

Dependency-based Reordering Model for Constituent Pairs in Hierarchical SMT

Arefeh Kazemi[†], Antonio Toral^{*}, Andy Way^{*}, Amirhassan Monadjemi[†],
Mohammadali Nematbakhsh[†]

[†] Department of Computer Engineering, University of Isfahan, Isfahan, Iran

{akazemi, monadjemi, nematbakhsh}@eng.ui.ac.ir

^{*} ADAPT Centre, School of Computing, Dublin City University, Ireland

{atoral, away}@computing.dcu.ie

Abstract

We propose a novel dependency-based reordering model for hierarchical SMT that predicts the translation order of two types of pairs of constituents of the source tree: head-dependent and dependent-dependent. Our model uses the dependency structure of the source sentence to capture the medium- and long-distance reorderings between these pairs of constituents. We describe our reordering model in detail and then apply it to a language pair in which the languages involved follow different word order patterns, English (SVO) and Farsi (free word order being SOV the most frequent pattern). Our model outperforms a baseline (standard hierarchical SMT) by 0.78 BLEU points absolute, statistically significant at $p = 0.01$.

1 Introduction

Reordering is a fundamental problem in machine translation (MT) that significantly affects translation quality, especially between languages with major differences in word order. While a great deal of work has been carried out to address this problem, none of the existing approaches can perform all the required types of reordering operations in a principled manner.

In general, there are four main approaches to address the reordering problem in statistical machine translation (SMT): distortion models, lexical phrase-based models, hierarchical phrase-based models and syntax-based models. Despite the relative success of each of these approaches in improving the overall performance of the SMT systems, they suffer from a number of shortcomings:

- *Inability to capturing long-distance reordering.* Distortion and lexical phrase-based models assign probability only to the adjacent word or phrase pairs, so they can only perform local reordering between adjacent units and fail to capture long distance reordering. This weakness has motivated research on tree-based models, such as the hierarchical phrase-based model (HPB). Although HPB models outperform phrase-based models (PB-SMT) on medium-range reordering, they still perform weakly on handling long distance reordering due to complexity constraints.
- *Sparsity.* Most of the approaches can perform the reordering of common words or phrases, but they usually cannot be generalized to unseen patterns which have the same linguistic structure. For example, if the object follows the verb in the source language and precedes the verb in the target language, we still need to see a particular instance of a verb and an object in the training data to be able to perform reordering between them.
- *Context insensitivity.* Lexical and hierarchical phrase-based models determine the ordering of the phrases based solely on the lexical items in those phrases. However, a phrase might have different orderings in different contexts, so it is essential to include more context in order to capture the reordering behaviour.
- *High complexity.* Compared to the other reordering models, syntax-based models have access to the necessary structural information to perform long-distance reordering. However, due to the complexity of the decoding algorithm, they have very low performance on large-scale translations.

In order to overcome some of these deficiencies,

© 2015 The authors. This article is licensed under a Creative Commons 3.0 licence, no derivative works, attribution, CC-BY-ND.

we propose a dependency-based reordering model for HPB-SMT. Our model uses the dependency structure of the source sentence to capture the medium- and long-distance reorderings between the dependent parts of the sentence. Unlike the syntax-based models that impose harsh syntactic limits on rule extraction and require serious efforts to be optimised (Wang et al., 2010), we use syntactic information only in the reordering model and augment the HPB model with soft dependency constraints. We report experimental results on a large-scale English-to-Farsi translation task.

The rest of this paper is organised as follows. Section 2 reviews the related work and contextualises our work. Section 3 outlines the main reordering issues due to syntactic differences between English and Farsi. Section 4 presents our reordering model, which is then evaluated in Section 5. Finally, Section 6 concludes the paper and outlines avenues of future work.

2 Related Work

Phrase-based systems can perform local (short distance) reordering inside the phrases but they are inherently weak at non-local (medium and long distance) reordering (Birch and Osborne, 2011). Previous work to address reordering in PB-SMT can generally be categorised into two groups. Approaches in the first group perform reordering in a pre-processing step (i.e. before decoding) by applying some reordering rules to the source sentences to make them in order more similar to that of the target language (Xia and McCord, 2004; Collins et al., 2005; Genzel, 2010). Although all these approaches have reported improvements, there is a fundamental problem with separating the reordering task into a pre-processing component as every faulty decision in the pre-processing step will be passed along as a hard decision to the translation system. This also violates the main principle behind statistical modelling in SMT, i.e. to avoid any hard choices and having the ability to reverse early faulty choices.

