar de su cuerpo .

Table 5: Out-of-domain translation examples.

cific translation systems. And it is worth it to use system combination for the out-of-domain case. Again, mention that we do not consider efficiency and computational cost comparison in this study.

In this paper, we only implemented a baseline neural MT. Further work would be to show how recent improvements in neural MT like the ones mentioned in previous: Byte Pair Encoding (BPE) (Sennrich et al., 2016), character-based (Costa-jussà and Fonollosa, 2016), coverage (Tu et al., 2016), language model (Gulcehre et al., 2017), multilingual (Firat et al., 2017) and other strategies (Wu et al., 2016) affect this language pair.

Acknowledgments

The author wants to thank the anonymous reviewers for their valuable feedback which helped improving this paper. This work is supported by the Spanish Ministerio de Economía y Competitividad and European Regional Development Fund, through the postdoctoral senior grant *Ramón y Cajal* and the contract TEC2015-69266-P (MINECO/FEDER, UE).

References

Juan Alonso Alonso. 2005. Machine Translation for Catalan-Spanish The real case for productive MT. . In *Proceedings of EAMT*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. volume abs/1409.0473.

Raul Canals-Marote, Anna Esteve-Guillén, Alicia Garrido-Alenda, M Guardiola-Savall, Amaia

Iturraspe-Bellver, Sandra Montserrat-Buendia, Sergio Ortiz-Rojas, Herminia Pastor-Pina, Pedro Pérez-Antón, and Mikel Forcada. 2001. The Spanish-Catalan machine translation system interNOSTRUM. In *Proceedings of MT Summit VIII, Santiago de Compostela*, pages 73–76.

David Chiang. 2007. Hierarchical phrase-based translation. *Comput. Linguist.*, 33(2):201–228, June.

Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1724–1734.

Marta R. Costa-jussà and José A. R. Fonollosa. 2016. Character-based neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 357–361, Berlin, Germany, August. Association for Computational Linguistics.

Marta R. Costa-jussà, Mireia Farrús, José B. Mariño, and José A.R. Fonollosa. 2012. Study and comparison of rule-based and statistical Catalan-Spanish machine translation systems. *Computing and Informatics Journal*, 31:1001–1026, April.

Marta R. Costa-jussà, March Poch, José A.R. Fonollosa, Mireia Farrús, and José B. Mariño. 2014. A large Spanish-Catalan parallel corpus release for Machine Translation. *Computing and Informatics Journal*, 33.

Douglas Arnold and Lorna Balkan and R. Lee Humphreys and Siety Meijer and Louisa Sadler. 1994. *Machine Translation: An Introductory Guide (Clear Business Studies)*.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia, June. Association for Computational Linguistics.

Nicola Ehling, Richard Zens, and Hermann Ney. 2007. Minimum bayes risk decoding for bleu. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 101–104, Prague, Czech Republic, June. Association for Computational Linguistics.

Orhan Firat, Kyunghyun Cho, Baskaran Sankaran, Fatos T. Yarman Vural, and Yoshua Bengio. 2017. Multi-Way, Multilingual Neural Machine Translation. *Accepted for publication in Computer Speech*

- and Language, *Special Issue in Deep learning for Machine Translation*.
- Mikel L. Forcada, 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.
- Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loic Barrault, Hwei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2017. On Integrating a Language Model Into Neural Machine Translation. *Accepted for publication in Computer Speech and Language, Special Issue in Deep learning for Machine Translation*.
- W. John Hutchins. 1986. *Machine Translation: Past, Present, Future*. John Wiley & Sons, Inc., New York, NY, USA.
- Sébastien Jean, Kyunghyun Cho, Roland Memisevic, and Yoshua Bengio. 2015. On using very large target vocabulary for neural machine translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1–10, Beijing, China, July. Association for Computational Linguistics.
- Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1700–1709, Seattle, Washington, USA, October. Association for Computational Linguistics.
- Philipp Koehn, 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*, NAACL '03, pages 48–54.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 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*, ACL '07, pages 177–180.
- Philipp Koehn. 2004. Statistical Significance Tests For Machine Translation Evaluation. In *Proceedings of EMNLP*, volume 4, pages 388–395.
- Alon Lavie and Abhaya Agarwal. 2007. Meteor: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT '07, pages 228–231.
- Kishore Papineni, 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*, ACL '02, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marc Poch, Mireia Fars, Marta R. Costa-jussà, José B. Mariño, Adolfo Hernández, Carlos A. Henríquez Q., and José A. R. Fonollosa. 2009. The TALP online Spanish-Catalan machine translation system. In *Speech and Language Technologies for Iberian Languages*, pages 105–105, September.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, August. Association for Computational Linguistics.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.
- Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. 2016. Coverage-based neural machine translation. *CoRR*, abs/1601.04811.
- David Vilar, Jan-T. Peter, and Hermann Ney. 2007. Can we translate letters? In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT '07, pages 33–39.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.