Noname manuscript No.
(will be inserted by the editor)

Chinese-Spanish Neural Machine Translation Enhanced with Character and Word Bitmap Fonts

Marta R. Costa-jussà · David Aldón · José A. R. Fonollosa

Received: date / Accepted: date

Abstract Recently, machine translation systems based on neural networks have reached state-of-the-art results for some pairs of languages (e.g. German-English). In this paper, we are investigating the performance of neural machine translation in Chinese-Spanish, which is a challenging language pair.

Given that the meaning of a Chinese word can be related to its graphical representation, this work aims to enhance neural machine translation by using as input a combination of: words or characters and their corresponding bitmap fonts. The fact of performing the interpretation of every word or character as a bitmap font generates more informed vectorial representations. Best results are obtained when using words plus their bitmap fonts obtaining an improvement (over a competitive neural MT baseline system) of almost 6 BLEU, 5 METEOR points and ranked coherently better in the human evaluation.

1 Introduction

Chinese-Spanish are two of the most spoken languages in the world with around 955 million native Chinese speakers and 405 million native Spanish speakers [1]. We can imagine that building accurate machine translation (MT) systems in particular for this pair of languages would have a high impact at the economic and social level. Although research in MT focuses on achieving algorithms independent of languages which could apply to any pair of languages, there are features from each language that are interesting to investigate.

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).

TALP Research Center
Universitat Politècnica de Catalunya
Jordi Girona 1-3, 08034, Barcelona
E-mail: marta.ruiz@upc.edu · E-mail: david.aldon@est.fib.upc.edu · E-mail: jose.fonollosa@upc.edu

[1] https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

The main contribution of this paper consists on the enhancement of Chinese-Spanish MT by introducing bitmap fonts. One Chinese word can consist of either one single character or a sequence of characters. But in either case, the meaning of a word can be related to its character(s) [Liu et al., 2016]. For example, " 汽车" (automobile), " 卡车" (truck), " 摩托车" (motocycle) are all kinds of " 车"(vehicle).

In this paper, we take advantage of this graphical information to feed the MT system. Although standard statistical MT systems (like hierarchical [Chiang, 2007] or phrase-based [Koehn et al., 2003]) may be capable of offering multimodal frameworks [Hitschler et al., 2016], neural MT [Cho et al., 2014a] provides a flexible pipeline to incorporate different format sources of information. We propose to use bitmap fonts to initialize the embeddings used in the neural MT system. This initialization allows to start the embeddings training with a more informative space representation where similar words (in terms of semantics) are closer to each other.

Our research extends the first neural MT publication on this language pair [Aldón et al., 2016, Aldón, 2016]. The current paper adds further theoretical and founded motivational details on the approach, new experiments in a large corpus and further analysis of the results (including a comparison with a standard phrase-based system and a human evaluation).

The rest of the paper is organised as follows. Section 2 describes the related work in Chinese-Spanish MT and in neural MT. Section 3 briefly reports the baseline neural MT system. Section 4 explains motivation and details about integrating information of bitmap fonts in neural MT. Section 5 details the experimental framework by showing corpus statistics and reporting preprocessing techniques. Section 6 and 7 discuss the automatic and human evaluation results, respectively. Both evaluations show significant improvements when using bitmap fonts. Finally, section 8 explains the main conclusions together with some further work.

2 Related work

In this section, we review work that has been done in MT for Chinese-Spanish and, also, previous work in neural MT.

2.1 Chinese-Spanish MT

Surprisingly, there are few works in Chinese-Spanish MT despite being two of the most spoken languages in the world as mentioned in the introduction. Initial works, [Banchs et al., 2006], face Chinese-Spanish challenge through several pivot techniques using English. Later, in 2008, there were two tasks organised by the popular IWSLT evaluation campaign[2] (International Workshop on Spoken Language Translation) between these two languages [Paul, 2008]. The first task was based on a direct translation for Chinese-Spanish. This second task provided corpus in Chinese-English and English-Spanish and asked participants to provide Chinese-Spanish translation

[2] http://iwslt2010.fbk.eu

through pivot techniques. The second task obtained better results than direct translation because of the larger corpus provided.

