Discriminative Reordering with Chinese Grammatical Relations Features

Pi-Chuan Chang[a], **Huihsin Tseng**[b], **Dan Jurafsky**[a], and **Christopher D. Manning**[a]
[a]Computer Science Department, Stanford University, Stanford, CA 94305
[b]Yahoo! Inc., Santa Clara, CA 95054
{pichuan,jurafsky,manning}@stanford.edu, huihui@yahoo-inc.com

Abstract

The prevalence in Chinese of grammatical structures that translate into English in different word orders is an important cause of translation difficulty. While previous work has used phrase-structure parses to deal with such ordering problems, we introduce a richer set of Chinese grammatical relations that describes more semantically abstract relations between words. Using these Chinese grammatical relations, we improve a phrase orientation classifier (introduced by Zens and Ney (2006)) that decides the ordering of two phrases when translated into English by adding path features designed over the Chinese typed dependencies. We then apply the log probability of the phrase orientation classifier as an extra feature in a phrase-based MT system, and get significant BLEU point gains on three test sets: MT02 (+0.59), MT03 (+1.00) and MT05 (+0.77). Our Chinese grammatical relations are also likely to be useful for other NLP tasks.

1 Introduction

Structural differences between Chinese and English are a major factor in the difficulty of machine translation from Chinese to English. The wide variety of such Chinese-English differences include the ordering of head nouns and relative clauses, and the ordering of prepositional phrases and the heads they modify. Previous studies have shown that using syntactic structures from the source side can help MT performance on these constructions. Most of the previous syntactic MT work has used phrase structure parses in various ways, either by doing syntax-directed translation to directly translate parse trees into strings in the target language (Huang et al., 2006), or by using source-side parses to preprocess the source sentences (Wang et al., 2007).

One intuition for using syntax is to capture different Chinese structures that might have the same meaning and hence the same translation in English. But it turns out that phrase structure (and linear order) are not sufficient to capture this meaning relation. Two sentences with the same meaning can have different phrase structures and linear orders. In the example in Figure 1, sentences (*a*) and (*b*) have the same meaning, but different *linear orders* and different *phrase structure parses*. The translation of sentence (*a*) is: "*In the past three years* these municipalities have collectively put together investment in fixed assets in the amount of 12 billion yuan." In sentence (*b*), "in the past three years" has moved its

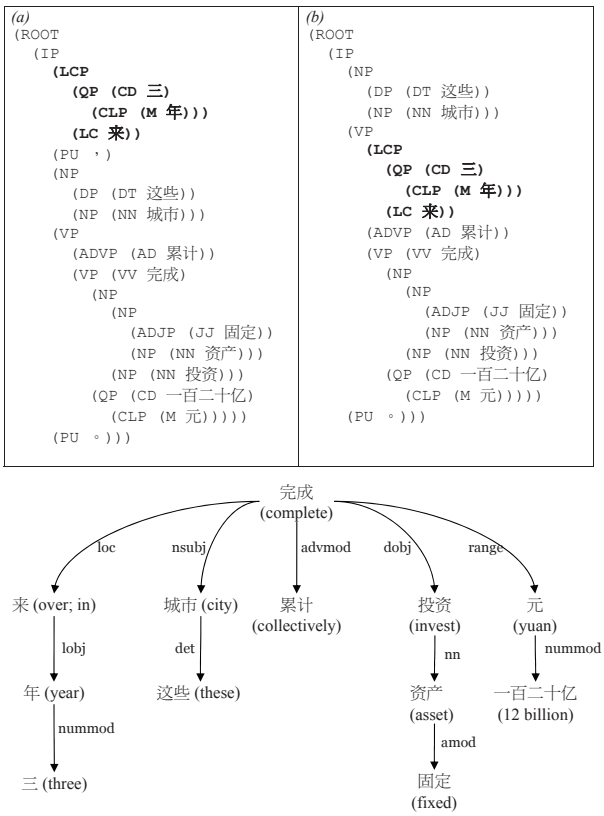


Figure 1: Sentences (*a*) and (*b*) have the same meaning, but different phrase structure parses. Both sentences, however, have the same typed dependencies shown at the bottom of the figure.

Proceedings of SSST-3, Third Workshop on Syntax and Structure in Statistical Translation, pages 51–59,
Boulder, Colorado, June 2009.

position. The temporal adverbial "三 年 来" (in the past three years) has different linear positions in the sentences. The phrase structures are different too: in (*a*) the LCP is immediately under IP while in (*b*) it is under VP.

