# Detecting and Translating Dropped Pronouns in Neural Machine Translation

Xin Tan[1], Shaohui Kuang[2], and Deyi Xiong[1](✉)

[1] School of Computer Science and Technology, Soochow University, Suzhou, China
xtan@stu.suda.edu.cn, dyxiong@suda.edu.cn
[2] Machine Intelligence Technology Lab, Alibaba Group, Hangzhou, China
shaohuikuang@foxmail.com

**Abstract.** Pronouns are commonly omitted in Chinese as well as other pro-drop languages, which causes a significant challenge to neural machine translation (NMT) between pro-drop and non-pro-drop languages. In this work, we propose a method to both automatically detect the dropped pronouns (DPs) and recover their translation equivalences rather than their original forms in source sentences. The detection and recovery are simultaneously performed as a sequence labeling task on source sentences. The recovered translation equivalences of DPs are incorporated into NMT as external lexical knowledge via a tagging mechanism. Experimental results on a large-scale Chinese-English dialogue translation corpus demonstrate that the proposed method is able to achieve a significant improvement over a strong baseline and is better than the method of recovering the original forms of DPs.

**Keywords:** Neural machine translation · Dropped pronouns · Tagging mechanism.

## 1 Introduction

In languages like Chinese and Japanese, there is a habitual phenomenon where if the pronouns are possible to be inferred from the surrounding context or dialog, most pronouns will be omitted to make sentences brief and clear. Such languages are known as pro-drop languages. Although the omissions of these pronouns are generally not problematic for human, they are very challenging for machine, especially when a machine translation system is used to translate dialogue and conversation text from pro-drop languages to non-pro-drop languages. This is illustrated by the examples shown in Fig. 1.

According to our statistics on a large Chinese-English dialogue corpus, about 26% pronouns in Chinese are omitted. And around 72% of them cannot be recovered and correctly translated by our strong NMT [1] baseline system. The failure in translating these omitted pronouns will seriously degrade the fluency and readability of translations in non-pro-drop languages (e.g., English). Translating these DPs is different from translating other words which are already in source sentences. We need to first infer omitted pronouns in the source language according to the context and discourse of them.

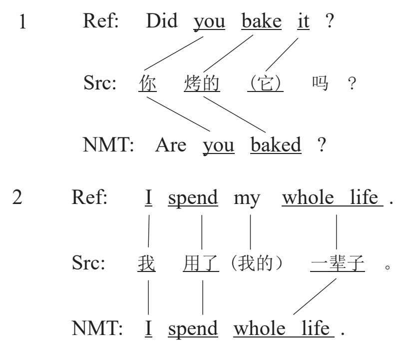


**Fig. 1.** Examples of dropped pronouns in Chinese-English translation.

A variety of efforts have been made for DPs translation in the context of statistical machine translation. These efforts recover the omitted pronouns in the source language either manually or automatically. The automatic methods normally use a small-scale source-language dataset, where DPs are manually recovered and annotated, as the training corpus to construct a DPs recovery model. There are three issues with this source-side DPs annotation method. First, it is time-consuming to build such an annotation corpus. Second, as the size of the manually built DPs annotation corpus is normally not big due to the cost, the accuracy of the DPs recovery model trained on this corpus is normally not high and the model is not easy to be adapted to different domains. Finally, the recovered DPs in pro-drop languages may be ambiguous for being translated into non-pro-drop languages. For example, if we translate from Chinese to English, '我' in Chinese can be corresponding to both 'me' and 'I' in English, but only one is suitable for specific sentence components.