[Costa-jussà et al., 2012] show a comparison between two types of standard pivots (pseudo corpus and cascade) using English and the direct system. These results show that the pivot and direct techniques do not differ much in their results (for the same amount of corpus), but that the technical pivot cascade is slightly better than the pseudo corpus.

Differently from previous approaches, which were all statistical MT systems, [Costa-jussà and Centelles, 2016] presents the first rule-based MT system for Chinese to Spanish. Authors describe a hybrid method for constructing this system taking advantage of available resources such as parallel corpora that are used to extract dictionaries and lexical and structural transfer rules.

Additionally, to all this research, there are products as Google[3] and Bing[4] translators, and a less popular one like the Chispa Android application and web service[5], that can be useful to tourists or traveling between countries, which use these languages [Centelles et al., 2014].

2.2 Neural MT

Early research on this neural MT can be found on works like [Forcada and Ñeco, 1997, Castaño and Casacuberta, 1997], which were mainly limited by the computational power and short data by means of a Recursive Auto-Associative Memory. Recently, proposed neural MT models used the above explained architecture of encoder-decoders [Sutskever et al., 2014, Cho et al., 2014b, Kalchbrenner and Blunsom, 2013]. This architecture allows for encoding the source text into a fixed-length vector and decoding this fixed-length vector into the target text. To address the long sentence issues, i.e. mainly caused by encoding the input sentence into a single fixed-lentgh vector, [Bahdanau et al., 2015] encode the input sentence into a sequence of vectors and choose a subset of these vectors dynamically when decoding. This frees the neural translation model from having to keep all source sentence information, regardless of its length, into a fixed-length vector, and to deal better with long sentences. This is the baseline system that we are considering in this work, and which will be briefly described in Section 3. Works in neural MT have hugely increased in the last two years. Some research directions are: using character-aware architectures [Costa-jussà and Fonollosa, 2016, Luong and Manning, 2016], going towards a multilingual system [Firat et al., 2017], or even using multimodal translation [Elliott et al., 2015].

As mentioned and as far as we are concerned, this paper is the first work for Chinese-Spanish in neural MT, which extends [Aldón et al., 2016] by providing a justified motivation for this work, giving a more detailed theoretical description of the approach, experimenting with a larger corpus and providing further results analysis (adding a comparison with the standard phrase-based system and a human evaluation).

[3] https://translate.google.com/
[4] https://www.bing.com/translator
[5] http://www.chispa.me

3 Neural MT description

Neural MT uses a neural network approach to compute the conditional probability of the target sentence given the source sentence [Cho et al., 2014b, Bahdanau et al., 2015]. The approach used in this work [Bahdanau et al., 2015] follows the encoder-decoder architecture.

First, the encoder reads the source sentence $s = (s_1, ..s_I)$ and encodes it into a sequence of hidden states $h = (h_1, ..h_I)$. Then, the decoder generates a corresponding translation $t = t_1, ..., t_J$ based on the encoded sequence of hidden states h. Both encoder and decoder are jointly trained to maximize the conditional log-probability of the correct translation.

This baseline autoencoder architecture is improved with a attention-based mechanism [Bahdanau et al., 2015, Luong et al., 2015], in which the encoder uses a bi-directional gated recurrent unit (GRU). This GRU allows for a better performance with long sentences. The decoder also becomes a GRU and each word t_j is predicted based on a recurrent hidden state, the previously predicted word t_{j-1}, and a context vector. This context vector is obtained from the weighted sum of the annotations h_k, which in turn, is computed through an alignment model α_{jk} (a feedforward neural network).

