國 立 清 華 大 學

碩 士 論 文

題目　統計式機器翻譯之句法式辭彙重排

Learning Syntactical Reordering of Source Sentences for Statistical Machine Translation

系別　資訊系統與應用研究所　　組別　甲組

學號姓名　946706 陳維德 (Wei-Teh Chen)

指導教授　張俊盛 博士 (Dr. Jason S. Chang)

中華民國九十六年六月

摘要

在本論文中，我們提出一個由原文句的句法結構剖析樹，學習句法式的辭彙重排模組。我們的方法先利用句法剖析工具，對原文句進行句法剖析，並使用單字對應工具，從句對應的雙語資料中，產生單字對應的資訊。針對所產生的句法樹，我們經由節點所對應到的目標句單詞位置，決定節點重新排列到目標與順序的機率函數。此一機率函數是透過句法樹節點的各種特徵值，以及所標記的順序標記，輸入到機器學習的工具以自動學習重新排列機率。在測試時，我們先將測試句經由句法剖析工具，產生句法樹，再藉由訓練模型，對句法樹中的節點進行機率預測，產生重新排序的剖析樹，以及此句法樹的葉節點所組成的重排句。最後，我們將重排句輸入到現成的機器翻譯系統，產生翻譯句。

我們實際撰寫了程式，由香港語料中選出訓練及測試資料，比較片語式統計機器翻譯作法顯示在搭配我們的辭彙重排模型，是否有翻譯提昇的效果。實驗的結果，以 BLEU 分數做評估，搭配我們的模組，比未搭配高出了 1%。實驗顯示搭配了我們的方法，對於統計式機器翻譯模組的辭彙重排，有正面的幫助，並改善了機器翻譯的品質。

關鍵詞： 詞彙排序、統計式機器翻譯、句法剖析樹

Abstract

We present a method for learning to perform syntactical reordering in machine translation. In our approach, source sentences are parsed into parse trees aimed at reordering source parse trees into reordered parse trees closer to target language structure. The method involves aligning words, parsing source sentences into parse trees, determining tree nodes reordering operation, and training a probability model using tree node features via machine learning model. At run-time, we parse the test sentence to obtain the parse trees, estimating reordering operation for each tree node using the trained model, and returning the sequence of words in reordered source parse tree to obtain reordered source sentence. We submit reordered source sentence to a state-of-the-art machine translation system for evaluation. We describe the implementation of the method using parallel Hong Kong corpus. The experiment results show that phrase-based machine translation model with our reordering model outperforms machine translation model without our reordering model in terms of BLEU score. Our methodology is clearly a step forward for producing more fluent and grammatical translation.

Keyword: Syntactical Reordering, Statistical Machine Translation, Parse Tree

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Machine translation has become more and more popular in recent years, and more and better qualify machine translation services on the Web are increasingly available. It is used to translate sentences or article from one language to another language more speedily than human translation. Most notable machine translation systems include Google Translate (http://www.google.com/translate_t) and AltaVista Babel Fish Translation (http://babelfish.altavista.com).

Figure 1: A translation sample from Google Translate. Submitting "A fire occurred in my neighborhood after midnight" which is not reordered

Machine translation systems such as Google Translate often treat source sentences as a sequence of phrases and produce a target sentence with word order similar to the original sentence. However, statistical machine translation often fails to effectively reorder phrases globally, leading to syntactical incorrect target sentence. Consider the English sentence "*A fire occurred in my*

neighborhood after midnight", Google Translate translates it into "*火災 發生 在 我的 鄰居 在 午夜 後*" (Figure 1). The ordering for short phrases, like"*我的 鄰居*" and "午夜 後", are acceptable while the global order of phrases are not. These sentences can be translated into more fluent and grammatical target sentence if syntactical information is taken in consideration.

For the sentence "*A fire occurred in my neighborhood after midnight*" again, parse tree can be obtained though a syntactical parser which is shown in Figure 2(a). By rotating (*S* (*NP*) (*VP*)) into (*S* (*VP*) (*NP*)) ,and rotating (*VP* (*VBD*) (*PP*) (*PP*)) into (*VP* (*PP*) (*PP*) (*VBD*)), and rotating (*PP* (*IN*) (*NP* (*NN*))) into (*PP* (*NP* (*NN*)) (*IN*)), a new parse tree closer to the Chinese sentence order is obtained (Figure 2(b)). After that, we can process the reordered English sentence "*midnight after in my neighborhood occurred a fire*" with a state-of-the art machine translation system. Finally, a machine translation system is more likely to produce the precise Chinese sentence "*午夜 後 我 家 附近 發生 火災*". Submitting this reordered English sentence to Google Translate, it is translated into "*午夜 後 , 我 在 附近 發生 火災*" which is more influent than Google Translate's result for the original sentence. (Figure 3)

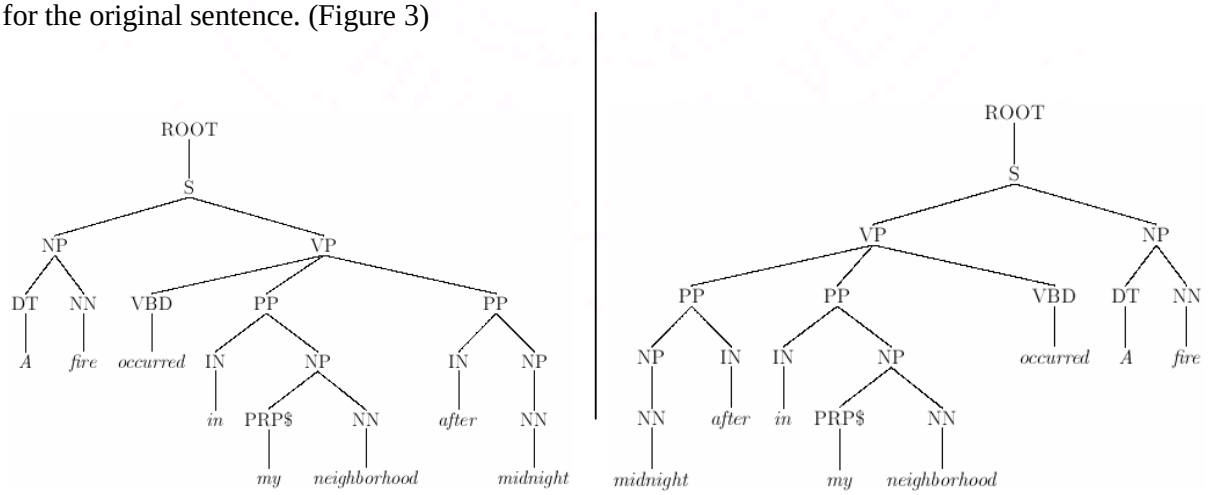

Figure 2: Original (left) and reordered (right) parse trees for the English sentence "*A fire occurred in my neighborhood after midnight.*"

We present and evaluate a new syntactical reordering method that processes source sentences into a word order closer to target sentences. For a given source sentence, we obtain the parse tree using a syntactic parser. Then we estimate reordering probability for every non-terminal tree nodes using a number of features via Conditional Random Field (CRF), a type of discriminative probabilistic model most often used for the labeling data. After predicting reordering probability, we could reorder the tree and produce a reordered sentence closer to target language word order. We can then feed this reordered sentence into a phrase-based statistical model to produce a target language sentence with better word order. We describe our method in details in Chapter 3.

Figure 3: Result for submitting reordered sentence to Google Translate

The rest of the paper is organized as follows. We review the related work in the next section. Then we present our method for training reordering probability via CRF (Chapter 3). For our evaluation, we compare the BLEU score of the translated sentences against translations produced by other state-of-the-art systems (Chapter 4). Finally, we conclude with a summary and future work (Chapter 5).

Chapter 2 Related Work

Statistical machine translation (SMT) has been an area of active research, since Brown et al. (1990) proposes to model a statistical approach to machine translation. The relation of two languages translation process is based on the noisy channel model. And the processing of translation is similar to a decoder where we choose the most possible target language sentence which optimizes probabilities related to channel and language models, the two models: in the noisy channel language model. And these two models operate independently. This model has become the standard framework of statistical machine translation.

More specially, we focus on the special part of SMT, translating with syntax information that is likely to obtain target sentences with grammatical structure. Translating with syntax information has long been an active topic of SMT research. The body of the SMT research (Yamada and Knight, 2003) in a way focuses on translating with syntax-based statistical translation model. A source language parse trees is taken into consideration. By using the operations of reordering, insertion and word translation, a target sentence is produced. In our proposed model, we also parse the source sentence and process with source language parse tree. But the process of actual translation is done using a state-of-the-art phrase-based statistical machine translation system in our model. An interesting approach presented by (Marcu, Wang, Echihabi, and Knight, 2006) descries how to exploiting feature functions for choosing target translation phrases from source language phrases, and use syntax-based translation. With composing rules, target language sub models can be trained to assemble target phrases into well-formed, grammatical outputs. Their evaluation results show that a syntax-based system can produce results that are better than those produced by a strong phrase-based system in experimental conditions. And the number of syntax-based rules used in their models is smaller than the number of phrase-based rules in a typical phrase-base statistical machine

translation system.