In order to handle the issues mentioned above and inspired by Wang *et al.* [13], we propose a new approach to automatically recover and translate DPs in the source language. Instead of recovering DPs in their original forms, we automatically recover their translations in appropriate positions in source sentences. On a large-scale word-aligned bilingual training corpus, we can easily detect the translations of source-side DPs in the target non-pro-drop language. These translations can be further aligned to placeholders in the source language where the omitted pronouns should be inserted. In this way, we can recover DPs' translations in the source language. This will allow us to train a new DPs recovery model that recovers the target counterparts of DPs, rather than themselves. We refer to this model as the DPs equivalence recovery model (DP_ERM). Since the equivalence recovery procedure can be automatically performed on word-aligned bilingual corpus, we can easily obtain a large-scale corpus to train our model which can be cast as a sequence labeling model. The manual annotation of DPs is completely not necessary in our approach. Source sentences with recovered DPs translations are then feed into an NMT model. Since source sentences are now mixed with both the source and target language, we treat the translations of DPs as external lexical knowledge, which is then incorporated into the NMT model via a tagging mechanism.

We examine the effect of the proposed method on Chinese-English translation task. Experimental results on large-scale subtitle corpora show that our approach can significantly improve the translation performance in terms of translating DPs. Furthermore, the proposed DPs equivalence recovery approach is better than the conventional DPs recovery in NMT. Interestingly, the better the recovered translations of DPs, the larger the performance gap between the proposed approach and the conventional method.

## 2 Related Work

One line of work that is closely related to the dropped pronoun resolution is zero pronoun resolution (ZR) which is a sub-direction of co-reference resolution (CR). The difference between these DPs and ZR tasks is that ZR contains three steps (namely zero pronoun detection, anaphoricity determination and co-reference linking) whereas the dropped pronoun resolution task only contains detection and recovery. Some studies use the ZR approaches to address the dropped pronoun resolution by using a rule-based procedure (based on full constituency parses) to identify DPs slots and candidate antecedents. Zhao and Ng [17] develop such a method that uses a decision tree classifier to assign DPs to antecedents. Furthermore, Yang *et al.* [15] employ a similar approach, where they use a more sophisticated rule-based approach (based on verbal logic valence theory) to identify dropped pronoun slots. Chen and Ng [4] propose an SVM classifier with 32 features including lexical, syntactical rules to detect DPs.

Another line that is related to dropped pronoun resolution is Empty Category (EC) [3] detection and resolution as DPs can be considered as one type of empty categories. EC resolution aims to recover long-distance dependencies and certain dropped elements [14]. Kong and Zhou [6] follow the idea of EC resolution to develop a method that recursively applies a "linear tagger" to tag each word with a single empty category or none so as to tackle the dropped pronoun problem.

Both zero pronoun and empty category based resolutions have made great progress. However, more and more recent efforts pay attention to DPs and treat the dropped pronoun resolution as an independent task. Taira *et al.* [11] try to improve Japanese-English translation by inserting DPs into input sentences via simple rule-based methods. Yang *et al.* [16] first propose to recover DPs in Chinese text message. They train a 17-class maximum entropy classifier to assign words to one of 16 types of DPs or "none". Each assigned label indicates whether a corresponding dropped pronoun is preceding the word. Their classifier explores lexical, part-of-speech tags, and parse-based features. Wang *et al.* [13] propose to label DPs with parallel training data. All these efforts have improved translation quality by recovering DPs. Our work is significantly different from them in that we recover the translation equivalences of DPs rather than their original forms in the source language. This allows to avoid the translation ambiguities where a source pronoun can be translated differently.

## 3 Background: Attention-Based NMT Architecture

The attention-based NMT is based on an RNN Encoder-Decoder architecture. It contains two components: one is an encoder part and the other is a decoder one. Here, we briefly describe the whole framework.

For the encoder part, an encoder first reads a sequence of vectors $X = (x_1, x_2, ..., x_T)$ which represents a sentence and among it, $X$ is the input sentence that we want to translate, $x_j$ is the $j^{th}$ word embedding in the sentence. Given an input $x_t$ and the previous hidden state $h_{t-1}$, the RNN encoder can be formulated as follows:

$$h_t = f(x_t, h_{t-1}) \tag{1}$$

$$c_t = \sum_{j=1}^{T_x} \alpha_{tj} h_j \tag{2}$$

$$\alpha_{tj} = \frac{exp(e_{tj})}{\sum_{k=1}^{T_x} exp(e_{tk})} \tag{3}$$

$$e_{tj} = a(s_{t-1}, h_j) \tag{4}$$

where, $c_t$ is the context vector, $\alpha_{tj}$ is the weight of $h_j$ computed by considering its relevance to the predicted target word, and $e_{tj}$ is an alignment model.