Approaches in the second group try to handle reordering in the decoding step, as a part of the translation process. They implement a probabilistic reordering model that can be used in combination with the other models in SMT to find the best translation. These approaches range from distortion models (Koehn et al., 2003) to lexical reordering models (Tillmann, 2004).

Distortion models generally prefer monotone translation which, while may work for related languages, is not a realistic assumption for translating between languages with different grammatical structure. On top of this limitation, these models do not take the content into consideration, and thus they do not generalise well.

Lexical reordering models take content into account and condition reordering on actual phrases. They try to learn local orientations for each adjacent phrase from training data. Despite the satisfactory performance of lexical models, they have two important limitations (Birch, 2011). First, since these models are conditioned on actual phrases, they have no ability to be generalised to unseen phrases. Second, these models still fail to capture long- and even medium-distance reorderings, since they try to find suitable reorderings only between adjacent phrases. The first limitation can be alleviated by using features of the phrase pair instead of the phrase itself (Xiong et al., 2006) while the second limitation can be tackled with hierarchical phrase reordering models (Galley and Manning, 2008).

HPB models (Chiang, 2005) should lead to better reordering than PB-SMT models by allowing phrases to contain gaps. In fact, this approach outperforms PB-SMT in medium-distance reordering, but it is equally weak in long-distance reordering (Birch et al., 2009). Common approaches to reordering in HPB models include pre-processing (Xu et al., 2009) and adding syntax to translation rules. The first approach results in improvements but suffers from the same issues presented above for pre-processing reordering in PB-SMT. The second introduces additional complexities and increases data sparsity (Hanneman and Lavie, 2013).

Our work falls into the recent research line that uses an external reordering model in hierarchical SMT. These models use source syntax to improve reordering without having to annotate translation rules with source syntax. Work in this line has so far looked at predicting the translation order of different types of source elements, pairs of words (Huang et al., 2013), constituents such as head and dependent words (Gao et al., 2011) and predicate-argument structures (Xiong et al., 2012; Li et al., 2013). It is worth noting that all these approaches have been applied solely to one language pair so far, Chinese-to-English.

This paper contributes to this research line on two dimensions. First, we extend the work of (Gao et al., 2011), who studied reordering of head-dependent pairs (i.e. parent and child elements in the dependency tree), and consider also the reordering of pairs of dependents (i.e. sibling elements in the dependency tree). Second, this is the first paper in this line of work to be applied to a language pair other than Chinese-to-English. Our language pair, English-to-Farsi, is comparatively challenging because (i) the target language is free word-order and morphologically rich, and (ii) it is comparatively under-resourced.

3 Word Order Differences between English and Farsi

This section provides a brief survey of the word order differences between the two languages of our case study. The main aim of this section is to make the reader familiar with the Farsi language, and specifically, to its word order peculiarities. That said, it should be noted that despite there being works that try to find specific syntactic reordering patterns for specific language pairs, e.g. (Collins et al., 2005), we have not used the syntactic information covered in this section in the proposed model as our model is language-independent.

There are two major differences between the word order in English and Farsi. First, English sentences follow the SVO (subject-verb-object) order while Farsi sentences follow, in most cases, the SOV order (Moghaddam, 2001). Second, English has strict word order while Farsi allows for free word order. In Farsi, the preferred word order is SOV, but all of the other orders are also correct.

Table 1 provides further details on word order differences by determining the element pairs that should be reordered in the translation process. In order to categorise word order differences we use the element pairs presented by Dryer (1992). Dryer has shown that these pairs can be used to distinguish SOV and SVO languages.