4 Integration of Bitmap Fonts

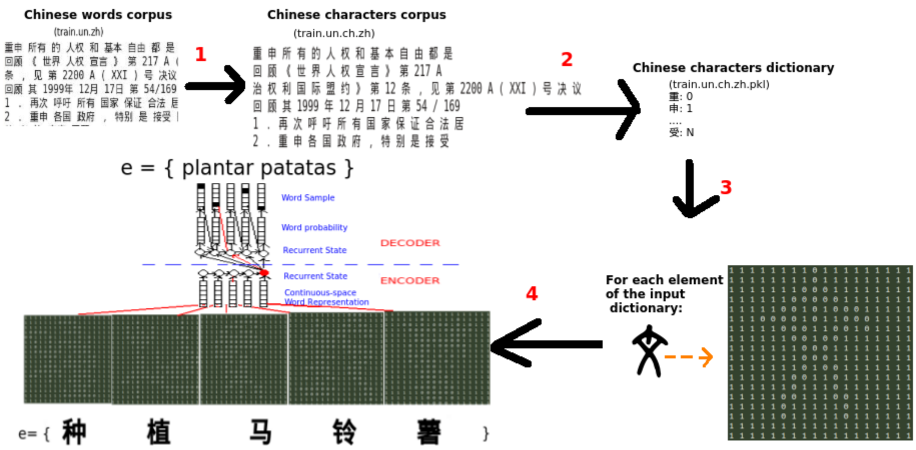

Fig. 1 Integration of Bitmap Fonts

As mentioned in the introduction, Chinese characters have meaning by themselves and similar Chinese words (with one of multiple Chinese characters) can have similar meaning. In fact, each Chinese character is composed of one or more radicals, which contain the semantic radical (carries the meaning) and the phonetic radical (indicates the pronunciation). Chinese characters with the same semantic radicals have related meanings. In this work, we do decompose words into characters or logograms, but not into radicals, which is left for further work.

We enhance the system from previous section 3 to use word bitmap fonts. Like this, we add further information to the neural MT system. Given that the baseline representations of word vectors (embeddings) are random (gaussian), we propose to use bitmap fonts to initialize the embeddings used in the neural MT system. This initialization starts the embeddings training not randomly but with similar words (in terms of semantics) closer to each other.

The integration is done in four steps, which are described as follows and depicted in Figure 1:

Step 1 This step consists of segmenting Chinese words into Chinese characters. This step is only done when using Chinese characters as minimal unit, but it is not necessary in case of using Chinese words.

Step 2 We extract a dictionary from the corpus of Chinese words (or characters), which is ordered from high frequency words to low.

Step 3 For each word in the Chinese dictionary extracted from our corpus, we represent Chinese words (or characters) by means of 2-dimensional bitmap which reflects the shape of the written words (or characters). Like this, we are converting Chinese words (or characters) to bitmap fonts. Then we can get the vector of bits representing the bitmap fonts obtained from each word (or character).

Step 4 The new bitmap font vector becomes the initialization of the embeddings used in the encoder. In this way, we are providing more information to the system. See the integration of this new encoding in the system in step 4 from Figure 1, note that the figure is showing a simplified scheme of the neural network system proposed by [Bahdanau et al., 2015] which is mainly based on a encoder-decoder with the attention-based mechanism. For further simplification, the Figure does not show the attention-based mechanism.

5 Experimental Framework

In this section, we describe the data used for experimentation together with the corresponding preprocessing. In addition, we detail the neural MT parameters chosen.

5.1 Data and Preprocessing

One of the main contributions of this work is using the neural MT approach for the Chinese-Spanish language pair. In the last years, there has appeared more and more resources for this language pair available in [Ziemski et al., 2016a] or from TAUS

Table 1 Corpus Statistics. Number of sentences (S),words (W), vocabulary (V). M stands for millions and K stands for thousands.

L	Set		S	W	V
ES	Train		3.0M	51.7M	207.5K
	Development		990	43.4K	5.4k
	Test	UN	1K	44.2K	5.5K
		BTEC	729	4.7K	888
ZH Words	Train		3.0M	43.9M	373.5K
	Development		990	33K	3.7K
	Test	UN	1K	33.7K	3.8K
		BTEC	729	4.1K	737
ZH Chars	Train		3.0M	71.1M	43.7K
	Development		990	53.9K	1.7K
	Test	UN	1K	55.1K	1.7K
		BTEC	729	5.5K	668

corporation[6]. Therefore, differently from previous works on this pair of languages, we can test our approach in a large set.

The large training corpus is composed of the UN corpus, the TAUS corpus, the Bible corpus [Chew et al., 2006] and the BTEC (Basic Traveller Expressions Corpus) [Takezawa, 2006]. The TAUS corpus is around 2,890,000 sentences, the Bible corpus about 30,000 sentences and the BTEC corpus about 20,000 sentences.