As for the decoder, it consists of another RNN network. Given the context vector $c_t$ calculated from the encoder and all the previously predicted target words $\{y_0, y_1, ..., y_{t-1}\}$, the target translation $Y$ can be predicted by

$$P(Y) = \prod_{t=1}^{T} p(y_t | \{y_0, y_1, y_2, ..., y_{t-1}\}, c_t) \tag{5}$$

where $Y = (y_0, y_1, y_2, ..., y_T)$.

The probability for predicting each target word is computed as follows:

$$p(y_t | \{y_0, y_1, y_2, ..., y_{t-1}\}, c_t) = g(y_{t-1}, s_t, c_t) \tag{6}$$

among which, $g$ often uses a softmax function to compute and $s_t$ is the hidden state of the decoder RNN which is computed by $s_t = f(y_{t-1}, s_{t-1}, c_t)$.

## 4 DPs Equivalence Recovery Model

In this section, we describe the DPs equivalence recovery model (DP_ERM) in detail. We also introduce the training and inference process of the DP_ERM.

### 4.1 The Model

The detection of DPs and the recovery of their translation equivalences can be considered as a sequence labeling task. The translation equivalences of all DPs are in a finite set, which are to be predicted in DP_ERM. Following Lample *et al.* [7], we use the combination of a bidirectional long short-term memory model (BiLSTM) and a conditional random field model (CRF) to deal with the DPs sequence labeling task, which we refer to as BiLSTM-CRF model. The architecture of the combined model is shown in Fig. 2.

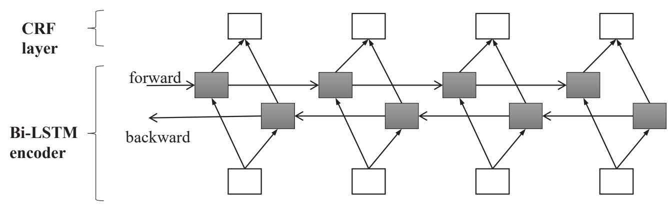


**Fig. 2.** The BiLSTM-CRF model.

On the one hand, the BiLSTM model is able to capture the left and right contextual information for each word through the forward and backward LSTM RNN. On the other hand, the CRF model is capable of exploring arbitrary features that capture relations between labels in neighborhoods making joint and globally optimal decisions instead of independent decisions on each individual position. The combination of BiLSTM and CRF enables DP_ERM to preserve these two strengths for recovering DPs equivalences. Similar to Lample *et al.* [7], the BiLSTM layer obtains the preliminary results $P_{i,j}$, which corresponds to $j^{th}$ tag score of the $i^{th}$ word in a sentence. The CRF network is used as the second layer and utilizing the features extracted by BiLSTM layer to perform the sentence level tagging. The parameter of the CRF layer is a matrix $A$, where $A_{i,j}$ is the score of a transition from the tag $i$ to the tag $j$. Given a sentence $X$, if $y = (y_0, y_1, y_2, ..., y_n)$ is the label sequence of the sentence $X$, then the score of the label sequence $y$ is computed as follows:

$$s(X, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \tag{7}$$

where $y_0$ and $y_n$ correspond to *start* and *end* tags of a sentence separately. Finally, a softmax function is used to determine the probability of the label sequence $y$, which is defined as follows:

$$p(y|X) = \frac{e^{s(X,y)}}{\sum e^{s(X,y)}} \tag{8}$$

.