In most of the SMT systems, the processing of translation is done by considering translation model and language model at the same time. A method for syntactic reordering has been presented in Wang, Collins, and Koehn (2007), which describe an approach with a set of hand crafted syntactic reordering rules that exploit systematic differences between source and target word order. By reordering some tree nodes with rules, a new, reordered parse is presented where the sequence of leaf words from a reordered sentence. And the reordered source sentences are submitted to a phrase-based SMT system to obtain target language with suitable order influent and grammatical. While they use hand crafted reordering rules, we proposes to learn reordering rules and probability automatically from the training data.

Wu (1997) described an Inversion Transduction Grammar (ITG) to model translation. ITG is a bilingual context-free grammar that generates a synchronous parse tree for a pair of aligned sentences. Any ITG can be converted into two normal forms, where the productions are either binary-fanout non-terminal productions or lexical productions. ITG allows for two reordering operations: straight or inverted. By training from a parallel corpus, a small number of rules are adopted which the accuracy for each rule is reliable. While ITG consider reordering with out even using syntactic labels and binary-fanout, we add more features into our proposed model for modeling reordering and we allow for fanout of more than 2. And with more features combination, the rules we obtain are much more effective than rules generated by ITG rules.

Chiang (2005) introduced lexicalized labelless hierarchical bilingual phrase structure to model translation without any linguistic commitment. Since he does not create any syntactic information category for hierarchical phrase pairs, very large number of rules is generated for specific words without any linguistics motivation. These lexicalized rewrite rules, however, record any differences

in hierarchical structure of two languages. The size of the rules is huge (2.2M rules). These rules do not make use of syntactical information such as part of speech tags like noun, verb, or adjective, instead they work with specific words. The results show that these rules improve translation accuracy compared with a state-of-the-art phrase-based system. Chiang also experimented with incorporation of syntactic information, which, however, did not provide a statistically significant gain. We integrate the features and syntactical structure in our model.

Most recently, the traditional SMT systems have begun to use phrase-based approach. Koehn (2004, 2006) describes a beam search decoder for two phase-based statistical machine translation tools, Pharaoh and Moses. This approach would translate source sentences into translated sentences with appropriate phrase translation and reordering. With a beam search, the hypotheses are placed in stacks. The size of stacks has been limited; some of the hypothesis in stacks may be pruned if the probability of these hypotheses is lower than a threshold. The processing time directly related to the pruning threshold. The reordering model adopted by Pharaoh and Moses are fairly weak. The cost of target phrases reordering is related to the distance between source phrase and target translated phrases. The cost with long distance reordering may be quite high, thus the probabilities of global reordering hypotheses are usually low and often pruned during beam search. The difference between Pharaoh and Moses is that Moses adds factored presentation, such as surface, forms, lemma, part-of-speech and other syntactic features, which Pharaoh only considers the translation probability and language probability. Our proposed model acts as a pre-processor for a phrase-based SMT, and we are likely to handle global reordering of translation more effectively.

In contrast to the previous research in phrase-based and syntax-based machine translation systems, we present a model that automatically learns syntactical reordering of source sentences, reordered tree nodes with probable reordering, and obtain the sequence of leaf words on the reordered tree. We exploit the features combination and training model learning to determine the

reordering of tree nodes. Our goal is to obtain reordered sentences with the goal of producing more fluent and grammatical target sentences.

Chapter 3 Syntactical Reordering Model

Submitting a source sentence (e.g., "*A fire occurred in my neighborhood after midnight*") to machine translation systems often produces reasonably useful translation. Unfortunately, machine translation systems typically order phrases monotonously, and may not produce fluent and syntactically correct target sentences. To obtain a translation with better reordering of words, a promising approach is to parse and reorder the source sentence into a new source sentence parse tree more consistent with the target language word order.

3.1 Language Model in Machine Translation

In this section, we present a new model of machine translation for determining the reordering probability for each parse tree node. Before describing the proposed method, we introduce basic concepts related to statistical machine translation.

In the noisy channel model (Brown et al., 1993), a source language sentence s enters in the noisy channel and translates into target language sentence t. The channel is governed by $P(t|s)$. In machine translation via the noisy channel, we decode the target sentence to give the most likely source sentence. By applying the Bayes rule, we have $P(t|s)= P(s)P(s|t)/P(s)$, where $P(s)$ is constant for the given source sentence s. This equation reformulates the translation processing into a translation component $P(s|t)$ and a language model component $P(t)$.

The probability of a given sentence $S=w_1w_2\ldots w_m$, the Visible Markov model is applied on $P(S) = P(w_1)P(w_2|w_1)P(w_3|w_1,w_2)\ldots P(w_m|w_1,w_2,\ldots,w_{m-1})$ without loss of generality. For the reason of lack of substring probability $P(w_k|w_1,w_2,\ldots,w_{k-1})$, it is applied on a n-gram language model which is an

application of n^{th} order Visible Markov model :

$$P(S) = P(w_1)P(w_2 \mid w_1)\ldots P(w_n \mid w_1, w_2, \ldots w_{n-1})\ldots P(w_m \mid w_{m-n+1}, w_{m-n+2}, \ldots w_{m-1}) \quad (1)$$

where the probability for each w_i depends only on the substring of $w_{i-n+1}w_{i-n+2}\ldots w_{i-1}$. N-gram language model is significantly helpful and easy to use in phrases ordering for phrase-based machine translation systems (e.g. Pharaoh). For a machine translation decoder, N-gram language model is used to estimate the probability of target sentences. Unfortunately, the problem of machine translation decoding is NP-complete (Knight, 1999) and thus state-of-the-art decoders limit the size of ordering hypotheses for translated sentences. Local reordering is typical very strong in machine translation systems since global reordering is very computationally intensive when many words are skipped over. However, global reordering is often needed for fluent and grammatical translations.

To overcome the limitation of local reordering, we use the syntactical parse tree of the source sentence to derive global reordering that is closer to target language word order. With that in mind, we rewrite the Noisy channel formulation as follows:

$$P(t \mid s) = P(t \mid s')P(\prod' \mid \prod)P(\prod \mid s) \quad \text{where } s'=Str(\Pi') \quad (2)$$

In equation (2), a parse tree Π for the sentence s is used to obtain the reordered sentence s'. There is a reorder operation applied to each node of Π, producing a new tree Π'. We denote $Str(\Pi')$ as the sequence of leaf words of parse tree Π' which is produced by a reordered parse tree Π' from original parse tree Π. The reordered source sentence feeds into a statistical MT system to produce target sentence t. Figure 4 shows an example describing formula (2) in a practical way. We will give our problem in a formal description in next section, and how to estimate the probability of reordering parse tree in (Section 3.3).