We propose to use *typed dependency* parses instead of phrase structure parses. Typed dependency parses give information about grammatical relations between words, instead of constituency information. They capture syntactic relations, such as *nsubj* (nominal subject) and *dobj* (direct object) , but also encode semantic information such as in the *loc* (localizer) relation. For the example in Figure 1, if we look at the sentence structure from the typed dependency parse (bottom of Figure 1), "三 年 来" is connected to the main verb 完成 (finish) by a *loc* (localizer) relation, and the structure is the same for sentences (*a*) and (*b*). This suggests that this kind of semantic and syntactic representation could have more benefit than phrase structure parses.

Our Chinese typed dependencies are automatically extracted from phrase structure parses. In English, this kind of typed dependencies has been introduced by de Marneffe and Manning (2008) and de Marneffe et al. (2006). Using typed dependencies, it is easier to read out relations between words, and thus the typed dependencies have been used in meaning extraction tasks.

We design features over the Chinese typed dependencies and use them in a phrase-based MT system when deciding whether one chunk of Chinese words (MT system statistical phrase) should appear before or after another. To achieve this, we train a discriminative phrase orientation classifier following the work by Zens and Ney (2006), and we use the grammatical relations between words as extra features to build the classifier. We then apply the phrase orientation classifier as a feature in a phrase-based MT system to help reordering.

2 Discriminative Reordering Model

Basic reordering models in phrase-based systems use linear distance as the cost for phrase movements (Koehn et al., 2003). The disadvantage of these models is their insensitivity to the content of the words or phrases. More recent work (Tillman, 2004; Och et al., 2004; Koehn et al., 2007) has introduced lexicalized reordering models which estimate reordering probabilities conditioned on the actual phrases. Lexicalized reordering models have brought significant gains over the baseline reordering models, but one concern is that data sparseness can make estimation less reliable. Zens and Ney (2006) proposed a discriminatively trained phrase orientation model and evaluated its performance as a classifier and when plugged into a phrase-based MT system. Their framework allows us to easily add in extra features. Therefore we use it as a testbed to see if we can effectively use features from Chinese typed dependency structures to help reordering in MT.

2.1 Phrase Orientation Classifier

We build up the target language (English) translation from left to right. The phrase orientation classifier predicts the start position of the next phrase in the source sentence. In our work, we use the simplest class definition where we group the start positions into two classes: one class for a position to the left of the previous phrase (*reversed*) and one for a position to the right (*ordered*).

Let $c_{j,j'}$ be the class denoting the movement from source position j to source position j' of the next phrase. The definition is:

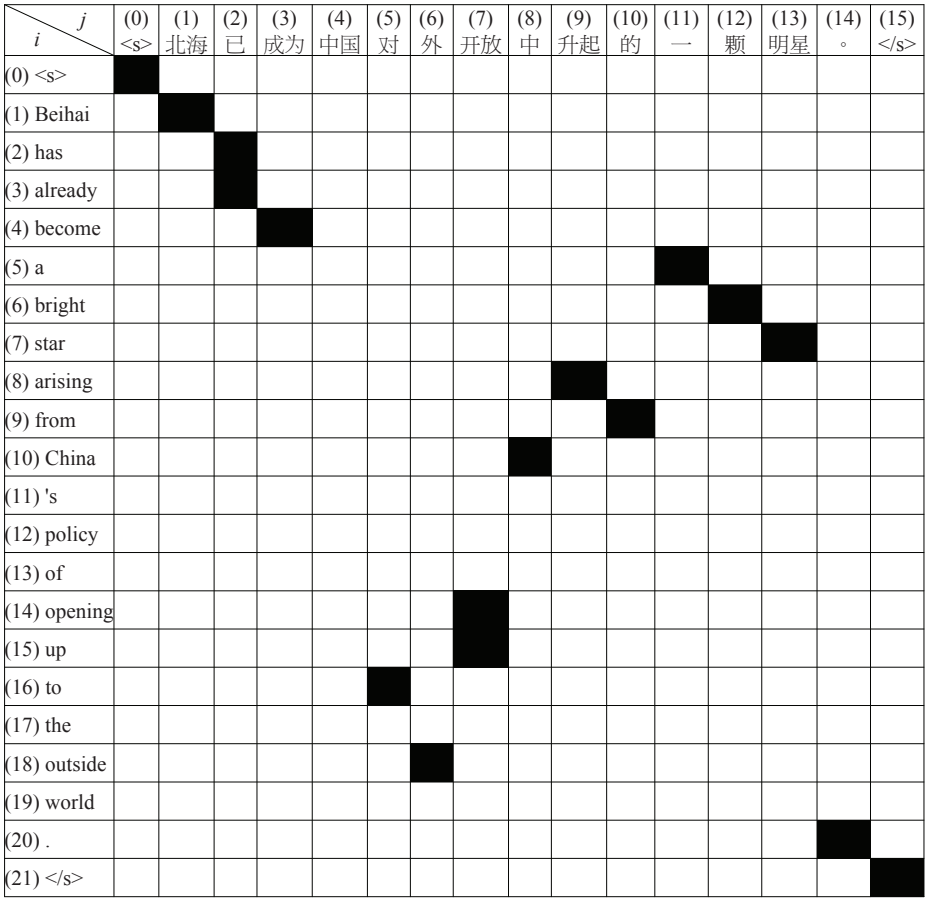
$$c_{j,j'} = \begin{cases} \textit{reversed} & \text{if } j' < j \\ \textit{ordered} & \text{if } j' > j \end{cases}$$