### 4.2 Training and Inference Process of the DP_ERM

In order to train DP_ERM, we need to obtain a training corpus where each translation equivalence is recovered in source sentences. Given a parallel corpus, we first use Giza++ [8] to get a word alignment between each source and target sentence. With word alignments, we can easily detect which pronouns on the target side are aligned to null on the source side. These null-aligned target pronouns are candidates of translation equivalences for DPs on the source side. Next, we detect the exact positions of these DPs in source sentences via the null-aligned target pronouns, we find that it is possible to first detect an approximate position for a DP in the source sentence. If the target words before and after an unaligned target pronoun are aligned to source words, we consider the approximate position of the DPs corresponding to the unaligned target pronoun in-between the source words that are aligned to the target words proceeding and succeeding the unaligned target pronoun, just like examples shown in Fig. 3.

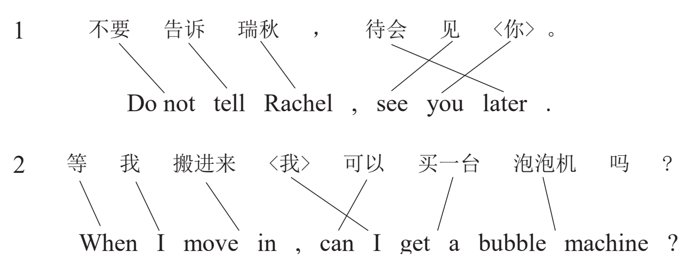


**Fig. 3.** Examples of word alignments between DPs and their translation equivalences.

After finding the possible positions of DPs, we put all source pronouns corresponding to those unaligned target pronouns into every possible position separately. In this way, we generate multiple source sentences with recovered DPs, which all correspond to the same source sentence with pronouns omitted. We then employ an n-gram language model (LM) [2] which pre-trained on a large-scale source corpus to score these candidate positions and select the lowest perplexity one as the final sequence to insert the translation equivalences of DPs. After that, we use the processed training data to train the DP_ERM.

For inference process, we train a BiLSTM-CRF model [5] on the corpus created above and use the pre-trained model to recover translation equivalence of each dropped pronoun for each source sentence of the test data. We regard the DPs translation recovery on the test data as a sequence labeling problem where labels are pronoun translations. There are 32 labels (i.e., none, I, me, you, he, him, she, her, it, we, us, they, them, my, mine, your, yours, his, hers, its, our, ours, their, theirs, myself, yourself, himself, herself, itself, ourselves, yourselves, and themselves) in total.

## 5 Translating Source Sentences with Translation Equivalences of DPs

We use a tagging mechanism to translate source sentences with annotated target pronouns in NMT. The tagging mechanism requires change neither in the NMT network architecture nor in the decoding algorithm. We just add two markers “<tag>” and “</tag>” to the beginning and the end of each DP equivalence (DPE) automatically annotated on the source side by the pre-trained BiLSTM-CRF model. Similarly, we add these markers to each DPE on the target side accordingly. By using such tagged instances in training data, we suspect that NMT model can automatically learn translation patterns triggered by these tags. Once the markers appear, NMT model considers that a special zone begins and copy the special zone into target translation surrounding by “<tag>” and “</tag>” according to the learned patterns. The tagging mechanism we introduce is illustrated in Fig. 4.

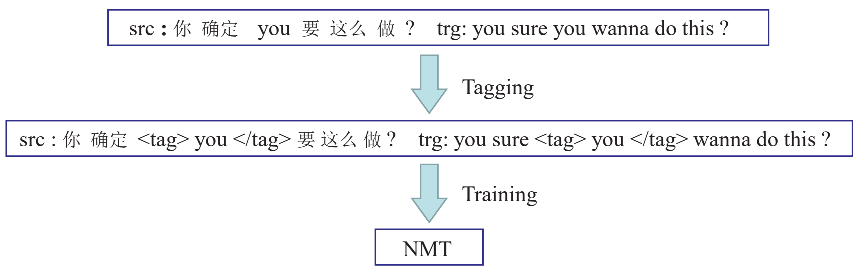


**Fig. 4.** NMT training process with the tagging mechanism.

## 6 Experiments

Experiments were conducted to evaluate the proposed method on a large-scale Chinese-English dialog corpus [12] with more than two million sentence pairs (movie or TV episode subtitles). The detailed statistics of data are listed in Table 1.