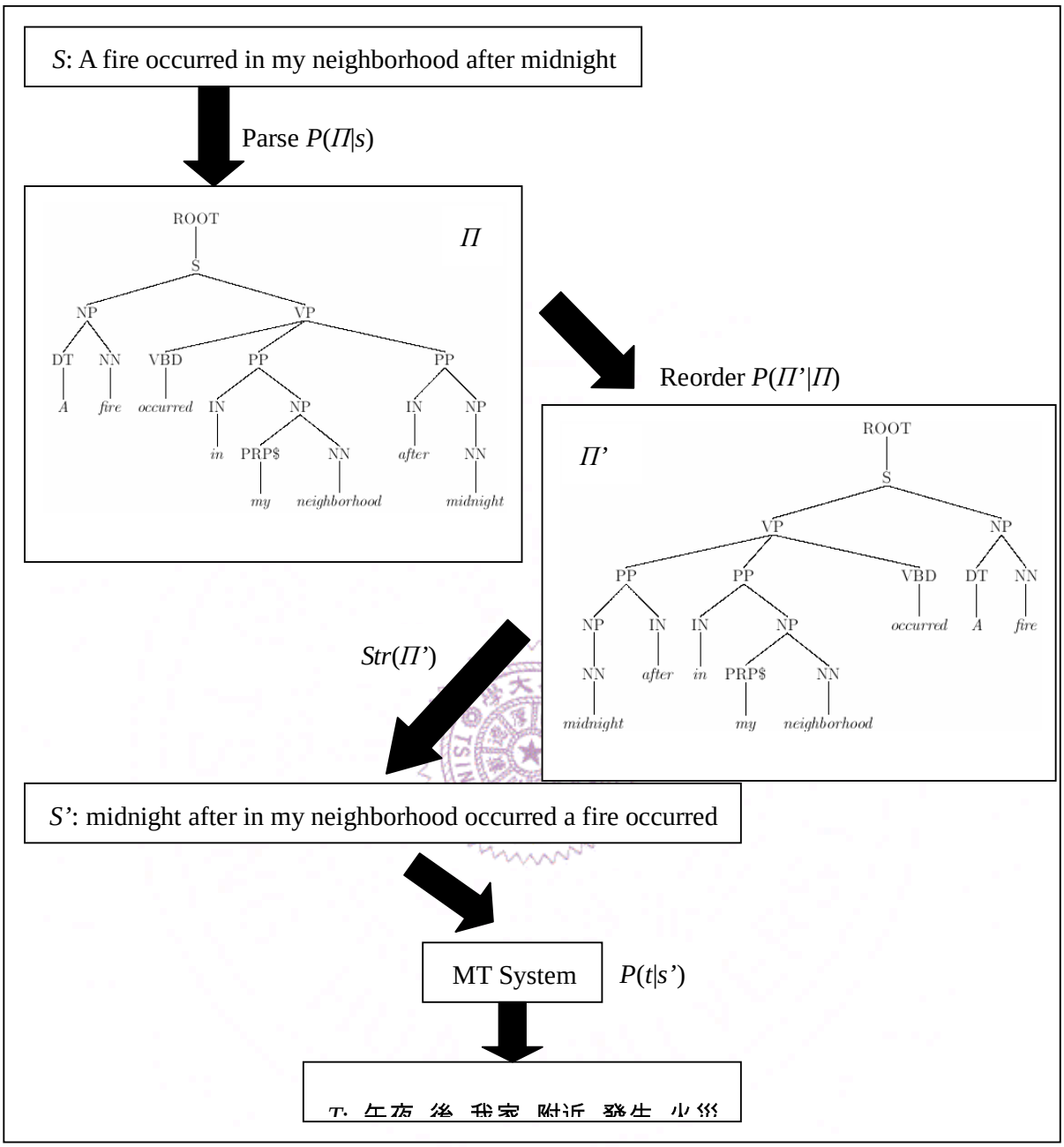

Figure 4: Overview of our proposed model

3.2 Problem Statement

We focus on the subtask of machine translation: reordering source language parse tree so as to produce a tree closer to target language word order. We then traverse the tree in infix order to

produce a reordered sentence. The returned sentence can be passed on to a state-of-the-art machine translation system to produce translation. Our goal is to reorder each tree node so that the MT system can produce more fluent and grammatical target sentence. Formal statement for this problem is as follows.

Problem Statement: We are given a general purpose natural language parser *PAR* (e.g., Stanford Parser) and a parallel corpus *C*. By parsing a source language sentence via *PAR*, we obtain the syntactic parse tree *ST* with tree nodes *TNs*. Our goal is to reorder each *TN* that is likely to transform *ST* to a parse tree *RT* with a syntactic structure similar to the target sentence. For this, we represent each *TN* as a set of features, $f_1,\ldots,f_m$, to determine the reorder probability for each tree node.

In the rest of this section, we describe our solution to this problem. First we describe on our training process of learning how to transform source parse tree (Section 3.3). Then we show the runtime process of reordering and translating a given sentence. (Section 3.4)

3.3 Training Process

We attempt to find transformation from source language parse tree into reordered parse tree to produce a sentence with word order similar to the target sentence. Our proposed training process is shown in Figure 5.

1. Aligning words in sentence pairs of bilingual corpus and generating the parse tree for each sentence (Section 3.3.1)
2. Determining word reordering operation information for each node (Section 3.3.2)
3. Training a CRF model using all features(Section 3.3.3)

Figure 5: Outline of the training process

3.3.1 Preprocessing for Training Data

In the first stage of the training process (Step (1) in Figure 5), we generate a set of source language parse trees with leaf words annotated with its translation counterpart, and then identify the reordering operation on each tree nodes that will lead to a tree where word translations are in order. For example, consider the sub-tree in Figure 6. Intuitively, if we reverse the left and right child note, the word translations, 張(1) 先生(2), will be in order. And the reverse decision may have to do with leaf word left child "Mr."

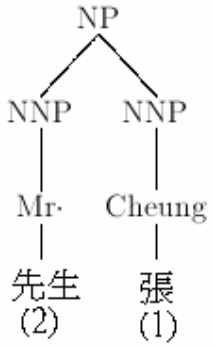

Figure 6: Sub-tree for "*Mr. Cheung*"

The input to this stage is a set of parallel source-target language sentences. These aligned sentences constitute the training data for learning to reorder parse trees.

The output of this stage is a set of source sentences annotated parse trees with word translations and positions that can be used to quickly determine reordering operations for each tree nodes. A sample parse tree, automatically generated using a state-of-the-art parser and word-alignment tool, is shown in Figure 7(b).

We describe the steps of data handling in this section. First, we run an existing word alignment on *C*. Then we parse each source sentence *s* using *PAR*. For each sentence pair (*s*,*t*) in bilingual corpus *C*, consider *s* = "*A fire occurred in Mr. Cheung's neighborhood after midnight*" and *t* = "*午夜 後 張 先生 家 附近 發生 火災*". For this sentence pair, we obtain the result of word alignment in Table I. Then, we annotate each leaf node with alignment target words and target word position. The results of annotated parse tree are shown in Figure 7(b).

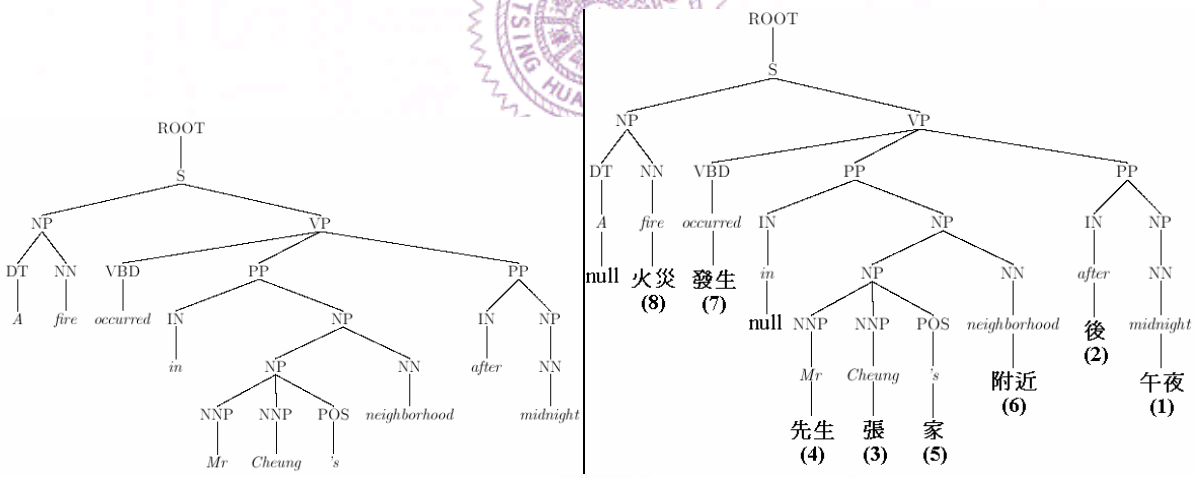

Figure 7: Original parse tree (left) and annotated parse tree (right). Source sentence is "*a fire occurred in Mr. Cheung's neighborhood after midnight*" with target sentence "*午夜*$_1$ *後*$_2$ *張*$_3$ *先生*$_4$ *家*$_5$ *附近*$_6$ *發生*$_7$ *火災*$_8$"

The training process can be illustrated using the procedure in Figure 8. We will describe the

determination of the reordering operation for each tree node in the next section.

Source Position	Source Word	Target Position	Target Word
1	*a*	0	*NULL*
2	*fire*	8	*火災*
3	*occurred*	7	*發生*
4	*in*	0	*NULL*
5	*Mr.*	4	*先生*
6	*Cheung*	3	*張*
7	*'s*	5	*家*
8	*neighborhood*	6	*附近*
9	*after*	2	*後*
10	*midnight*	1	*午夜*

Table I: Word-Alignment result for "*a fire occurred in Mr. Cheung's neighborhood after midnight*" with "*午夜*$_1$ *後*$_2$ *張*$_3$ *先生*$_4$ *家*$_5$ *附近*$_6$ *發生*$_7$ *火災*$_8$"

Procedure Preprocessing (Bilingual Corpus $C=\{(s_1,t_1), (s_2,t_2), \ldots, (s_n,t_n)\}$)

$S = \{s_1,s_2,\ldots,s_n\}$
(1) Syntactical parse tree *set ST* = *parse*(*S*)
(2) A = word_align(C)
for each syntactical parse tree Π_i in *ST*
for each leaf word e_j in Π_i
(3) annotateTreeLeafNodeAndPosition(e_j, f_{aj}, a_j)

Figure 8: Procedure for preprocessing for training data

3.3.2 Determining Word Reordering Operation

In the second stage of training process (Step (2) in Figure 5), we describe how to determine the reordering operation for each non-terminal node in the parse tree. We apply a straightforward procedure to determine reordering operations based on word positions. Figure 9 shows an example of determining reordering operation for each node in an annotated parse tree. We use permutation of 1 to n to denote the reordering operations. For example, the sub-tree (*PP* (*IN after-後*,2) (NP (NN *midnight-午夜*,1))) on the right of Figure 9 is label as "*2 1*" indicating the target words are reversed in the target sentence. And the node (*NP* (*NNP Mr.-先生*,4) (*NNP Cheung-張*,3) (*POS 's-家*,5) is labeled as "2 1 3" indicating that the second translation appears first in the target sentence followed by those of the first and third translations.