The phrase orientation classifier model is in the log-linear form:

$$p_{\lambda_1^N}(c_{j,j'}|f_1^J, e_1^I, i, j) = \frac{\exp\left(\sum_{n=1}^N \lambda_n h_n(f_1^J, e_1^I, i, j, c_{j,j'})\right)}{\sum_{c'} \exp\left(\sum_{n=1}^N \lambda_n h_n(f_1^J, e_1^I, i, j, c')\right)}$$

i is the target position of the current phrase, and f_1^J and e_1^I denote the source and target sentences respectively. c' represents possible categories of $c_{j,j'}$.

We can train this log-linear model on lots of labeled examples extracted from all of the aligned MT training data. Figure 2 is an example of an aligned sentence pair and the labeled examples that can be extracted from it. Also, different from conventional MERT training, we can have a large number of binary features for the discriminative phrase orientation classifier. The experimental setting will be described in Section 4.1.



i	j	j'	class
0	0	1	ordered
1	1	2	ordered
3	2	3	ordered
4	3	11	ordered
5	11	12	ordered
6	12	13	ordered
7	**13**	**9**	**reversed**
8	9	10	ordered
9	**10**	**8**	**reversed**
10	**8**	**7**	**reversed**
15	**7**	**5**	**reversed**
16	5	6	ordered
18	6	14	ordered
20	14	15	ordered

Figure 2: An illustration of an alignment grid between a Chinese sentence and its English translation along with the labeled examples for the phrase orientation classifier. Note that the alignment grid in this example is automatically generated.

The basic feature functions are similar to what Zens and Ney (2006) used in their MT experiments. The basic binary features are source words within a window of size 3 ($d \in -1, 0, 1$) around the current source position j, and target words within a window of size 3 around the current target position i. In the classifier experiments in Zens and Ney (2006) they also use word classes to introduce generalization capabilities. In the MT setting it's harder to incorporate the part-of-speech information on the target language. Zens and Ney (2006) also exclude word class information in the MT experiments. In our work we will simply use the word features as basic features for the classification experiments as well. As a concrete example, we look at the labeled example ($i = 4, j = 3, j' = 11$) in Figure 2. We include the word features in a window of size 3 around j and i as in Zens and Ney (2006), we also include words around j' as features. So we will have nine word features for ($i = 4, j = 3, j' = 11$):

Src_{-1}:已	Src_0:成为	Src_1:中国
$Src2_{-1}$:的	$Src2_0$:一	$Src2_1$:颗
Tgt_{-1}:already	Tgt_0:become	Tgt_1:a

2.2 Path Features Using Typed Dependencies

Assuming we have parsed the Chinese sentence that we want to translate and have extracted the grammatical relations in the sentence, we design features using the grammatical relations. We use the path between the two words annotated by the grammatical relations. Using this feature helps the model learn about what the relation is between the two chunks of Chinese words. The feature is defined as follows: for two words at positions p and q in the Chinese

Shared relations	Chinese	English
nn	15.48%	6.81%
punct	12.71%	9.64%
nsubj	6.87%	4.46%
rcmod	2.74%	0.44%
dobj	6.09%	3.89%
advmod	4.93%	2.73%
conj	6.34%	4.50%
num/nummod	3.36%	1.65%
attr	0.62%	0.01%
tmod	0.79%	0.25%
ccomp	1.30%	0.84%
xsubj	0.22%	0.34%
cop	0.07%	0.85%
cc	2.06%	3.73%
amod	3.14%	7.83%
prep	3.66%	10.73%
det	1.30%	8.57%
pobj	2.82%	10.49%

Table 1: The percentage of typed dependencies in files 1–325 in Chinese (CTB6) and English (English-Chinese Translation Treebank)

sentence ($p < q$), we find the shortest path in the typed dependency parse from p to q, concatenate all the relations on the path and use that as a feature.