4 Dependency-based Reordering Model

Our reordering model is based on the source dependency tree, an example of which is shown in Figure 1. The dependency tree of a sentence shows the grammatical relations between the head and dependent words of that sentence. For example in Figure 1, the arrow from “he” to “bought” with label “nsubj”, expresses that the

dependent word “he” is the subject of the head word “bought”. Under the assumption that constituents move as a whole (Quirk et al., 2005), our proposed reordering model aims to predict the orientation of each dependent word with respect to its head (*head-dependent*), and also with respect to the other dependents of that head (*dependent-dependent* orientation). For example, for the sentence in Figure 1 we try to predict the appropriate orientations between the head-dependent and dependent-dependent pairs shown in Tables 2 and 3, respectively.

Our motivation for using dependency structure as the basis of our reordering model is based on the assumption that, if it is the case that a reordering pattern is employed for one English–Farsi sentence pair with a specific dependency structure, then another sentence pair containing the same dependency structure will follow the same reordering pattern. For example, in translating from English to Farsi, all of the following English sentences have the same word order in Farsi: “he puts the book on the table”, “they put the desk on the ground”, “he put his hand on my shoulder”. In general, almost all the English sentences following the structure “subject” put “object” on “preposition-on” follow the same word order pattern in their Farsi translations.

We generate the dependency parse tree of the source sentence and perform word alignment between the source and target words in the parallel corpus. Having obtained both the source dependency tree and the word alignments, we extract the orientation type (monotone or swap) between each dependent word with respect to its head and the other dependents of that head. With the alignment points (p_{S1}, p_{T1}) and (p_{S2}, p_{T2}) for two source words $S1$ and $S2$ and their aligned target words $T1$ and $T2$, we define orientation types (*ori*) as in Equation 1.

$$ori = \begin{cases} \textit{monotone}, & \\ \textit{if } (p_{S1} - p_{S2}) \times (p_{T1} - p_{T2}) > 0 & \\ \textit{swap}, & \\ \textit{otherwise}. & \end{cases} \quad (1)$$

When a source word is aligned to multiple target words, we only consider the last aligned target word in determining the orientation type. For example, given the alignments for the head word *bought* with alignment point (1, 7) and dependent word *camels* with alignment point (2, 2) in Figure

Element Pairs	Example (English)	Word Order (English)	Word Order (Farsi)
subject, object and verb	Mary gave the book to John	SVO	SOV
noun and genitive	Mary's Book	noun + genitive	genitive + noun
verb and adpositional	He slept on the ground	verb + adp.	adp. + verb
verb and manner adverb	He ran slowly	Verb + m. adverb	m. adverb + verb
copula and predicate	She is a teacher	copula + predicate	predicate + copula
noun and adjective	Green Book	adjective + Noun	Noun+ Adjective
possessive affix and noun	My book	possessive + noun	noun + possessive

Table 1: word order differences between Farsi and English

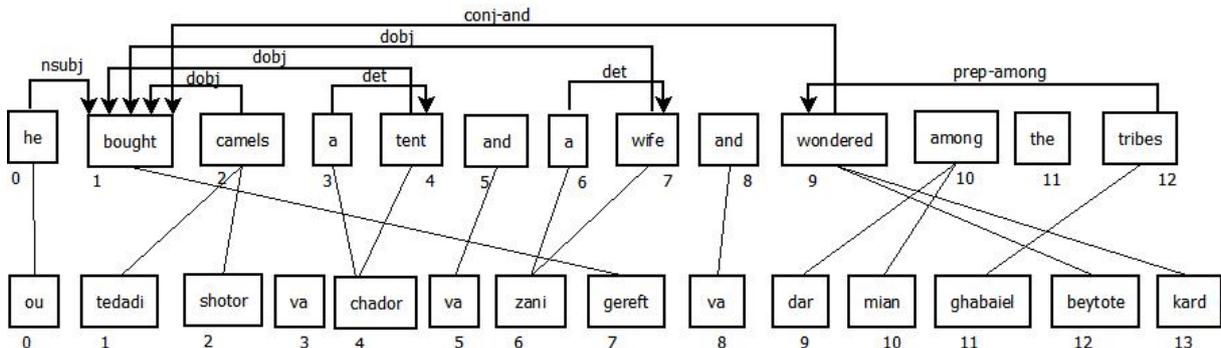


Figure 1: An example dependency tree for an English source sentence, its translation in Farsi and the word alignments

1, we consider swap orientation between *bought* and *camels* based on Equation 1.