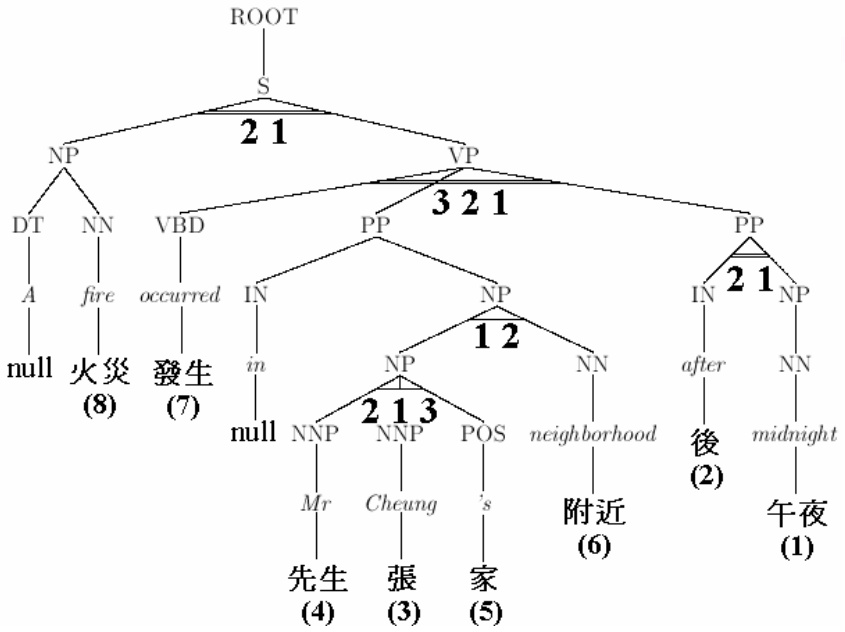

Figure 9: Determining reorder options for an annotated parse tree

We use a bottom-up tree traversal to determining reordering operation (RO) for each tree node.

First we determine *RO* on base phrase tree node. Since the target word positions have been annotated on parse tree in the first stage, it is easy to derive *RO* based on word translation positions. For instance, the permutation of “2 1 3” will change the word positions of the node for “Mr. Cheung ‘s” from (4 3 5) to (3 4 5) in order.

For tree node on higher of parse tree, we label *RO* after all child nodes have been labeled. An example of reordered tree is shown in Figure 10. The three child nodes have been labeled and reorder in target sentence position. In order to reorder this sub-tree node also into target sentence order, we label the root with “3 2 1” labeling to the sequence of translation positions of (1 2 3 4 5 6 7) which is the same on the target word positions. And the reorder operations in Figure 9 are shown in Table II.

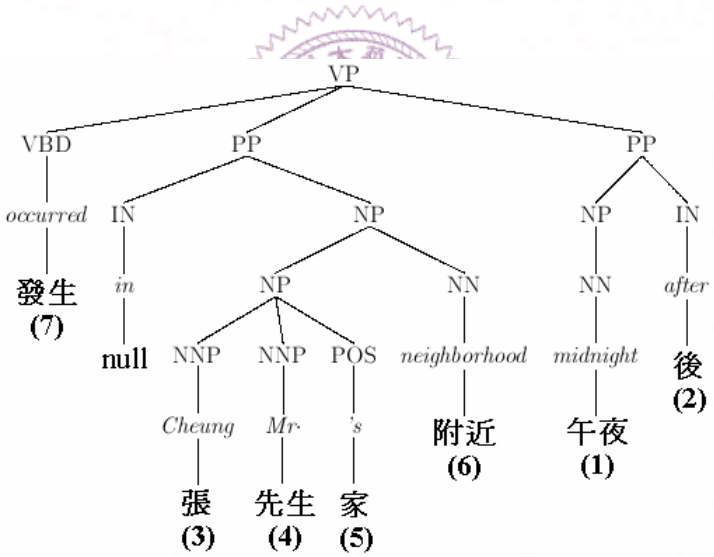

Figure 10: Example sub-tree for reorder operation determining

Sub-Tree	POS of Root	RO
π_1	*S*	*2 1*
(S (NP VP))		
π_2	*NP*	*NULL*
(NP (DT NN))		
π_3	*VP*	*3 2 1*
(VP (VBD PP PP))		
π_4	*PP*	*NULL*
(PP (IN NP))		
π_5	*NP*	*1 2*
(NP (NP NN))		
π_6	*NP*	*213*
(NP (NNP NNP POS))		
π_7	*PP*	*2 1*
(PP (IN NP))		

Table II: Reorder Operation of sub-tree. π_i stands for sub-tree by traversing tree nodes in preorder

We label reorder operation for non-terminal nodes describing in above section. Unfortunately, the number of children in non-terminal nodes is various; therefore, a non-terminal node with n children is mostly having n! types of ordering. We only label for non-terminal nodes with 2 or 3 children with our notation above. For sub-tree nodes with more than three children, we only consider straight and inverted cases and label them as “1 2” and “2 1” for simplicity.

By labeling all tree node nodes using the above *RO* notation, we annotate the reordering operation for each non-terminal node with a straightforward procedure. The ordering for these nodes will be the training data for probability estimation, which we will describe in details in the next section.

3.3.3 Training a CRF Model Using All Features

In the final training stage, we take the lexical and syntactic features of each node into consideration, and use a machine learning tool, Conditional Random Field, to estimate reordering probability conditioned on all features. The input of this stage is a set of parse trees annotated with reordering operations.

Consider the formula (2) in section 3.1. In this formula, we obtain a reordered parse tree *Π'* corresponding to a reordered source sentence *s'*. We rewrite this mathematical operation as follows:

$$P(\Pi' \mid \Pi) = \prod_{\pi_j} P(r_j \mid \pi_j) = P(R \mid \Pi) \tag{3}$$

where the annotated parse tree *Π* is composed of sub-tree π_j, and *R* denoting a set of reordering operations r_j for π_j while transform *Π* into a new parse tree *Π'*. The probability of *R* can be modeled using the label set *R* conditioned on the parse tree *Π*. We focus on the procedure $P(R|\Pi)$. The probability estimation is similar to a tagging task, and we can use existing probability estimation tools to train a tagger.

One of the tools for our purpose is Conditional Random Field (Lafferty, etc. 2001) (CRF). It is a framework for building probabilistic models for labeling and segmenting structural data, such as sequences, trees, and lattices. Conditional Random Field provides several advantages over hidden Markov model and related model. CRF has the ability to relax strong independence assumptions made in these models. A CRF is a form of undirected graphical model that defines a single log-linear distribution over the label sequences given a particular observation sequence (Wallach, 2004). We introduce CRF in details with the graphical model below.

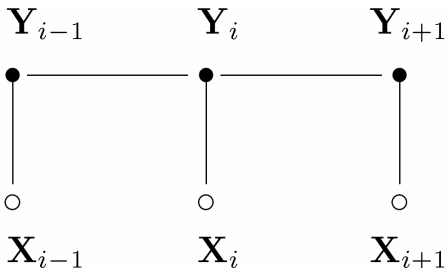

Figure 11: Graphical structures of simple CRF for sequences. An open circle indicates that the variable is not generated by the model.

First, we define two variables, *X* and *Y*, where *X* is a random variable over data sequences, and *Y* is a random variable over label sequences. Thus *X* is the observation variable and *Y* is the unseen variable. Then the graphical model for CRF is shown in Figure 11. X_i is observation variable in position *i* of data sequence *X*, and Y_i is the same expression. In Figure 11, a label Y_i is condition on two components, Y_j and X_i where Y_j is the neighborhood of Y_i and X_i is the observation variable for Y_i. Thus we separate the probability estimation into two parts, state function and transition function. The definition of the probability in simple form is as follows:

$$p(Y \mid X) \propto \exp\left(\sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i)\right) \tag{4}$$

where $t_j(y_{i-1},y_i,x,i)$ is a transition feature function of the entire observation sequence and the labels at position *i* and *i-1* in the label sequence; $s_k(y_i,x,i)$ is a state feature function of the label at position *i* and observation sequence; and λ_j and μ_j are parameters to be estimated from the training data. All feature functions are real-valued. We construct a set of real-valued features $b(x,i)$ of the observation to expresses characteristic empirical distribution of the training data that should also hold of the model distribution. An example for POS labeling of such a feature is

$$b(x,i) = \begin{cases} 1 & \text{if the observation at position } i \text{ is the word "September"} \\ 0 & \text{otherwise} \end{cases}$$

Each feature function takes on the value of one of these real-valued observation features $b(x,i)$ the current state (in the case of a state function) or previous and current states (in the case of a transition function).

Finally, notation is simplified by writing

$$s(y_i, x, i) = s(y_{i-1}, y_i, x, i)$$

and

$$F_j(y, x) = \sum_{i=1}^{n} f_j(y_{i-1}, y_i, x, i)$$

where each $f_j(y_{i-1},y_i,x,i)$ is either a state function s or a transition function t. Thus the general CRF formalism is below:

$$P(Y \mid X) \propto \exp\left(\sum_k \lambda_k \cdot F_k(Y, X)\right) \tag{5}$$

We ignore the details for parameter estimation and other aspects about CRF.