Corpus statistics are shown in Table 1. Statistics for Chinese are shown both with word and character segmentations. In the case of word segmentation, the size of the vocabulary is similar to the target vocabulary, while in the case of using Chinese characters, we have a much lower vocabulary. Development is taken from UN corpus. We use two different test sets corresponding to two different corpus: UN and BTEC.

Corpus preprocessing consisted in tokenization, filtering empty sentences and longer than 50 words, Chinese segmentation by means of the ZhSeg [Dyer, 2016], Spanish lowercasing, filtering pairs of sentences with more than 10% of non-Chinese characters in the Chinese side and more than 10% of non-Spanish characters in the Spanish side.

5.2 System Parameters

The neural MT system was built using the software available in github[7]. We used the following settings. Regarding vocabulary limitation, we used a vocabulary size of 90,000 in Spanish and in Chinese when using words, and we reduced the Chinese vocabulary to 38,000 when using characters. We replaced out-of-vocabulary words (UNKs) using the standard methodology [Jean et al., 2015]: we used the word-to-word translation model learned with 'fast-align' [Dyer et al., 2013] or, if not available, the aligned source word (given the different alphabets we only used the source word when it was not a Chinese word or character).

[6] http://www.taus.net

[7] http://github.com/nyu-dl/dl4mt-tutorial/

(Font: 金文大篆体.ttf, Jin_Wen_Da_Zhuan_Ti.7z)

```
import cairo

# adapted from
# http://heuristically.wordpress.com/2011/01/31/pycairo-hello-world/

# setup a place to draw
surface = cairo.ImageSurface(cairo.FORMAT_ARGB32, 100, 100)
ctx = cairo.Context (surface)

# paint background
ctx.set_source_rgb(1, 1, 1)
ctx.rectangle(0, 0, 100, 100)
ctx.fill()

# draw text
ctx.select_font_face('金文大篆体')
ctx.set_font_size(80)
ctx.move_to(12,80)
ctx.set_source_rgb(0, 0, 0)
ctx.show_text('我')

# finish up
ctx.stroke() # commit to surface
surface.write_to_png('我.gif')
```

Fig. 2 Example of Generation of Bitmap.

Networks have an embedding of 510 when using words and of 529 when using characters. These embeddings are inspired on previous work [Bahdanau et al., 2015] and adapted to sizes from bitmap fonts: 10×51 in the case of words (rectangular bitmap font) and 23×23 in the case of characters (square bitmap font). Other sizes (larger and smaller) from bitmap fonts were experimented with worse results than previous sizes. Embeddings are kept the same for both corpus sizes. To transform each element from the dictionary into a bitmap font, we installed a special package, which is used by the python library *cairo*, see Figure 2 for an example. Note that the generation of bitmaps is hybrid, meaning that when there is no bitmap font available (i.e. it happens for special symbols), we use a random vector. At the end, we use random vectors for less than 5% of the words.

Additionally, we have built a standard phrase-based MT system to contrast results. The phrase-based MT system has been build using standard *Moses* [Koehn et al., 2007] trained using default parameters, which include: grow-diag-final 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 optimisation.

6 Automatic Evaluation

Table 2 shows the results in terms of BLEU [Papineni et al., 2002] and METEOR [Lavie and Denkowski, 2009]. Results show that using bitmap fonts as initialization is much better than using a random initialization, since much more information is provided to the neural system. The improvement holds for both when using single Chinese characters as minimal translation units or using Chinese words.

Table 2 BLEU and METEOR results. In bold, best neural MT results.

System	UN		BTEC	
	BLEU	METEOR	BLEU	METEOR
Characters	17.88	36.29	3.91	18.66
Characters +Bitmap	20.50	39.60	7.34	23.36
Words	21.79	41.60	5.55	22.10
Words +Bitmap	**27.48**	**46.48**	**12.13**	**29.62**
Phrase-based	40.14	56.98	24.44	40.63

The improvement is large for both character bitmap fonts: +2.6 BLEU points, for the UN test, and +3.4 BLEU points for the BTEC test set; and word bitmap fonts: almost +5.7 BLEU points for both test sets). In any case, it is observed that it is better to use words than characters as translation units. METEOR results are always coherent with BLEU results.