A concrete example is the sentences in Figure 3, where the alignment grid and labeled examples are shown in Figure 2. The glosses of the Chinese words in the sentence are in Figure 3, and the English translation is "Beihai has already become a bright star arising from China's policy of opening up to the outside world." which is also listed in Figure 2.

For the labeled example $(i = 4, j = 3, j' = 11)$, we look at the typed dependency parse to find the path feature between 成为 and 一. The relevant dependencies are: *dobj*(成为, 明星), *clf*(明星, 颗) and *nummod*(颗, 一). Therefore the path feature is *PATH:dobjR-clfR-nummodR*. We also use the directionality: we add an *R* to the dependency name if it's going against the direction of the arrow.

3 Chinese Grammatical Relations

Our Chinese grammatical relations are designed to be very similar to the Stanford English typed dependencies (de Marneffe and Manning, 2008; de Marneffe et al., 2006).

3.1 Description

There are 45 named grammatical relations, and a default 46th relation *dep* (dependent). If a dependency matches no patterns, it will have the most generic relation *dep*. The descriptions of the 45 grammatical relations are listed in Table 2 ordered by their frequencies in files 1–325 of CTB6 (LDC2007T36). The total number of dependencies is 85748, and other than the ones that fall into the 45 grammatical relations, there are also 7470 dependencies (8.71% of all dependencies) that do not match any patterns, and therefore keep the generic name *dep*.

3.2 Chinese Specific Structures

Although we designed the typed dependencies to show structures that exist both in Chinese and English, there are many other syntactic structures that only exist in Chinese. The typed dependencies we designed also cover those Chinese specific structures. For example, the usage of "的" (DE) is one thing that could lead to different English translations. In the Chinese typed dependencies, there are relations such as *cpm* (DE as complementizer) or *assm* (DE as associative marker) that are used to mark these different structures. The Chinese-specific "把" (BA) construction also has a relation *ba* dedicated to it.

The typed dependencies annotate these Chinese specific relations, but do not directly provide a mapping onto how they are translated into English. It becomes more obvious how those structures affect the ordering when Chinese sentences are translated into English when we apply the typed dependencies as features in the phrase orientation classifier. This will be further discussed in Section 4.4.

3.3 Comparison with English

To compare the distribution of Chinese typed dependencies with English, we extracted the English typed dependencies from the translation of files 1–325 in the English Chinese Translation Treebank 1.0 (LDC2007T02), which correspond to files 1–325 in CTB6. The English typed dependencies are extracted using the Stanford Parser.

There are 116,799 total English dependencies, and 85,748 Chinese ones. On the corpus we use, there are 45 distinct dependency types (not including *dep*) in Chinese, and 50 in English. The coverage of named relations is 91.29% in Chinese and 90.48% in English; the remainder are the unnamed relation *dep*. We looked at the 18 shared relations

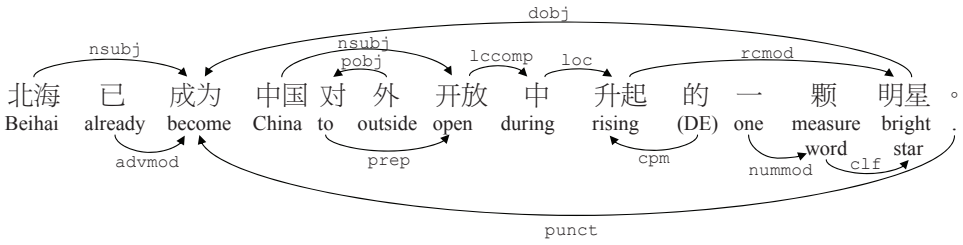


Figure 3: A Chinese example sentence labeled with typed dependencies

between Chinese and English in Table 1. Chinese has more *nn*, *punct*, *nsubj*, *rcmod*, *dobj*, *advmod*, *conj*, *nummod*, *attr*, *tmod*, and *ccomp* while English uses more *pobj*, *det*, *prep*, *amod*, *cc*, *cop*, and *xsubj*, due mainly to grammatical differences between Chinese and English. For example, some determiners in English (e.g., "the" in (1b)) are not mandatory in Chinese:

(1a) 进出口/import and export 总额/total value
(1b) The total value of imports and exports

In another difference, English uses adjectives (*amod*) to modify a noun ("financial" in (2b)) where Chinese can use noun compounds ("金融/finance" in (2a)).