After extracting the orientation for all the pairs in the training set, we train a Naive Bayes classifier to estimate the probability of a source dependent word being translated in a monotone or swap manner with respect to its head and the other dependent words of that head.

Making the strong independence assumption that each word is ordered in the sentence independently, the reordering probability for a sentence can be split into the reordering probability of its constitutive (*head,dependent*) and (*dependent,dependent*) pairs. Hence, we define the dependency-based reordering (*DBR*) feature-function score for a translation hypothesis as the sum of the log orientation probabilities for its constitutive pairs as in Equation 2, where H is the translation hypothesis and $Pairs(H)$ is the set of the pair components of H .

$$score_{DBR}(H) = \sum_{pair_i \in Pairs(H)} \log(P_{DBR}(ori|pair_i)) \quad (2)$$

We implemented the reordering model P_{DBR} as a feature-function and combined it with the other feature-functions in the log-linear framework of the HPB model. This feature-function is made of four components: *monotone*, *swap*, *dependency*

coherence and *unaligned pairs*. The components *monotone* and *swap* compute the sum of orientation probabilities of those pairs which are translated in monotone and swap orientation, respectively. *Dependency coherence* counts the number of translated pairs in a hypothesis and encourages concurrent translation of constituents based on the assumption that constituents move together in translation (Quirk et al., 2005). *Unaligned pairs* counts the number of pairs with at least one unaligned source word, as the other three components can not be applied to unaligned pairs.

Various features can be used to reflect the local information of each translation hypothesis H to model $P_{DBR}(orientation|pair_i)$. Finally, we chose the following features to describe the translation hypothesis H for head-dependent pairs:

- The surface forms of the head word $Lex(head)$ and the dependent word $Lex(dep)$
- The dependency relation of the dependent word $depRel(dep)$

we chose following features for dependent-dependent pairs:

- The surface forms of the mutual head word $Lex(head)$, the first dependent word $Lex(dep1)$ and the second dependent word $Lex(dep2)$

head	bought	bought	bought	bought	bought	wife	tent	tribes
dependent	He	camels	tent	Wife	wandered	a	a	wandered

Table 2: head-dependent pairs for the sentence in Figure 1.

dep1	He	He	He	He	camels	camels	camels	tent	tent	wife
dep2	camels	tent	wife	wandered	Tent	wife	wandered	wife	wandered	wandered

Table 3: dependent-dependent pairs for the sentence in Figure 1.

- The dependency relations of the first dependent word $\text{depRel}(\text{dep1})$ and the second dependent word $\text{depRel}(\text{dep2})$

As an example, consider the pair *bought* and *camels* in our example in Figure 1. The model attempts to predict the orientation between these two words as described in Equation 3.

$$P_{DBR}(\text{ori}|\text{lex}(\text{head}), \text{lex}(\text{dep}), (\text{depRel}(\text{dep})))$$

$$\text{ori} \in \{\text{monotone}, \text{swap}\}$$

$$(3)$$

where $\text{lex}(\text{head})=\text{bought}$, $\text{lex}(\text{dep})=\text{camels}$ and $\text{depRel}(\text{dep})=\text{obj}$. The orientation probabilities for *bought* and *camels* (0.21 and 0.79 for monotone and swap, respectively) encourage the swap orientation between them, which supports the required reordering of the object and verb, when translating from English-to-Farsi. Despite the limitations of this model, it can capture the general linguistic reordering patterns that are not available to other reordering models. For instance, it can learn that when translating between SVO and SOV languages, the object and the verb should be reordered, while the subject and the object should be translated in monotone order.

5 Experiments

5.1 Experimental Setup

We used the Mizan parallel English–Farsi corpus¹ (Supreme Council of Information and Communication Technology, 2013) which contains nearly 1 million sentence pairs. This corpus is extracted from English novel books (mostly in their classical literature domain) and their translations in Farsi. 3,000 sentence pairs were held out for development and 1,000 for testing. These sentence pairs were randomly selected from the corpus. The remaining content of the corpus is used for training. Table 4 presents the details about this dataset.