In order to obtain good word alignments, we ran Giza++ [8] on the created training data together with another larger parallel subtitle corpora[3]. Furthermore, we pre-trained a tri-gram language model using SRI Language Toolkit [10]. Also, we used the FoolNLTK Toolkit[4] to train the BiLSTM-CRF sequence labeling model on the training corpus. We then used the pre-trained model to assign DPEs to proper positions of source sentences. Almost 90% of DPs were recovered thanks to the alignment information of parallel training corpus.

We used the FoolNLTK Toolkit which contains the BiLSTM and CRF model as mentioned before to train the DP_ERM on the training corpus as shown in Table 1. We then used the pre-trained DP_ERM to assign 32 labels (as mentioned in Section 4.2) to each position on the source sentences of both the development(tune) set and test set, as shown in Table 1.

[3] The data were obtained from the website http://opus.nlpl.eu/.
[4] An Open-source toolkit at https://github.com/rockyzhengwu/FoolNLTK.

**Table 1.** Statistics of the experimental datasets.

| Data | Number of sentences | Number of Zh prons | Number of En prons |
|---|---|---|---|
| Train | 2.15M | 12.1M | 16.6M |
| Tune | 1.09K | 6.67K | 9.25K |
| Test | 1.15K | 6.71K | 9.49K |

To train the NMT model with the tagging mechanism introduced in Section 5. We limit the vocabularies to most frequent 30K words in both Chinese and English, covering approximately 97.3% and 99.3% of the words in the two languages separately, and then merge the two vocabularies. The maximum length of sentences is set to be longer than 50 for both the source and target side due to the insertion of extra tagging labels <tag> and </tag>, and thus we have the same number of training sentences as for the baseline. Except that, all the settings are the same as those in our baseline model RNNSearch. The dimension of word embedding is 620 and the size of the hidden layer is set to 1000. Mini-batches were shuffled during training process with a mini-batch size of 80. Additionally, during decoding process, we use the beam-search algorithm to optimize the prediction process and the beam size is set to 10.

For end-to-end evaluation, case-insensitive BLEU [9] is used to measure translation performance and manual evaluation is used to measure recovered DP_ERM quality. We evaluate the numbers and corresponding rate for recovered pronouns using the DP_ERM, the most frequent 3 kinds of recovered DPs in the training and test data together with their corresponding distributions are shown in Table 2.

**Table 2.** Percentages of recovered pronouns in the training and test set.

| recovered DPs (training set) | Numbers (ratio) |
|---|---|
| "it" | 85250 (22%) |
| "i" | 85029 (21%) |
| "you" | 78717 (20%) |
| **recovered DPs (test set)** | **Numbers (ratio)** |
| "you" | 48 (47%) |
| "it" | 29 (19%) |
| "i" | 19 (18%) |

Additionally, we also evaluated the accuracy of translating tagged pronouns using NMT with the tagging mechanism in it and find out that the accuracy of translating DPs with the tagging mechanism based NMT is 96.1%.

# 7 Results and Analysis

## 7.1 Overall Results

Table 3 summarizes the results of translation performance on the Chinese-English corpus with DPs. "Baseline" was trained and evaluated on the original training and test data. "+ DPEs_manual" indicates that the system is trained on the training data with DPEs being automatically annotated according to word alignments and tested on the test set with manually annotated DPEs. And the "+ DPs_manual" indicates a system trained and tested with DPs (automatically annotated on the training set and manually annotated on the test set) rather than DPEs. The suffix "_ref" represents a system that is evaluated on the test set annotated with DPs or translation equivalences of DPs according to reference translations. And the suffix "_seqlabel" indicates systems evaluated on the test set annotated with DPs (method of Wang *et al.* [13]) or DPEs via the pre-trained sequence labeling model described in Section 4. It can be clearly observed that the proposed DP_ERM which recovers translation equivalences of DPs and translates DPE-annotated source sentences with the tagging mechanism is significantly better than the method that recovers source DPs rather than DPEs in all cases.
As shown in Table 3, the proposed method which recovers the translation equivalences of DPs and translates the DPE-annotated source sentences with the tagging mechanism can significantly improves the translation quality in all cases over recovering source DPs. However, the baseline only achieves 32.04 in BLEU score on the test set, where there are 3 references per source sentence. From the results, machine translation of dialogue from a pro-drop language to a non-pro-drop language is still a challenge for NMT.