(2a) 西藏/Tibet 金融/finance 体制/system 改革/reform
(2b) the reform in Tibet 's financial system

We also noticed some larger differences between the English and Chinese typed dependency distributions. We looked at specific examples and provide the following explanations.

prep* and *pobj English has much more uses of *prep* and *pobj*. We examined the data and found three major reasons:

1. Chinese uses both prepositions and postpositions while English only has prepositions. "After" is used as a postposition in Chinese example (3a), but a preposition in English (3b):
   (3a) 九七/1997 之後/after
   (3b) after 1997
2. Chinese uses noun phrases in some cases where English uses prepositions. For example, "之间" (period, or during) is used as a noun phrase in (4a), but it's a preposition in English.
   (4a) 九七/1997 到/to 九八/1998 之间 /period
   (4b) during 1997-1998
3. Chinese can use noun phrase modification in situations where English uses prepositions. In example (5a), Chinese does not use any prepositions between "apple company" and "new product", but English requires use of either "of" or "from".
   (5a) 苹果公司/apple company 新产品/new product
   (5b) the new product of (or from) Apple
   The Chinese DE constructions are also often translated into prepositions in English.

cc* and *punct The Chinese sentences contain more punctuation (*punct*) while the English translation has more conjunctions (*cc*), because English uses conjunctions to link clauses ("and" in (6b)) while Chinese tends to use only punctuation (",," in (6a)).

(6a) 这些/these 城市/city 社会/social 经济/economic 发展/development 迅速/rapid ， 地方/local 经济/economic 实力/strength 明显/clearly 增强/enhance
(6b) In these municipalities the social and economic development has been rapid, and the local economic strength has clearly been enhanced

rcmod* and *ccomp There are more *rcmod* and *ccomp* in the Chinese sentences and less in the English translation, because of the following reasons:

1. Some English adjectives act as verbs in Chinese. For example, 新 (new) is an adjectival predicate in Chinese and the relation between 新 (new) and 制度 (system) is *rcmod*. But "new" is an adjective in English and the English relation between "new" and "system" is *amod*. This difference contributes to more *rcmod* in Chinese.
   (7a) 新/new 的/(DE) 核销/verify and write off
   (7b) a new sales verification system
2. Chinese has two special verbs (VC): 是 (SHI) and 为 (WEI) which English doesn't use. For