We parsed the source side (English) of the corpus using the Stanford dependency parser (Chen

¹<http://dadegan.ir/catalog/mizan>

	unit	English	Farsi
Train	sentences	1,016,758	1,016,758
	words	13,919,071	14,043,499
Tune	sentences	3,000	3,000
	words	40,831	41,670
Test	sentences	1,000	1,000
	words	13,165	13,444

Table 4: Mizan parallel corpus statistics

and Manning, 2014) and used the “collapsed representation” of the parser output to obtain direct dependencies between the words in the source sentences. We used GIZA++ (Och and Ney, 2003) to align the words in the corpus. Then we extracted 6,391,255 *head–dependent* pairs and 5,247,137 *dependent–dependent* pairs from train dataset and determined the orientation for each pair based on Equation 1.

In order to measure the impact of different features on the accuracy of our reordering model (as will be described in Section 5.2), we used the Naive Bayes classifier with standard settings from the Weka machine learning toolkit (Hall et al., 2009). We trained the classifier separately for *head–dependent* and *dependent–dependent* pairs.

Our baseline MT system was the Moses implementation of HPM model with default settings (Hoang et al., 2009). We used a 5-gram target language model trained on the Farsi side of the training data. In all experiments, the weights of our reordering feature-function and the built-in feature-functions was tuned with MERT (Och, 2003).

5.2 Impact of different features

Since the proposed reordering model has to classify the *head–dependent* and *dependent–dependent* pairs into their correct monotone or swap orientation classes, its task can be seen as a binary classification task. We used the Naive Bayes algorithm to build such an orientation classifier. We then used different

feature sets in each classification experiment to determine their impact on the accuracy of the model.

The features that were examined in this paper are shown in Table 5. All of these features are entirely based on the source sentence and source dependency parse, so we performed dependency parsing and feature extraction as pre-processing steps so as not to slow down the decoding phase.

Surface form of head	$Lex(head)$
Surface form of 1 st dependent	$Lex(dep1)$
Surface form of 2 nd dependent	$Lex(dep2)$
Dep. relation of 1 st dependent	$depRel(dep1)$
Dep. relation of 2 nd dependent	$depRel(dep2)$

Table 5: Features

We used 10-fold cross validation on the train data set (as described in Table 4) to evaluate the classifier. Table 6 shows the performance of the Naive Bayes classifier for monotone and swap orientation for head-dependent (rows *hd*) and dependent-dependent (rows *dd*) pairs. We use two different feature sets, with (rows *ws*) and without (rows *wos*) surface forms. The features used in each classifier are then as follows:

- hd-wos. $depRel(dep)$
- hd-ws. $Lex(head), Lex(dep), depRel(dep)$
- dd-wos. $depRel(dep1), depRel(dep2)$
- dd-ws. $Lex(head), Lex(dep1), Lex(dep2), depRel(dep1), depRel(dep2)$

Features	Accuracy
hd-wos	64.75%
hd-ws	67.37%
dd-wos	70.85%
dd-ws	71.38%

Table 6: Classification results on head-dep and dep-dep relations

For both types of constituent pairs (*hd* and *dd*), the use of surface forms results in a slight improvement (2.62% and 0.53% absolute for *hd* and *dd*, respectively).

As this is the first paper that attempts to model reordering of *dd* pairs (*hd* has been attempted before (Gao et al., 2011)), we are especially inter-

²The mutual head between dependent word 1 and dependent word 2. For example, in Figure 1 *bought* is the mutual head between the dependent words *he* and *camels*

ested in the results for *dd*. The fact that the classification accuracy for *dd* is higher than for *hd* (71.38% vs 67.37%) motivates us to model the reordering of *dd* constituents in MT, for which we present results in the next Section.

5.3 MT Results

We build six MT systems, four according to the constituent pairs and features examined (cf. Table 6) and two additional systems that model the reordering for both types of constituent pairs (rows *all*) with (*ws*) and without (*wos*) surface forms. We compare our systems to two baselines, a standard HPB-SMT system (HPB) and a HPB-SMT system with added swap glue grammar rule (HPB_sgg) as in Equation 4. The swap glue rule allows adjacent phrases to be reversed.

$$X \rightarrow (X_1X_2, X_2X_1) \quad (4)$$

Table 7 shows the results obtained by each of the MT systems according to four automatic evaluation metrics: BLEU, NIST (Doddington, 2002), TER (Snover et al., 2006) and METEOR (Denkowski and Lavie, 2014). For each system and evaluation metric we show its relative improvement over the baseline HPB (columns diff).