Although, results for this particular dataset and language pair do not reach the phrase-based MT system, the goal of the paper is to show that the integration of bitmap fonts enhances the neural MT system. Our approach is helping neural MT towards achieving state-of-the-art results confirming its promising results [8].

7 Human Evaluation and Translation Output Analysis

We compared using human evaluation the two best systems on the large dataset and only for the UN test corpus: the baseline system with words as minimal translation units (21.79 BLEU) and the same system with the bitmap fonts initialization (27.48 BLEU).

Figure 3 shows the framework on which we did the human evaluation. We asked four independent evaluators to compare 97 random sentences from the test set mentioning which translation output was better. Ties were allowed.

Results for each of the four evaluators are shown on Table 3. Note that all evaluators ranked the enhanced system with bitmap fonts as better. In average, almost 51% of the times the system with the bitmap fonts integrated was better than the baseline, 17% of the times was equal, and only 32% of the time it was worst. In order to confirm the quality of the performed human evaluation, we computed the inter-annotation agreement using Fleiss Kappa [Fleiss, 1971] and resulted into 0.21, which is considered fair agreement.

We did some manual analysis to see what kind of errors the integration of bitmap fonts solved in the translation. Table 4 shows some examples of the kind of improvements that the neural MT system with the new initialization is capable of. Examples

[8] https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

We show a reference translation and two corresponding translation outputs. For each one, please answer:
* 1 if ***option 1*** is the best quality translation
* 2 if ***option 2*** is the best quality translation
* 3 if both are of the same quality

Line 0
Reference:
60 / 68 . respuesta a las repercusiones negativas humanitarias y para el desarrollo de la fabricación , transferencia y circulación ilícitas de armas pequeñas y armas ligeras y su acumulación excesiva
Option 1:
60 / 68 . efectos negativos de la transferencia ilícita de armas pequeñas y armas ligeras , la transferencia y la producción y el desarrollo
Option 2:
60 / 68 . respuesta a los efectos negativos en la fabricación y el desarrollo ilícitos de armas pequeñas y armas ligeras en la fabricación de armas pequeñas y armas ligeras

Fig. 3 Human Evaluation Framework.

Table 3 Human evaluation results on 97 random sentences for 4 independent evaluators and the corresponding average. In bold, best results.

Evaluator	Words	Words +Bitmap	Equal
1	30	**52**	15
2	26	**45**	26
3	40	**45**	12
4	28	**55**	14
Average (%)	31.96	**50.77**	17.26

show how it improves the adequacy and fluency of the translations in general. Example 1 and 5 show that when using bitmap fonts initialization there are less content words/information missing in the translation and, as a consequence, the translation tends to be more fluent. Example 2 and 4 show a more adequate translation of source words. Example 3, in addition to chosing a more adequate verb (*apruebe*), it also shows less repetition of translated words, which is a tendency of the neural MT system.

8 Conclusions

This paper shows first experiments in using the neural MT approach for the Chinese-Spanish language pair. Basically, taking advantage of the graphical representation of the Chinese alphabet, we use bitmap fonts of the Chinese words (either words themselves or characters) to initialize the neural MT system. Our technique represents Chinese minimal translation units (words or characters) by means of 2-dimensional bitmap which reflects the shape of the written word. Like this, we are converting Chinese units into bitmap fonts. Then we get the vector of bits representing the bitmap

Table 4 Example Sentences from the UN test set. Source (Src), Baseline (Words), Bitmap fonts (+Bitmap), Reference (Ref)