abbreviation	short description	Chinese example	typed dependency	counts	percentage
nn	noun compound modifier	服务 中心	nn(中心, 服务)	13278	15.48%
punct	punctuation	海关 统计 表明 ,	punct(表明, ,)	10896	12.71%
nsubj	nominal subject	梅花 盛开	nsubj(盛开, 梅花)	5893	6.87%
conj	conjunct (links two conjuncts)	设备 和 原材料	conj(原材料, 设备)	5438	6.34%
dobj	direct object	浦东 颁布 了 七十一 件 文件	dobj(颁布, 文件)	5221	6.09%
advmod	adverbial modifier	部门 先 送上 文件	advmod(送上, 先)	4231	4.93%
prep	prepositional modifier	在 实践 中 逐步 完善	prep(完善, 在)	3138	3.66%
nummod	number modifier	七十一 件 文件	nummod(件, 七十一)	2885	3.36%
amod	adjectival modifier	跨世纪 工程	amod(工程, 跨世纪)	2691	3.14%
pobj	prepositional object	根据 有关 规定	pobj(根据, 规定)	2417	2.82%
rcmod	relative clause modifier	不 曾 遇到 过 的 情况	rcmod(情况, 遇到)	2348	2.74%
cpm	complementizer	开发 浦东 的 经济 活动	cpm(开发, 的)	2013	2.35%
assm	associative marker	企业 的 商品	assm(企业, 的)	1969	2.30%
assmod	associative modifier	企业 的 商品	assmod(商品, 企业)	1941	2.26%
cc	coordinating conjunction	设备 和 原材料	cc(原材料, 和)	1763	2.06%
clf	classifier modifier	七十一 件 文件	clf(文件, 件)	1558	1.82%
ccomp	clausal complement	银行 决定 先 取得 信用 评级	ccomp(决定, 取得)	1113	1.30%
det	determiner	这些 经济 活动	det(活动, 这些)	1113	1.30%
lobj	localizer object	近年 来	lobj(来, 近年)	1010	1.18%
range	dative object that is a quantifier phrase	成交 药品 一亿多 元	range(成交, 元)	891	1.04%
asp	aspect marker	发挥 了 作用	asp(发挥, 了)	857	1.00%
tmod	temporal modifier	以前 不 曾 遇到 过	tmod(遇到, 以前)	679	0.79%
plmod	localizer modifier of a preposition	在 这 片 热土 上	plmod(在, 上)	630	0.73%
attr	attributive	贸易额 为 二百亿 美元	attr(为, 美元)	534	0.62%
mmod	modal verb modifier	利益 能 得到 保障	mmod(得到, 能)	497	0.58%
loc	localizer	占 九成 以上	loc(占, 以上)	428	0.50%
top	topic	建筑 是 主要 活动	top(是, 建筑)	380	0.44%
pccomp	clausal complement of a preposition	据 有关 部门 介绍	pccomp(据, 介绍)	374	0.44%
etc	etc modifier	科技 、 文教 等 领域	etc(文教, 等)	295	0.34%
lccomp	clausal complement of a localizer	中国 对 外 开放 中 升起 的 明星	lccomp(中, 开放)	207	0.24%
ordmod	ordinal number modifier	第七 个 机构	ordmod(个, 第七)	199	0.23%
xsubj	controlling subject	银行 决定 先 取得 信用 评级	xsubj(取得, 银行)	192	0.22%
neg	negative modifier	以前 不 曾 遇到 过	neg(遇到, 不)	186	0.22%
rcomp	resultative complement	研究 成功	rcomp(研究, 成功)	176	0.21%
comod	coordinated verb compound modifier	颁布 实行	comod(颁布, 实行)	150	0.17%
vmod	verb modifier	其 在 支持 外商 企业 方面 的 作用	vmod(方面, 支持)	133	0.16%
prtmod	particles such as 所, 以, 来, 而	在 产业化 所 取得 的 成就	prtmod(取得, 所)	124	0.14%
ba	"ba" construction	把 注意力 转向 市场	ba(转向, 把)	95	0.11%
dvpm	manner DE(地) modifier	有效 地 防止 流失	dvpm(有效, 地)	73	0.09%
dvpmod	a "XP+DEV(地)" phrase that modifies VP	有效 地 防止 流失	dvpmod(防止, 有效)	69	0.08%
prnmod	parenthetical modifier	八五 期间 （ 1990 – 1995 ）	prnmod(期间, 1995)	67	0.08%
cop	copular	原 是 自给自足 的 经济	cop(自给自足, 是)	59	0.07%
pass	passive marker	被 认定 为 高 技术 产业	pass(认定, 被)	53	0.06%
nsubjpass	nominal passive subject	镍 被 称作 现代 工业 的 维生素	nsubjpass(称作, 镍)	14	0.02%

Table 2: Chinese grammatical relations and examples. The counts are from files 1–325 in CTB6.

example, there is an additional relation, *ccomp*, between the verb 是/(SHI) and 降低/reduce in (8a). The relation is not necessary in English, since 是/SHI is not translated.

(8a) 二/second 是/(SHI) 一九九六年/1996
中国/China 大幅度/substantially
降低/reduce 关税/tariff

(8b) Second, China reduced tax substantially in 1996.

conj There are more *conj* in Chinese than in English for three major reasons. First, sometimes one complete Chinese sentence is translated into several English sentences. Our *conj* is defined for two grammatical roles occurring in the same sentence, and therefore, when a sentence breaks into multiple ones, the original relation does not apply. Second, we define the two grammatical roles linked by the *conj* relation to be in the same word class. However, words which are in the same word class in Chinese may not be in the same word class in English. For example, adjective predicates act as verbs in Chinese, but as adjectives in English. Third, certain constructions with two verbs are described differently between the two languages: verb pairs are described as coordinations in a serial verb construction in Chinese, but as the second verb being the complement

of the first verb in English.

4 Experimental Results

4.1 Experimental Setting

We use various Chinese-English parallel corpora[1] for both training the phrase orientation classifier, and for extracting statistical phrases for the phrase-based MT system. The parallel data contains $1,560,071$ sentence pairs from various parallel corpora. There are $12,259,997$ words on the English side. Chinese word segmentation is done by the Stanford Chinese segmenter (Chang et al., 2008). After segmentation, there are 11,061,792 words on the Chinese side. The alignment is done by the Berkeley word aligner (Liang et al., 2006) and then we symmetrized the word alignment using the grow-diag heuristic.

For the phrase orientation classifier experiments, we extracted labeled examples using the parallel data and the alignment as in Figure 2. We extracted $9,194,193$ total valid examples: 86.09% of them are *ordered* and the other 13.91% are *reversed*. To evaluate the classifier performance, we split these examples into training, dev and test set $(8:1:1)$. The phrase orientation classifier used in MT experiments is trained with all of the available labeled examples.