The scores obtained by systems that implement our novel reordering between pairs of dependents (columns *dd*) are better than those of the baseline, both with (*ws*) and without (*wos*) surface forms, across all the four evaluation metrics. The same is true for models that implement reordering between both pairs of constituent types (columns *all*), except for the system *all_wos* according to BLEU. The results for systems that perform reordering between pairs of head and dependent offer a mixed picture, with some metrics indicating improvement (e.g. BLEU) and some others deterioration (e.g. TER).

The use of surface forms leads to better results in most cases (except for *hd* systems in terms of NIST, TER and METEOR, and *dd* systems in terms of NIST and TER), confirming the trends shown in the classification experiment, cf. Table 6.

As stated earlier in the paper, Farsi is a free word-order language. When compiling the results of our experiments, we only had a single reference available against which the output from our various systems could be compared. Computing au-

System	BLEU	diff	NIST	diff	TER	diff	METEOR	diff
HPB	0.1083		3.7625		0.8005		0.1683	
HPB_sgg	0.1087	0.37%	3.7299	-0.87%	0.8009	0.05%	0.1665	-1.10%
dd_ws	0.1161 ‡	7.20%	3.8303†	1.80%	0.7937	-0.85%	0.1733	2.96%
dd_wos	0.1097	1.29%	3.8381†	2.01%	0.7929	-0.95%	0.1728	2.65%
hd_ws	0.1095	1.11%	3.6548	-2.86%	0.8155	1.88%	0.1643	-2.39%
hd_wos	0.1095	1.11%	3.7413	-0.56%	0.8061	0.70%	0.1687	0.21%
all_ws	0.1091	0.74%	3.8614 ‡	2.63%	0.7858	-1.84%	0.1727	2.60%
all_wos	0.1054	-2.68%	3.8374†	1.99%	0.7902	-1.29%	0.1708	1.44%

Table 7: Scores of the MT systems according to different automatic metrics. The best score according to each metric is shown in bold. Statistically significant results, calculated with paired bootstrap resampling (Koehn, 2004) for BLEU and NIST, are indicated with symbols ‡ ($p = 0.01$) and † ($p = 0.05$).

omatic evaluation scores when translating into a free word-order language in the single-reference scenario is somewhat arbitrary. The fact that the four evaluation metrics used follow slightly different trends reflects this arbitrariness. We would expect a manual evaluation on a subset of sentences to confirm that the output translations are somewhat better than the automatic evaluation scores suggest.

6 Conclusions

This paper has proposed a dependency-based reordering model for HPB-SMT that predicts the translation order of two types of pairs of constituents of the source tree: dependent-dependent and head-dependent. Our model uses the dependency structure of the source sentence to capture the medium- and long-distance reorderings between these pairs of constituents.

It is worth mentioning that this is the first paper where a dependency-based reordering model is applied to a language pair other than Chinese-to-English. Our language pair, English-to-Farsi, is comparatively challenging because (i) the target language is free-word order and morphologically rich, and (ii) it is comparatively under-resourced.

We have evaluated our model against two baselines: standard HPB-SMT and HPB-SMT with swap glue grammar rules. Our model that reorders pairs of dependents outperforms both baselines (> 0.7 absolute in terms of BLEU), with the improvement being statistically significant ($p = 0.01$ in terms of BLEU).

As for future work, several directions are worth considering. First, the use of features that hold linguistic information, such as part-of-speech tags or semantic classes. Second, an in-depth analysis of the output translations produced by our models to

discern which reordering cases it succeeds at and for which other cases it fails. Third, an improved reordering model based on the findings of the previous line of work.

Acknowledgments

This research is supported by Science Foundation Ireland through the CNGL Programme (Grant 12/CE/I2267) in the ADAPT Centre (www.adaptcentre.ie) at Dublin City University, the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement PIAP-GA-2012-324414 (Abu-MaTran) and by University of Isfahan.