Therefore we cast our sub-tree labeling problem as a CRF. By treat the label variable R as the unseen variable Y and the parse tree Π as the observation variable X, we rewrite formula (5) as below:

$$P(R \mid \Pi) \propto \exp\left(\sum_k \lambda_k \cdot F_k(R, \Pi)\right) \tag{6}$$

We also rewrite the feature function F as below:

$$F_j(R,\Pi)=\sum_{j=1}^{n} f_j(R_{k'},R_k,\Pi,k) \tag{7}$$

For the sub tree node π_j, we survey four features in $b(x,i)$ which are likely to affect the reordering: (1) Part of speech tags of the node and child nodes (2) Source word of child nodes (3) Head word of child nodes (4) Height of the node in the tree. We describe them in the following.

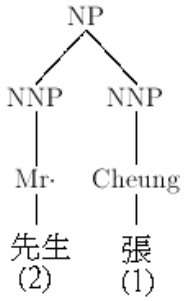

Figure 12: Example sub-tree of “Mr. Cheung” with alignment target word “張 $_1$ 先生 $_2$”

Intuitively, part of speech tags are the basic component of parse trees and may influence the reordering of a node. So we take POS tags of the tree node and its child nodes into consideration. For example, the sub-tree of “Mr. Cheung” is shown in Figure 12, where the POS of this sub-tree is *NP*, with child nodes POS, *NNP* and *NNP*. Second, the source word of child nodes is also a useful feature, for its potential to effect the reordering of the tree node. In Figure 12, the source word of “*Mr.*” of the left child node is a title which often triggers inversion. We also consider head word of child nodes as another feature for the similar reason. We define head words in a straightforward way. Head word of a sub-tree is the head word of the child nodes with the same POS tag as the sub-tree root. If more than one child has such a POS, we choose the right most child node. Finally, we also select height of the node in the tree as a feature. We define the height of leaf word is 1. After we calculate the height of all child nodes, the height of sub-tree is one plus the maximal tree heights of child nodes.

Sometimes the reordering operation of a tree node can be attributed to combination of features. For example, we combine features with source word left child (e.g. "*Mr.*") and *POS* of the right child (e.g. "*NNP*"). In most cases of "Mr." followed by a proper noun, the reordering operation is "2 1". We use many reasonable combinations of features in training, such as POS_{node} / $POS_{left\ child}$ / $POS_{right\ child}$, Source $Word_{left\ child}$ / $POS_{right\ child}$, etc. We list all of our training feature combinations with tree nodes with of 2 or more child nodes in Appendix A.

Since traditional CRF is a simple chain graphical model, we need to reformulate our tree model into a chain model to match the CRF format for training data. In Figure 13, the label for sub-tree (*VP* (*VBD PP PP*)) is condition on parent node *S*, and the three child nodes *VBD*, *PP*, *PP*. To change this tree model, we use a tree-to-chain procedure. We traverse the parse tree in infix order to obtain the infix traversal of nodes, and transform the tree into a sequence of nodes. In Figure 13, we show an example for turning a parse tree into chain form (*NP S VP PP NP NP VP PP*). To compensate for the loss of information related to tree structure, we add features of child nodes to their parent node. An example of conditional random field training data for the parse tree is shown in Table III.

The result of this final training stage is a set of tree nodes, features, and reordering operations. Recall that we preprocess the parallel corpus into annotated parse tree, determining reordering operations, and estimating the probability of reordering operations conditioned on different feature information. Then with the trained model, we can automatically reorder a parse tree *Π* into reordered parse tree *Π'* whose leaf node sequence is closer to the word order in target language. We introduce the runtime process in the next section.

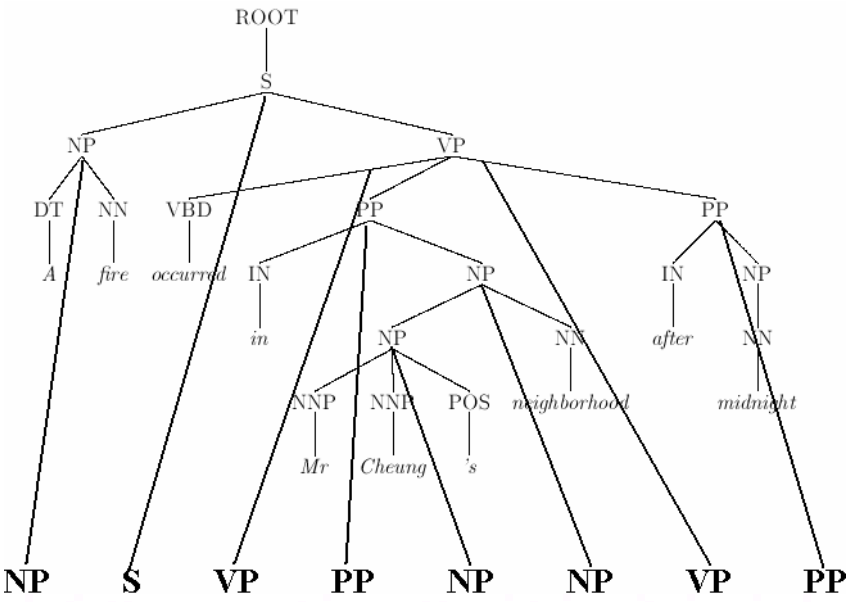

Figure 13: Example for reformulating parse tree with chain form

POS	Head Word	LPOS	LH	LW	RPOS	RH	RW	H	RO
NP	*fire*	DT	*a*	*a*	NN	*fire*	*fire*	*2*	*12*
S	*fire*	NP	*fire*	*N.A.*	VP	*occur*	*N.A.*	*6*	*21*
VP	*occur*	VBD	*occur*	*N.A.*	PP	*in*	*N.A.*	*5*	*321*
PP	*in*	IN	*in*	*in*	NP	*neighborhood*	*N.A.*	*3*	*N.A.*
NP	*Cheung*	NNP	*Mr.*	*Mr.*	POS	*'s*	*'s*	*2*	*213*
NP	*neighborhood*	NP	*Cheung*	*N.A.*	NN	*neighborhood*	*neighborhood*	*3*	*12*
VP	*midnight*	PP	*in*	*N.A.*	PP	*after*	*N.A.*	*4*	*321*
PP	*after*	IN	*after*	*after*	NP	*midnight*	*midnight*	*2*	*21*

Table III : CRF training data from parse tree in Figure 9. *LPOS*: Left child POS; *LH*: Left child head word; *LW*: Left child source word; *RPOS*: Right child POS; *RH*: Right child head word; *RW*: Right child source word; *H*: Tree node height on tree; *RO*: Reordering Operation

3.4 Run-Time Reordering Estimation

We obtain a CRF model for the probability of reorder operation conditioned on features. At run-time, we applied the CRF training model to reorder the given sentence *s*. Our run-time process

is shown in Figure 14.

Procedure ReorderingCorrespond (source sentence E)

ST = parser(E)
for each sub-tree π_j in ST
(1) r_j=labelEstimate (*ST*, F_j)
(2) ST' = ReorderedParseTree(*ST*, r_1, r_2, … r_k)
(3) Return TreeLeafWord(ST')

Figure 14: Evaluating source sentences at run-time

In Step (1), we parse the given source sentence using *PAR* and obtain parse tree *ST*. In Step (2), for each tree node in *ST*, we decompose parse tree *ST* into sub-tree π_i, and determine the best reordering operation r_i. Then in Step (3), we estimate the most possible reordering operation for π_i using the CRF model on the parse tree *ST*, and feature set for π_i. In Step (4), we generate a reordered parse tree *ST'* from *ST* by applying all reorder operations. Finally we visit the leaf node in infix order and return the sequence of words in infix order.

Chapter 4 Experimental Results

Our reordering model was designed to reorder phrases in the parse tree with features and reordering operations. As such, this model will be trained and evaluated over a bilingual corpus. Furthermore, since the goal of our model is to improve fluency of the translation, we evaluate our model based on machine translation metrics. In this section, we first present the details of training a reordering model for the evaluation (Section 4.1). Then, Section 4.2 lists the machine translation systems that we use in our comparison. Section 4.3 introduces the evaluation metrics, BLEU, which is the common metrics standard for machine translation systems. Finally, we describe the test data and evaluation results in Section 4.4.

4.1 Experimental Setup

To process the training data, we used the parser produced by Stanford University Natural Language Processing Group (http://nlp.stnford.edu/software/lex-parser.shtml) to parse English sentences. We also use the Giza++ Version 2 developed by Josef Och (http://www.fjoch.com/GIZA++.html) to obtain word alignment information for the training data.