Our MT experiments use a re-implementation of Moses (Koehn et al., 2003) called *Phrasal*, which provides an easier API for adding features. We use a 5-gram language model trained on the Xinhua and AFP sections of the Gigaword corpus (LDC2007T40) and also the English side of all the LDC parallel data permissible under the NIST08 rules. Documents of Gigaword released during the epochs of MT02, MT03, MT05, and MT06 were removed. For features in MT experiments, we incorporate Moses' standard eight features as well as the lexicalized reordering features. To have a more comparable setting with (Zens and Ney, 2006), we also have a baseline experiment with only the standard eight features. Parameter tuning is done with Minimum Error Rate Training (MERT) (Och, 2003). The tuning set for MERT is the NIST MT06 data set, which includes 1664 sentences. We evaluate the result with MT02 (878 sentences), MT03 (919 sentences), and MT05 (1082 sentences).

[1] LDC2002E18, LDC2003E07, LDC2003E14, LDC2005E83, LDC2005T06, LDC2006E26, LDC2006E85, LDC2002L27 and LDC2005T34.

4.2 Phrase Orientation Classifier

Feature Sets	#features	Train. Acc. Acc. (%)	Train. Macro-F	Dev Acc. (%)	Dev Macro-F
Majority class	-	86.09	-	86.09	-
Src	1483696	89.02	71.33	88.14	69.03
Src+Tgt	2976108	92.47	82.52	91.29	79.80
Src+Src2+Tgt	4440492	95.03	88.76	93.64	85.58
Src+Src2+Tgt+PATH	4691887	96.01	91.15	94.27	87.22

Table 3: Feature engineering of the phrase orientation classifier. Accuracy is defined as (#correctly labeled examples) divided by (#all examples). The macro-F is an average of the accuracies of the two classes.

The basic source word features described in Section 2 are referred to as Src, and the target word features as Tgt. The feature set that Zens and Ney (2006) used in their MT experiments is Src+Tgt. In addition to that, we also experimented with source word features Src2 which are similar to Src, but take a window of 3 around j' instead of j. In Table 3 we can see that adding the Src2 features increased the total number of features by almost 50%, but also improved the performance. The PATH features add fewer total number of features than the lexical features, but still provide a 10% error reduction and 1.63 on the macro-F1 on the dev set. We use the best feature sets from the feature engineering in Table 3 and test it on the test set. We get 94.28% accuracy and 87.17 macro-F1. The overall improvement of accuracy over the baseline is 8.19 absolute points.

4.3 MT Experiments

In the MT setting, we use the log probability from the phrase orientation classifier as an extra feature. The weight of this discriminative reordering feature is also tuned by MERT, along with other Moses features. In order to understand how much the PATH features add value to the MT experiments, we trained two phrase orientation classifiers with different features: one with the Src+Src2+Tgt feature set, and the other one with Src+Src2+Tgt+PATH. The results are listed in Table 4. We compared to two different baselines: one is Moses8Features which has a distance-based reordering model, the other is Baseline which also includes lexicalized reordering features. From the table we can see that using the discriminative reordering model with PATH features gives significant improvement over both base-

Setting	#MERT features	MT06(tune)	MT02	MT03	MT05
Moses8Features	8	31.49	31.63	31.26	30.26
Moses8Features+DiscrimRereorderNoPATH	9	31.76(+0.27)	31.86(+0.23)	**32.09**(+**0.83**)	**31.14**(+**0.88**)
Moses8Features+DiscrimRereorderWithPATH	9	**32.34**(+**0.85**)	**32.59**(+**0.96**)	**32.70**(+**1.44**)	**31.84**(+**1.58**)
Baseline (Moses with lexicalized reordering)	16	32.55	32.56	32.65	31.89
Baseline+DiscrimRereorderNoPATH	17	32.73(+0.18)	32.58(+0.02)	32.99(+0.34)	31.80(−0.09)
Baseline+DiscrimRereorderWithPATH	17	**32.97**(+**0.42**)	**33.15**(+**0.59**)	**33.65**(+**1.00**)	**32.66**(+**0.77**)

Table 4: MT experiments of different settings on various NIST MT evaluation datasets. All differences marked in bold are significant at the level of 0.05 with the approximate randomization test in (Riezler and Maxwell, 2005).