	Type	Sentence
1	Src	并强调报告篇幅的任何缩减均不得影响报告的列报质量或报告内容
	Words	y destacando que los informes no se pueden comparar con la calidad y el informe de los informes
	+Bitmap	y destaca que **ninguna reducción en la escala** de los informes **no afectará** a la presentación
		de la calidad ni el contenido de la presentación de informe
presentación	Ref	y destaca que la reducción en la longitud de los informes no debería afectar a la calidad de la ni al contenido de los informes
2	Src	5 . 鼓励在国家一级在提高妇女地位的国家机构与负责制订 执行和协调老龄政策和方案的政府实体之间建立机构联系；
	Words	5 . alienta a los órganos nacionales a que , en el contexto de los países , las entidades gubernamentales y los programas gubernamentales ,
		así como en las entidades gubernamentales , las entidades gubernamentales y los programas gubernamentales competentes ;
	+Bitmap	5 . alienta a las instituciones nacionales a cooperar con las autoridades nacionales en la formulación , la ejecución y la armonización de las políticas y **los programas relativos al envejecimiento** ;
	Ref	5 . alienta el establecimiento de vínculos institucionales en el plano nacional entre los mecanismos nacionales para el adelanto de la mujer y las entidades gubernamentales responsables de desarrollar , ejecutar y coordinar políticas y programas relacionados con el envejecimiento ;
3	Src	吁请国会通过关于国家赔偿方案的法律草案
	Words	exhorta a la asamblea general a presentar un proyecto de administración de los programas nacionales sobre los programas nacionales
	+Bitmap	exhorta al parlamento **a que apruebe el proyecto de ley sobre el programa de las naciones unidas**
	Ref	exhorta al congreso a que apruebe el proyecto de ley sobre el programa nacional de resarcimiento
4	Src	各缔约国均应当考虑在同专家协商的情况下
	Words	los estados partes tomarán la posibilidad de examinar
	+Bitmap	cada estado parte considerará la posibilidad de **analizar**
	Ref	cada estado parte considerará la posibilidad de analizar
5	Src	确认小武器和轻武器的泛滥对索马里人道主义局势和发展造成了负面影响
	Words	reconociendo el aumento negativo de las armas pequeñas y armas ligeras en las armas pequeñas y en las armas convencionales
	+Bitmap	reconociendo **los efectos negativos de la proliferación de armas pequeñas y armas ligeras en la situación humanitaria y en el desarrollo en somalia**
	Ref	reconociendo los efectos negativos de la proliferación de armas pequeñas y armas ligeras para la situación humanitaria y el desarrollo de somalia

fonts. This information improves the standard initialization of the word vectors (embeddings).

This extension to the standard neural MT system presented by [Bahdanau et al., 2015] shows almost 6 point BLEU improvement. Human evaluation shows that our system is in average 51% better than the baseline, 17% equal to it and only 31% worse than it.

As further work, we plan to use only the semantic radical of the logogram to initialize embeddings; use the character-based neural MT system [Luong and Manning, 2016, Costa-jussà and Fonollosa, 2016] and use larger datasets, recently published [Ziemski et al., 2016b].

References

[Aldón, 2016] Aldón, D. (2016). *Sistema de Traducción Neuronal Usando Bitmaps*. B.s. thesis, Universitat Politècnica de Catalunya.

[Aldón et al., 2016] Aldón, D., Costa-jussà, M. R., and Fonollosa, J. A. R. (2016). Neural machine translation using bitmap fonts. In *EAMT Workshop on Hybrid Approaches to Translation (HyTRA)*.

[Bahdanau et al., 2015] Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473.

[Banchs et al., 2006] Banchs, R., Crego, J. M., Lambert, P., and Mariño, J. B. (2006). A feasibility study for chinese-spanish statistical machine translation. In *in Procedings of the 5th International Symposium on Chinese Spoken Language Processing (ISCSLP)*.

[Castaño and Casacuberta, 1997] Castaño, M. A. and Casacuberta, F. (1997). A connectionist approach to mt. In *Proc. of the EUROSPEECH Conference*.

[Centelles et al., 2014] Centelles, J., Costa-jussà, M. R., and Banchs, R. E. (2014). Chispa on the go: A mobile chinese-spanish translation service for travellers in trouble. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 33–36.

[Chew et al., 2006] Chew, P. A., Verzi, S. J., Bauer, T. L., and McClain, J. T. (2006). Evaluation Of The Bible As A Resource For Cross-language Information Retrieval. In *Proceedings of the Workshop on Multilingual Language Resources and Interoperability*, pages 68–74.

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

[Cho et al., 2014a] Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014a). On the properties of neural machine translation: Encoder–decoder approaches. In *CoRR*.

[Cho et al., 2014b] Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014b). 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.