We used a collection of 94,622 aligned English-Chinese sentences for training. We obtained these sentences from Hong Kong Parallel Corpus (HKPC) with some filtering results in the parameters shown in Table IV. We kept source sentences between *mimSLen* to *MaxSLen* words. For example, we consider “A fire occurred in Mr. Cheung’s neighborhood after midnight” with “午夜 後 張 先生 家 附近 發生 火災” as an alignment sentence from the parallel corpus. Therefore we check if the length of English sentence, 9, is between *mimSLen* to *maxSLen*. And we also kept

only those sentences with most words aligned with a translation with a ratio higher than *minAratio*, and with an alignment rate higher than the nouns, verbs, or adjectives words aligned ratio *mimNVJraitio*. Figure 15 shows a sample parse tree. The alignment ratio for this example is 8/10 over all and 6/6 for content words. We also left out the sentences with independent clauses. This method for selecting sentences can produce many parse trees, potentially including a significant number of parse trees that are too specific to be widely acceptable. We need high accuracy and ratio in word alignment for training, because the order for each node is determined by position of alignment target words. If parse tree is presented with too many null alignment leaf, we would not know the order in many tree nodes. Table V shows some samples of the parallel sentences after the filtering process.

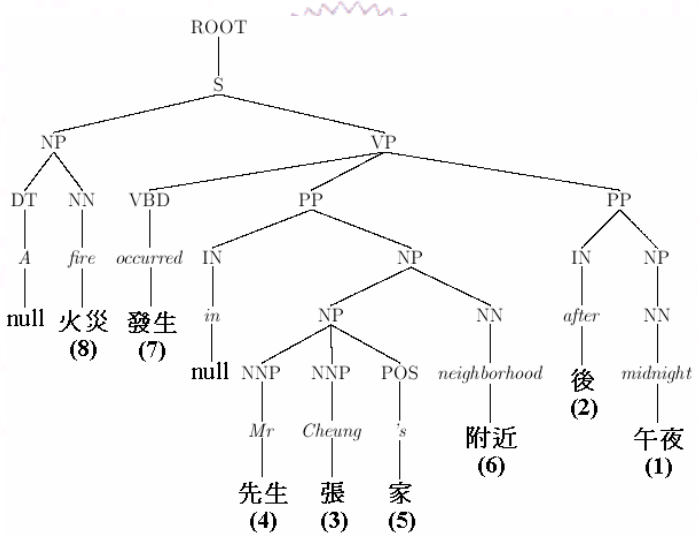

Figure 15: Parse tree annotated with target words and positions. Source sentence is "*A fire occurred in Mr. Cheung's neighborhood after midnight*" with target sentence "*午夜*$_1$ *後*$_2$ *張*$_3$ *先生*$_4$ *家*$_5$ *附近*$_6$ *發生*$_7$ *火災*$_8$"

Paramter	Value	Description
minSLen	5	Mim. length of English sentences
maxSLen	20	Max. length of English sentences
minAratio	0.5	Min. alignment ratio from English words to Chinese words
minNVJAratio	0.5	Min. alignment ratio with part-of-speech tag Noun, Verb, Adjective from English words to Chinese words

Table IV: Training Parameters

Date	Corpora	English Sentence	Chinese Sentence
12/21/1999	Hong Kong News	This arrangement also applies to persons holding valid travel documents of the Mainland.	這項安排亦適用於持有效旅行證件的內地人士。
08/09/2002	Hong Kong News	They will appear in Kowloon City Magistrates' Courts later today.	他們將於今日稍後在九龍城裁判法院應訊。
07/25/1990	Hong Kong Legislative Council	This has increased the cost more rapidly than that of other public services.	因此,醫療衛生服務比其他公共服務的成本增加更快。
06/29/1994	Hong Kong Legislative Council	Mr. PATTEN has been sending different signals to different audiences.	彭定康先生一直向不同聽眾發出不同的訊息。

Table V: Sample sentences from the training data

For training Conditional Random Fields parameters, we used CRF++ implemented by Taku Kudo (http://crfpp.sourceforge.net/). CRF++ is a simple and customizable implement of Conditional Random Fields for segmenting/labeling sequential data. By submitting our training data and feature definition to CRF++, we can train the parameters with high accuracy. After training, we

submit our testing data to the same tool and the most possible labels of reordering operation for each data will be produced.

4.2 Determine the Reordering Operations

Recall that our model starts with a source sentence, and reorder the parse tree nodes. The output of our model is a reordered English sentence closer to the target language structure. Finally, the output sentences are submitted to a state-of-the-art phrase base machine translation system such as Google Translate and Pharaoh for final translation.

Our experimental evaluation focuses on two common machine translation systems, Google Translate and Pharaoh. We compare the results produces by each of these systems with or without syntactical reordering using our model.

4.3 Testing Data and Evaluation Metrics

Machine translation systems are usually compared based on the quality of the translated sentence sets. This quality is often measured using the BLEU score(Kishore Papineni, etc. 2002). We are given a set of translated target sentences with a reference sentence for each source sentence in order to calculate the BLEU score.

We select 1,000 parallel sentences from Hong Kong Corpus as the test data. Length of these English sentences range from 5 to 20 words and sentences with independent clause are removed. Sample of the test data is included in Appendix A.

4.4 Evaluation Results

We report the results of the experimental evaluation using the data and methodology described in the previous section. First we report the results of compared machine translation systems and our model. Then we present of comparing the results of using different machine learning method with Conditional Random Fields and Maximal Entropy. And the results of feature selection differences are presented finally.

Table VI gives the result for the baseline and reordered model with different machine translation systems

	Pharaoh	**Google**
Baseline	19.97	24.67
Reordering	21.09	23.29
Gain	+1.12	-1.38

Table VI: BLEU score of the baseline and reordered model

As shown in the Table VI, reordering model is able to improve the BLEU score by 1.12 points for the Pharaoh system, but loss 1.38 points in Google Translate.

We also compare the influence of different machine learning method between tree-to-chain CRF model and Maximal Entropy. In training, CRF model refers to features of previous and subsequence states while Maximal Entropy only refers to the current state. Table VII shows the result of these two methods.

	Pharaoh	Google
CRF	21.09	23.29
ME	21.02	23.16
Diff	0.07	0.13

Table VII: BLEU score of Conditional Random Field and Maximal Entropy

We compare the BLEU score between these two machine learning methods. The differences are not very notable. We also evaluated the performance when different sets of feature were used. The results are shown in Table VIII.

Feature Pairs	Pharaoh	Google
POS	20.92	23.16
POS + Head Word	21.09	23.22
POS + Head Word + Source Word	21.09	23.29
POS + Head Word + Source Word + Tree Height	20.99	23.28

Table VIII: BLEU score of different features

As Table VIII shows, BLEU score improve after adding head word feature, both for Pharaoh and Google Translate. But the addition of source word has mixed results for Pharaoh and Google Translate. Finally, feature for tree height decrease the BLEU score.

Chapter 5 Conclusion and Future Work

As the results shown in Chapter Table VI, the BLEU score is increased by using statistical syntactic reordering model with Pharaoh, but decreased when working with Google Translate. It may be the training data we use is different from data set in Google Translate, but the training data for Pharaoh is the same with our statistical syntactic reordering model. And Google Translate may be very sensitive to the form of the source sentence. For example, the result of English sentence "the existing provision in section 80 of the district court ordinance is loosely word" with reference Chinese translated sentence "地 方 法 院 條 例 第 80 條 的 現 有 條 文 用 字 不 夠 嚴 密" is shown in Table IX. Our statistical syntactic reorder model reorder source sentence phrases into ordering closer to target sentence, but in Pharaoh, the ordering in translated sentence is still the same in the source sentence. Thus the BLEU score of Pharaoh with syntactic reordering is higher than using Pharaoh alone. Although the reordering of the source sentence we are given to Google Translate is closer to the target sentence, the translation of Google Translate is worse for some reason. For instance, the part of "in of the" should be translated at all while Google Translate produces "在向" in the target sentence. Other examples of evaluation results are included in Appendix B.

In Table VII, it is shown that the difference between Conditional Random Field and Maximal Entropy is not significant. Our possible reason may be that the performance of CRF and Maximal Entropy on learning syntactical reordering not very different. And our tree-to-chain procedure seems not to help too much. If we can alter the tree-to-chain procedure to keep more tree structure information of CRF may improve the BLEU score.

		Pharaoh	Google
Baseline	*Source*	the exist provision in section 80 of the district court ordinance be loosely word .	
	Target	現時的條文在第八成的地區法院被寬鬆的。該條例	現行條文第80條地方法院條例很鬆散的措辭.
	BLEU	11.62	32.37
Reordered	*Source*	in of the district court ordinance section 80 the exist provision be loosely word .	
	Target	在區域法院條例第百分之八十的現有條文是說，鬆散的。	在向地方法院條例第80條的現行規定，也很鬆散的措辭.
	BLEU	28.88	40.19
Reference Target		地方法院條例第80條的現有條文用字不夠嚴密。	

Table IX: The result of baseline model and our SSR model for “the exist provision in section 80 of the district court ordinance be loosely word ” translated into “地方法院條例第80條的現有條文用字不夠嚴密。”

We compare the efficiency of different features in Table VIII. Although the addition of head word feature improve the BLEU score of the translated sentences, leaf word feature is not useful for both Pharaoh and Google Translation. In training, we experimented with different features. However, there are too many features combinations leading to data sparseness problem. Furthermore, we only use 94,622 English sentences with highly reliable alignment information. The problem of data sparseness may hurt the BLEU score.