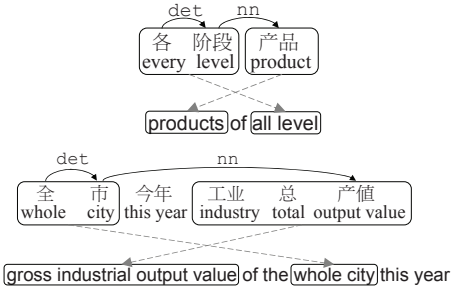


Figure 4: Two examples for the feature *PATH:det-nn* and how the reordering occurs.

lines. If we use the discriminative reordering model without PATH features and only with word features, we still get improvement over the Moses8Features baseline, but the MT performance is not significantly different from Baseline which uses lexicalized reordering features. From Table 4 we see that using the Src+Src2+Tgt+PATH features significantly outperforms both baselines. Also, if we compare between Src+Src2+Tgt and Src+Src2+Tgt+PATH, the differences are also statistically significant, which shows the effectiveness of the path features.

4.4 Analysis: Highly-weighted Features in the Phrase Orientation Model

There are a lot of features in the log-linear phrase orientation model. We looked at some highly-weighted PATH features to understand what kind of grammatical constructions were informative for phrase orientation. We found that many path features corresponded to our intuitions. For example, the feature *PATH:prep-dobjR* has a high weight for being *reversed*. This feature informs the model that in Chinese a PP usually appears before VP, but in English they should be reversed. Other features with high weights include features related to the DE construction that is more likely to translate to a relative clause, such as *PATH:advmod-rcmod* and *PATH:rcmod*. They also indicate the phrases are more likely to be chosen in reversed order. Another frequent pattern that has not been emphasized in the previous literature is *PATH:det-nn*, meaning that a [DT NP_1 NP_2] in Chinese is translated into English as [NP_2 DT NP_1]. Examples with this feature are in Figure 4. We can see that the important features decided by the phrase orientation model are also important from a linguistic perspective.

5 Conclusion

We introduced a set of Chinese typed dependencies that gives information about grammatical relations between words, and which may be useful in other NLP applications as well as MT. We used the typed dependencies to build path features and used them to improve a phrase orientation classifier. The path features gave a 10% error reduction on the accuracy of the classifier and 1.63 points on the macro-F1 score. We applied the log probability as an additional feature in a phrase-based MT system, which improved the BLEU score of the three test sets significantly (0.59 on MT02, 1.00 on MT03 and 0.77 on MT05). This shows that typed dependencies on the source side are informative for the reordering component in a phrase-based system. Whether typed dependencies can lead to improvements in other syntax-based MT systems remains a question for future research.

Acknowledgments

The authors would like to thank Marie-Catherine de Marneffe for her help on the typed dependencies, and Daniel Cer for building the decoder. This work is funded by a Stanford Graduate Fellowship to the first author and gift funding from Google for the project "Translating Chinese Correctly".

References

Pi-Chuan Chang, Michel Galley, and Christopher D. Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 224–232, Columbus, Ohio, June. Association for Computational Linguistics.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8, Manchester, UK, August. Coling 2008 Organizing Committee.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC-06*, pages 449–454.

Liang Huang, Kevin Knight, and Aravind Joshi. 2006. Statistical syntax-directed translation with extended domain of locality. In *Proceedings of AMTA*, Boston, MA.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proc. of NAACL-HLT*.

Philipp Koehn, Hieu Hoang, Alexandra Birch Mayne, Christopher Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Demonstration Session*.

Percy Liang, Ben Taskar, and Dan Klein. 2006. Alignment by agreement. In *Proceedings of HLT-NAACL*, pages 104–111, New York City, USA, June. Association for Computational Linguistics.

Franz Josef Och, Daniel Gildea, Sanjeev Khudanpur, Anoop Sarkar, Kenji Yamada, Alex Fraser, Shankar Kumar, Libin Shen, David Smith, Katherine Eng, Viren Jain, Zhen Jin, and Dragomir Radev. 2004. A smorgasbord of features for statistical machine translation. In *Proceedings of HLT-NAACL*.

Franz Josef Och. 2003. Minimum error rate training for statistical machine translation. In *ACL*.

Stefan Riezler and John T. Maxwell. 2005. On some pitfalls in automatic evaluation and significance testing for MT. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 57–64, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Christoph Tillman. 2004. A unigram orientation model for statistical machine translation. In *Proceedings of HLT-NAACL 2004: Short Papers*, pages 101–104.

Chao Wang, Michael Collins, and Philipp Koehn. 2007. Chinese syntactic reordering for statistical machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 737–745, Prague, Czech Republic, June. Association for Computational Linguistics.

Richard Zens and Hermann Ney. 2006. Discriminative reordering models for statistical machine translation. In *Proceedings on the Workshop on Statistical Machine Translation*, pages 55–63, New York City, June. Association for Computational Linguistics.