[Costa-jussà and Centelles, 2016] Costa-jussà, M. R. and Centelles, J. (2016). Description of the chinese-to-spanish rule-based machine translation system developed using a hybrid combination of human annotation and statistical techniques. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 15.

[Costa-jussà and Fonollosa, 2016] Costa-jussà, M. R. and Fonollosa, J. A. R. (2016). Character-based neural machine translation. In *Proc. of the ACL*.

[Costa-jussà et al., 2012] Costa-jussà, M. R., Q., C. A. H., and Banchs, R. E. (2012). Evaluating indirect strategies for chinese-spanish statistical machine translation. *Journal Of Artificial Intelligence Research*, 45:762–780.

[Dyer, 2016] Dyer, C. (2016). http://code.google.com/p/zhseg/.

[Dyer et al., 2013] Dyer, C., Chahuneau, V., and Smith, N. A. (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. Association for Computational Linguistics.

[Elliott et al., 2015] Elliott, D., Frank, S., and Hasler, E. (2015). Multi-language image description with neural sequence models. *CoRR*, abs/1510.04709.

[Firat et al., 2017] Firat, O., Cho, K., Sankaran, B., Vural, F. T., and Bengio, Y. (2017). Multi-Way, Multilingual Neural Machine Translation. *Accepted for publication in Computer Speech and Language, Special Issue in Deep learning for Machine Translation*.

[Fleiss, 1971] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulleting*, 76(5):378–382.

[Forcada and Ñeco, 1997] Forcada, M. L. and Ñeco, R. P. (1997). Recursive hetero-associative memories for translation. In *Proceedings of the International Work-Conference on Artificial and Natural Neural Networks: Biological and Artificial Computation: From Neuroscience to Technology*, IWANN '97, pages 453–462, London, UK, UK. Springer-Verlag.

[Hitschler et al., 2016] Hitschler, J., Schamoni, S., and Riezler, S. (2016). Multimodal pivots for image caption translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2399–2409, Berlin, Germany. Association for Computational Linguistics.

[Jean et al., 2015] Jean, S., Cho, K., Memisevic, R., and Bengio, Y. (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. Association for Computational Linguistics.

[Kalchbrenner and Blunsom, 2013] Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1700–1709, Seattle, Washington, USA. Association for Computational Linguistics.

[Koehn et al., 2007] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Koehn et al., 2003] Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical Phrase-Based Translation. In *Proc. of the ACL*.

[Lavie and Denkowski, 2009] Lavie, A. and Denkowski, M. J. (2009). The meteor metric for automatic evaluation of machine translation. *Machine Translation*, 23(2-3):105–115.

[Liu et al., 2016] Liu, M., Rus, V., Liao, Q., and Liu, L. (2016). Encoding and Ranking Similar Chinese Characters. Tech report, accessed online 11/2016, Chongqing University.

[Luong and Manning, 2016] Luong, M.-T. and Manning, C. D. (2016). Achieving open vocabulary neural machine translation with hybrid word-character models. In *Association for Computational Linguistics (ACL)*, Berlin, Germany.

[Luong et al., 2015] 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, Lisbon, Portugal. Association for Computational Linguistics.

[Papineni et al., 2002] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (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.

[Paul, 2008] Paul, M. (2008). Overview of the iwslt 2008 evaluation campaign. In *Proc. of the International Workshop on Spoken Language Translation*, pages 1–17, Hawaii, USA.

[Sutskever et al., 2014] Sutskever, I., Vinyals, O., and Le, Q. V. (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.

[Takezawa, 2006] Takezawa, T. (2006). Multilingual spoken language corpus development for communication research. In *Chinese Spoken Language Processing, 5th International Symposium, ISCSLP 2006, Singapore, December 13-16, 2006, Proceedings*, pages 781–791.

[Ziemski et al., 2016a] Ziemski, M., Junczys-Dowmunt, M., and Pouliquen, B. (2016a). The united nations parallel corpus v1.0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

[Ziemski et al., 2016b] Ziemski, M., Junczys-Dowmunt, M., and Pouliquen, B. (2016b). The united nations parallel corpus v1.0. In Chair), N. C. C., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).