Future research direction presents them. Currently, the training source sentences we used in evaluation are in the original order. Thus the data for word alignment and machine translation system is not the same. If we adjust the translation model and language model in machine translation systems, the effectiveness of machine translation approach with our reordered model may increase. Another future research direction would be only reorder tree nodes higher up 1. The reordering of base phrases seems work better using the phrase-based SMT directly. We probable should apply syntactic reordering only for global ordering, and the order of base phrases should be left for the underlying statistical machine translation systems.

In summary, we have introduced a method for learning syntactical reordering of source sentences that improves the translation quality when working a phrase-based machine translation system. The method involves parsing source sentences, determining reordering operation for tree nodes, estimating probability for training data, and finally reordering the given sentence with the trained model. We have implemented and evaluated the method as applied to state-of-the-art machine translation systems. In evaluation, we have shown that the method with machine translation approach outperforms the underlying pure machine translation approach. In addition, the proposed model can integrate easily with other machine translation approach.

Reference

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. `A Statistical Approach to Machine Translation`, in *Computational Linguistics*, volume 16(2): 79-85.

David Chiang. 2005. `A Hierarchical Phrase-Based Model for Statistical Machine Translation`, in *Proceedings of Association for Computational Linguistics* (*ACL-05*) pages 263-270

Dan Klein, and Christopher D. Manning. 2002. `Fast Exact Inference with a Factored Model for Natural Language Parsing`, in *Advances in Neural Information Processing Systems 15.*

Dan Klein, and Christopher D. Manning. 2003. `Accurate Unlexicalized Parsing`, in *Proceedings of 41st Meeting of the Association for Computational Linguistics* (ACL-03).

Kevin Knight. 1999. `Decoding Complexity in Word-Replacement Translation Models`, in *Computational Linguistics*, volume 25(4): 607-615.

Philipp Koehn. 2004. `Pharaoh: a Beam Search Decoder for Phrase-Based Statistical Machine Translation Models`, in *Proceeding of the Sixth Conference of the Association for Machine Translation in the Americas*, pages 115-124.

Philipp Koehn, 2006. `MOSES: A beam search decoder for factored phrase-based statistical translation models`. http://www.statmt.org/moses/.

John Lafferty, Andrew McCallum, Fernando Pereira. 2001. `Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data`, in *Proceedings of International Conference on Machine Learning* (ICML-01), pages 591-598

Daniel Marcu, Wei Wang, Abdessamad Echihabi, Kevin Knight. 2006. `SPMT: Statistical Machine Translation with Syntactified Target Language Phrases` in *Proceedings of EMNLP*, pages 44-52, Sydney, Australia.

Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. 2001. `BLEU: a Method for Automatic Evaluation of Machine Translation` in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (*ACL-012*), pages 311-318.

Chao Wang, Michael Collins, Philipp Koehn. 2007. `Chinese Syntactic Reordering for Statistical Machine Translation` in *Proceedings of EMNLP*, Prague, Czech Republic.

Kenji Yamada, Kevin Knight. 2001. `A syntax-based Statistical Translation Model` in *Proceedings of the 39th Annual Conference of the Association for Computational Linguistics* (*ACL*-01).

Hanna M. Wallach. 2004. `Conditional Random Fields: An Introduction`, in *Technical Report MS-CIS-04-21, Department of Computer and Information Science, University of Pennsylvania.*

Dekai Wu. 1997. `Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora` in *Computational Linguistics*, volume 23(3): 377-403

Appendix A - Features List

In the following table, the field *Feature No.* stands for the feature number for training data and *Features List* stands for the features used in this data combination. The indices number means this features refer to the position of the feature on the chain (0 for itself; 1 for next state; -1 for previous state).

Features of Conditional Random Field for Tree Nodes with 2 or more than 4 child nodes

Feature No.	***Features List***
01	POS_0
02	Head $Word_0$
03	$POS_{Left\ Child,\ 0}$
04	Source $Word_{Left\ Child,\ 0}$
05	Head $Word_{Left\ Child,\ 0}$
06	$POS_{Right\ Child,\ 0}$
07	Source $Word_{Right\ Child,\ 0}$
08	Head $Word_{LeftChild,\ 0}$
09	Tree $Height_0$
10	POS_0 / $POS_{Left\ Child,\ 0}$ / $POS_{Right\ Child,\ 0}$
11	POS_0 / $POS_{Left\ Child,\ 0}$ / Source $Word_{Right\ Child,\ 0}$
12	POS_0 / $POS_{Left\ Child,\ 0}$ / Head $Word_{LeftChild,\ 0}$
13	POS_0 / Source $Word_{Right\ Child,\ 0}$ / $POS_{Right\ Child,\ 0}$
14	POS_0 / Head $Word_{Left\ Child,\ 0}$ / $POS_{Right\ Child,\ 0}$
15	$POS_{Left\ Child,\ 0}$ / $POS_{Right\ Child,\ 0}$
16	POS_0 / Tree $Height_0$
17	POS_0 / $POS_{Left\ Child,\ 0}$ / $POS_{Right\ Child,\ 0}$ / Tree $Height_0$
18	$POS_{Left\ Child,\ 0}$ / $POS_{Right\ Child,\ 0}$ / Tree $Height_0$
19	POS_1
20	Head $Word_1$
21	$POS_{Left\ Child,\ 1}$

22	$\text{Source Word}_{\text{Left Child, 1}}$
23	$\text{Head Word}_{\text{Left Child, 1}}$
24	$\text{POS}_{\text{Right Child, 1}}$
25	$\text{Source Word}_{\text{Right Child, 1}}$
26	$\text{Head Word}_{\text{LeftChild, 1}}$
27	Tree Height_{1}
28	POS_{1} / $\text{POS}_{\text{Left Child, 1}}$ / $\text{POS}_{\text{Right Child, 1}}$
29	POS_{1} / $\text{POS}_{\text{Left Child, 1}}$ / $\text{Source Word}_{\text{Right Child, 1}}$
31	POS_{1} / $\text{POS}_{\text{Left Child, 1}}$ / $\text{Head Word}_{\text{LeftChild, 1}}$
31	POS_{1} / $\text{Source Word}_{\text{Right Child, 1}}$ / $\text{POS}_{\text{Right Child, 1}}$
32	POS_{1} / $\text{Head Word}_{\text{Left Child, 1}}$ / $\text{POS}_{\text{Right Child, 1}}$
33	$\text{POS}_{\text{Left Child, 1}}$ / $\text{POS}_{\text{Right Child, 1}}$
34	POS_{1} / Tree Height_{1}
35	POS_{1} / $\text{POS}_{\text{Left Child, 1}}$ / $\text{POS}_{\text{Right Child, 1}}$ / Tree Height_{1}
36	$\text{POS}_{\text{Left Child, 1}}$ / $\text{POS}_{\text{Right Child, 1}}$ / Tree Height_{1}
37	$\text{POS}_{\text{-1}}$
38	$\text{Head Word}_{\text{-1}}$
39	$\text{POS}_{\text{Left Child, -1}}$
40	$\text{Source Word}_{\text{Left Child, -1}}$
41	$\text{Head Word}_{\text{Left Child, -1}}$
42	$\text{POS}_{\text{Right Child, -1}}$
43	$\text{Source Word}_{\text{Right Child, -1}}$
44	$\text{Head Word}_{\text{LeftChild, -1}}$
45	$\text{Tree Height}_{\text{-1}}$
46	$\text{POS}_{\text{-1}}$ / $\text{POS}_{\text{Left Child, -1}}$ / $\text{POS}_{\text{Right Child, -1}}$
47	$\text{POS}_{\text{-1}}$ / $\text{POS}_{\text{Left Child, -1}}$ / $\text{Source Word}_{\text{Right Child, -1}}$
48	$\text{POS}_{\text{-1}}$ / $\text{POS}_{\text{Left Child, -1}}$ / $\text{Head Word}_{\text{LeftChild, -1}}$
49	$\text{POS}_{\text{-1}}$ / $\text{Source Word}_{\text{Right Child, -1}}$ / $\text{POS}_{\text{Right Child, -1}}$
50	$\text{POS}_{\text{-1}}$ / $\text{Head Word}_{\text{Left Child, -1}}$ / $\text{POS}_{\text{Right Child, -1}}$
52	$\text{POS}_{\text{Left Child, -1}}$ / $\text{POS}_{\text{Right Child, -1}}$
52	$\text{POS}_{\text{-1}}$ / $\text{Tree Height}_{\text{-1}}$
53	$\text{POS}_{\text{-1}}$ / $\text{POS}_{\text{Left Child, -1}}$ / $\text{POS}_{\text{Right Child, -1}}$ / $\text{Tree Height}_{\text{-1}}$
54	$\text{POS}_{\text{Left Child, -1}}$ / $\text{POS}_{\text{Right Child, -1}}$ / $\text{Tree Height}_{\text{-1}}$