We achieve + 4.18 BLEU points over the baseline if we manually recover source DPs and + 2.98 if we automatically recover them according to reference translations. These improvements go further to + 5.57 and + 3.94 BLEU points if we recover DPEs, about 1 BLEU point higher than those with DPs. If we perform DPE/DPs recovery via the fully automatic sequence labeling method in Section 4, we achieve improvements of + 1.17 and + 0.54 BLEU points. From the results, recovering DPEs is proved to be better than Wang *et al.*'s work [13] in recovering DPs.

## 7.2 Effect of Recovered DPEs

We further conducted three experiments to compare with three different methods. As shown in Table 4, note that "+ DPEs_manual" means annotating DPs with corresponding target equivalences, and "+ DPs_manual" just annotates DPs with their original forms in the source language, and "+ half_manual" means that we recover source DPs first and then manually translate them into the counterparts in the target language. From the results, using "+ half_manual" to recover the DPs can still gain 0.19 BLEU points over the "+ DPs_manual", which indicates the advantage of recovering DPEs over DPs.

**Table 3.** BLEU scores of different DPs recovery methods.

| System | Test (BLEU) | Δ |
|---|---|---|
| Baseline | 32.04 | — |
| + DPEs_manual | 37.61 | + 5.57 |
| + DPs_manual | 36.22 | + 4.18 |
| + DPEs_ref | 35.98 | + 3.94 |
| + DPs_ref | 35.02 | + 2.98 |
| + DPEs_seqlabel | 33.21 | + 1.17 |
| + DPs_seqlabel (Wang *et al.* [13]) | 32.58 | + 0.54 |

**Table 4.** BLEU scores of different methods recovering DPs on training data.

| System | Test (BLEU) | Δ |
|---|---|---|
| + DPEs_manual | 37.61 | + 5.57 |
| + half_manual | 36.41 | + 4.37 |
| + DPs_manual | 36.22 | + 4.18 |

### 7.3 Analysis on the DPEs Labeling Accuracy

We compare the labeling accuracy of different methods: manual recovery (MR), automatic recovery according to reference translations (RR) and completely automatic recovery (AR) via BiLSTM-CRF as shown in Table 5. We find that when treating recovery of DPs as sequence labeling problem, we achieve a relatively low F1 score. This suggests that automatically detecting DPEs in appropriate positions is nontrivial and challenging. Our DP_ERM can be further improved if we have better detected DPEs, and we will leave this to our future work.

The precisions and recalls of different DPE recovery methods are listed in Table 5. From the table, MR obtains the highest F1 score of 76%. The RR and AR have no alignment information. Therefore, they obtain a lower precision of 69% and 44% respectively. Furthermore, when treating the recovery of DPs as a sequence labeling problem, it can only recover DPs like ‘I’, ‘me’, ‘it’, ‘you’ and so on due to little information learned from the training process and mismatching with surrounding context. Except for ‘it’, other pronouns seriously depend on the surrounding information of a sentence which is the reason for the low BLEU score of the translation of DPs.

**Table 5.** Precisions and Recalls for different methods of recovering DPs.

| Method | Precision | Recall | F1 |
|---|---|---|---|
| manual recovery (MR) | 69% | 84% | 76% |
| recovery by references (RR) | 80% | 29% | 43% |
| fully automatic recovery (AR) | 44% | 12% | 19% |

### 7.4 Translation Examples

In this section, we present some examples of translating recovered dropped pronouns with our proposed method to show the actual effectiveness of the proposed method. The examples are shown in Table 6. From these examples, we can obviously find that the dropped pronouns in source sentences are successfully detected and recovered with their translation equivalences first and then translated into the target translations by NMT.