References

- Birch, Alexandra and Miles Osborne. 2011. Reordering Metrics for MT. In *Proceedings of the Association for Computational Linguistics*.
- Birch, Alexandra, Phil Blunsom, and Miles Osborne. 2009. A Quantitative Analysis of Reordering Phenomena. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 197–205, Athens, Greece.
- Birch, Alexandra. 2011. *Reordering Metrics for Statistical Machine Translation*. Ph.D. thesis, University of Edinburgh.
- Chen, Danqi and Christopher D Manning. 2014. A fast and accurate dependency parser using neural networks. In *Empirical Methods in Natural Language Processing (EMNLP)*.
- Chiang, David. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 263–270.
- Collins, Michael, Philipp Koehn, and Ivona Kučerová. 2005. Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 531–540.

- Denkowski, Michael and Alon Lavie. 2014. Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the EACL 2014 Workshop on Statistical Machine Translation*.
- Doddington, George. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the Second International Conference on Human Language Technology Research, HLT '02*, pages 138–145.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language*, 68(1):81–138.
- Galley, Michel and Christopher D. Manning. 2008. A simple and effective hierarchical phrase reordering model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 848–856.
- Gao, Yang, Philipp Koehn, and Alexandra Birch. 2011. Soft dependency constraints for reordering in hierarchical phrase-based translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 857–868.
- Genzel, Dmitriy. 2010. Automatically learning source-side reordering rules for large scale machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 376–384.
- Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.
- Hanneman, Greg and Alon Lavie. 2013. Improving syntax-augmented machine translation by coarsening the label set. In *Proceedings of the North American Chapter of the Association of Computational Linguistics*, pages 288–297.
- Hoang, Hieu, Philipp Koehn, and Adam Lopez. 2009. A unified framework for phrase-based, hierarchical, and syntax-based statistical machine translation. In *Proceedings of the International Workshop on Spoken Language Translation, IWSLT*, pages 152–159.
- Huang, Zhongqiang, Jacob Devlin, and Rabih Zbib. 2013. Factored soft source syntactic constraints for hierarchical machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 556–566.
- Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, pages 48–54.
- Koehn, Philipp. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395.
- Li, Junhui, Philip Resnik, and Hal Daumé III. 2013. Modeling syntactic and semantic structures in hierarchical phrase-based translation. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics*, pages 540–549.
- Moghaddam, Mohammad Dabir. 2001. Word order typology of iranian languages. *Journal of Humanities*, 8(2):17–24.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Och, Franz Josef. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 160–167.
- Quirk, Chris, Arul Menezes, and Colin Cherry. 2005. Dependency treelet translation: Syntactically informed phrasal smt. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 271–279.
- Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and Ralph Weischedel. 2006. A Study of Translation Error Rate with Targeted Human Annotation. In *Proceedings of the Association for Machine Translation in the Americas*.
- Supreme Council of Information and Communication Technology. 2013. Mizan English-Persian Parallel Corpus. Tehran, I.R. Iran.
- Tillmann, Christoph. 2004. A unigram orientation model for statistical machine translation. In *Proceedings of HLT-NAACL 2004: Short Papers*, pages 101–104.
- Wang, Wei, Jonathan May, Kevin Knight, and Daniel Marcu. 2010. Re-structuring, re-labeling, and re-aligning for syntax-based machine translation. *Comput. Linguist.*, 36(2):247–277.
- Xia, Fei and Michael McCord. 2004. Improving a statistical mt system with automatically learned rewrite patterns. In *Proceedings of the 20th International Conference on Computational Linguistics*.
- Xiong, Deyi, Qun Liu, and Shouxun Lin. 2006. Maximum entropy based phrase reordering model for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 521–528.
- Xiong, Deyi, Min Zhang, and Haizhou Li. 2012. Modeling the translation of predicate-argument structure for smt. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, pages 902–911.
- Xu, Peng, Jaeho Kang, Michael Ringgaard, and Franz Och. 2009. Using a dependency parser to improve smt for subject-object-verb languages. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 245–253.