**Table 6.** Examples of translations with DPEs recovered by DPERM.

| | |
|---|---|
| **Input** | (你) 想不想听一件奇怪的事? |
| **Ref** | do you want to hear something weird ? |
| **Baseline** | want to hear something weird ? |
| **+ DPEs** | <tag> you </tag> want to hear something weird ? |
| **Input** | 下次 (我们) 见到他们, (我们) 就告诉他们 |
| **Ref** | next time we see them , we 'll just tell them . |
| **Baseline** | see them next time . tell them . |
| **+ DPEs** | next time <tag> we </tag> see them , <tag> we </tag> 'll tell them . |

## 8 Conclusion

In this paper, we present a method to recover DPs for NMT from a pro-drop language to a non-pro-drop language. We train a sequence labeling style detector to automatically detect DPs and recover their translation equivalences rather than themselves. The detector is a BiLSTM-CRF model pre-trained on the training data, where dropped pronoun equivalences are recovered according to word alignments. The pre-trained detector is then used to infer the translation equivalences of DPs on test set. The recovered DPEs are translated into the target language via the tagging mechanism. Experiments on a large-scale Chinese-English dialog corpus show that recovering DPEs in source sentences has made a greater improvement than recovering DPs in the source sentence. In our future work, we plan to further improve the accuracy of recovering translation equivalences of DPs.

**Acknowledgements.** The authors would like to thank three anonymous reviewers for their comments on this paper. This research was supported by National Natural Science Foundation of China (Grant Nos. 61622209 and 61861130364).

## References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)

2. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. Journal of machine learning research **3**(Feb), 1137–1155 (2003)
3. Cai, S., Chiang, D., Goldberg, Y.: Language-independent parsing with empty elements. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. pp. 212–216. Association for Computational Linguistics (2011)
4. Chen, C., Ng, V.: Chinese zero pronoun resolution: Some recent advances. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 1360–1365 (2013)
5. Huang, Z., Xu, W., Yu, K.: Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
6. Kong, F., Zhou, G.: A clause-level hybrid approach to chinese empty element recovery. In: IJCAI. pp. 2113–2119 (2013)
7. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. Proceedings of NAACL 2016 (2016)
8. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. Computational linguistics **29**(1), 19–51 (2003)
9. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th ACL. pp. 311–318. Association for Computational Linguistics (2002)
10. Stolcke, A.: Srilm-an extensible language modeling toolkit. In: Seventh international conference on spoken language processing (2002)
11. Taira, H., Sudoh, K., Nagata, M.: Zero pronoun resolution can improve the quality of je translation. In: Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation. pp. 111–118. Association for Computational Linguistics (2012)
12. Wang, L., Tu, Z., Shi, S., Zhang, T., Graham, Y., Liu, Q.: Translating pro-drop languages with reconstruction models. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. pp. 1–9. AAAI Press, New Orleans, Louisiana, USA (2018)
13. Wang, L., Tu, Z., Zhang, X., Liu, S., Li, H., Way, A., Liu, Q.: A novel and robust approach for pro-drop language translation. Machine Translation **31**(1-2), 65–87 (2017)
14. Xiang, B., Luo, X., Zhou, B.: Enlisting the ghost: Modeling empty categories for machine translation. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 822–831 (2013)
15. Yang, W., Dai, R., Cui, X.: Zero pronoun resolution in chinese using machine learning plus shallow parsing. In: Information and Automation, 2008. ICIA 2008. International Conference on. pp. 905–910. IEEE (2008)
16. Yang, Y., Liu, Y., Xue, N.: Recovering dropped pronouns from chinese text messages. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). vol. 2, pp. 309–313 (2015)
17. Zhao, S., Ng, H.T.: Identification and resolution of chinese zero pronouns: A machine learning approach. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL) (2007)