Features of Conditional Random Field for Tree Nodes with 3 child nodes

Feature No.	Features List
01	POS
02	Head Word
03	Tree Height
04	POS_{Parent}
05	$Head\ Word_{Parent}$
06	$Tree\ Height_{Parent}$
07	$POS_{Left\ Child}$
08	$Source\ Word_{Left\ Child}$
09	$Head\ Word_{Left\ Child}$
10	$Tree\ Height_{Left\ Child}$
11	$POS_{Center\ Child}$
12	$Source\ Word_{Center\ Child}$
13	$Head\ Word_{Center\ Child}$
14	$Tree\ Height_{Center\ Child}$
15	$POS_{Right\ Child}$
16	$Source\ Word_{Right\ Child}$
17	$Head\ Word_{Right\ Child}$
18	$Tree\ Height_{RightChild}$
19	POS / $POS_{Left\ Child}$ / $POS_{Center\ Child}$ / $POS_{Right\ Child}$
20	Head Word / $POS_{Left\ Child}$ / $POS_{Center\ Child}$ / $POS_{Right\ Child}$
21	POS / $Head\ Word_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
22	Head Word / $Head\ Word_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
23	POS / $Head\ Word_{Left\ Child}$ / $POS_{Center\ Child}$ / $POS_{Right\ Child}$
24	POS / $POS_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $POS_{Right\ Child}$
25	POS / $POS_{Left\ Child}$ / $POS_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
26	POS / $POS_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
27	POS / $Head\ Word_{Left\ Child}$ / $POS_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
28	POS / $Head\ Word_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $POS_{Right\ Child}$
29	POS / $Head\ Word_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
30	Head Word / $Head\ Word_{Left\ Child}$ / $POS_{Center\ Child}$ / $POS_{Right\ Child}$
31	$Head\ Word_{t}$ / $POS_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $POS_{Right\ Child}$
32	Head Word / $POS_{Left\ Child}$ / $POS_{Center\ Child}$ / $Head\ Word_{Right\ Child}$
33	Head Word / $POS_{Left\ Child}$ / $Head\ Word_{Center\ Child}$ / $Head\ Word_{Right\ Child}$

34	Head Word / Head Word$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / Head Word$_{\text{Right Child}}$
35	Head Word / Head Word$_{\text{Left Child}}$ / Head Word$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
36	Head Word / Head Word$_{\text{Left Child}}$ / Head Word$_{\text{Center Child}}$ / Head Word$_{\text{Right Child}}$
37	POS / Source Word$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
38	POS / POS$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
39	POS / POS$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
40	POS / POS$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
41	POS / Source Word$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
42	POS / Source Word$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
43	POS / Source Word$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
44	Head Word / Source Word$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
45	Head Word / POS$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
46	Head Word / POS$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
47	Head Word / POS$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
48	Head Word / Source Word$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
49	Head Word / Source Word$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$
50	Head Word / Source Word$_{\text{Left Child}}$ / Source Word$_{\text{Center Child}}$ / Source Word$_{\text{Right Child}}$
51	POS$_{\text{Parent}}$ / POS
52	POS$_{\text{Parent}}$ / Head Word
53	Head Word$_{\text{Parent}}$ / POS
54	Head Word$_{\text{Parent}}$ / Head Word
55	POS$_{\text{Parent}}$ / POS / Tree Height
56	POS / POS$_{\text{Left Child}}$ / POS$_{\text{Center Child}}$ / POS$_{\text{Right Child}}$ / Tree Height
57	POS / Tree Height

Appendix B - Samples of Results

	Result Sentences
Original English	This international practice is followed in Hong Kong.
Reordered English	This international practice is followed in Hong Kong.
Original Chinese	本港目前正是採用這種國際上慣用的方法。
Original Pharaoh	這些國際守則是根據在香港的。
Reordered Pharaoh	這些國際守則是根據在香港的。
Original Google	這個國際慣例,其次是香港.
Reordered Google	這個國際慣例,其次是香港.
Original English	Firstly, the maintenance of the stability and prosperity of Hong Kong ranks above other considerations.
Reordered English	Firstly, of of Hong Kong the stability and prosperity the maintenance ranks above other considerations.
Original Chinese	第一，維持本港的繁榮和穩定是最重要的，比其他事情更為重要。
Original Pharaoh	第一，維持的穩定和繁榮的香港職級以上的考慮，香港的。
Reordered Pharaoh	第一，香港的安定和繁榮的維修職級以上的考慮，香港的。
Original Google	第一,維持繁榮和穩定,香港排名高於其他因素.
Reordered Google	第一,香港的穩定和繁榮,維護凌駕於其他因素.
Original English	All once important have collapsed.
Reordered English	All once important have collapsed.

Original Chinese	所有這些曾盛極一時的工業都已一蹶不振。
Original Pharaoh	首先，必須有崩潰。
Reordered Pharaoh	首先，必須有崩潰。
Original Google	一旦所有重要倒塌.
Reordered Google	一旦所有重要倒塌.
Original English	There were 12 cases in the past 12 months.
Reordered English	There were in the past 12 months 12 cases.
Original Chinese	在過去十二個月內總共有 12 宗。
Original Pharaoh	我們是十二宗在過去的十二月。
Reordered Pharaoh	我們在過去十二個月的是，。
Original Google	共有 12 例,在過去的 12 個月.
Reordered Google	有在過去的 12 個月 12 例.
Original English	Some of these difficulties are still unresolved.
Reordered English	Some of these difficulties are still unresolved.
Original Chinese	到目前為止，有些困難仍有待解決。
Original Pharaoh	這部分的問題是可以解決。
Reordered Pharaoh	這部分的問題是可以解決。
Original Google	有些困難仍然沒有得到解決.
Reordered Google	有些困難仍然沒有得到解決.

Original English	The average utilisation rate of each of these community halls varies.
Reordered English	of each of these community halls The average utilisation rate varies.
Original Chinese	每個社區會堂都有不同的平均使用率。
Original Pharaoh	的平均使用率低的分別的，大會堂的不同。
Reordered Pharaoh	所有這些社區會堂的平均使用率高，不同的。
Original Google	平均使用率,每個社區會堂各異.
Reordered Google	每個社區會堂的平均使用率是不同的.
Original English	- Nuclear Island civil works.
Reordered English	- Nuclear Island civil works.
Original Chinese	核能島土木工程。
Original Pharaoh	—核能島的工作。
Reordered Pharaoh	—核能島的工作。
Original Google	-核島土木工程.
Reordered Google	-核島土木工程.
Original English	Sir, the above are my remarks.
Reordered English	Sir, the above are my remarks.
Original Chinese	主席先生，本人謹此陳辭。
Original Pharaoh	先生，以上是我的意見。
Reordered Pharaoh	先生，以上是我的意見。
Original Google	主席先生,以上是我的話.

Reordered Google	主席先生,以上是我的話.
Original English	I will prefer the last option.
Reordered English	I will prefer the last option.
Original Chinese	我選擇最後者。
Original Pharaoh	我想提出的最後建議。
Reordered Pharaoh	我想提出的最後建議。
Original Google	我會選擇最後.
Reordered Google	我會選擇最後.
Original English	Sir, with these remarks, I support the motion.
Reordered English	Sir, with these remarks, I support the motion.
Original Chinese	主席先生，我謹此陳辭，支持動議。
Original Pharaoh	先生，謹此陳辭，支持我的議案。
Reordered Pharaoh	先生，謹此陳辭，支持我的議案。
Original Google	主席先生,我謹此陳辭,支持這項議案.
Reordered Google	主席先生,我謹此陳辭,支持這項議案.
Original English	Daya Bay must be both safe and competitive.
Reordered English	Daya Bay must be both safe and competitive.
Original Chinese	大亞灣核電廠必須既安全又收費不高。
Original Pharaoh	大亞灣要無論是安全和競爭力。
Reordered Pharaoh	大亞灣要無論是安全和競爭力。

Original Google	大亞灣必須既安全又具競爭力.
Reordered Google	大亞灣必須既安全又具競爭力.
Original English	Over the years, Hong Kong has devoted huge amount of resources on public housing programmes.
Reordered English	Over the years, Hong Kong has on public housing programs devoted huge amount of resources.
Original Chinese	長期以來，本港已在公屋計劃上投入極為龐大的資源。
Original Pharaoh	過去數年，香港已投入大量資源在公共房屋計劃。大量的香港。
Reordered Pharaoh	過去數年，香港公共房屋計劃投入龐大的資源。金額已在香港的
Original Google	多年來,香港已投入大量資源,公共房屋計劃.
Reordered Google	多年來,香港的公營房屋計劃投入大量的資源.
Original English	Localisation Sir, I now turn to the Civil Service.
Reordered English	Localisation Sir, I now turn to the Civil Service.
Original Chinese	主席先生，現在我轉而談到公務員的問題。
Original Pharaoh	化先生，我現在要談談公務員的服務。
Reordered Pharaoh	化先生，我現在要談談公務員的服務。
Original Google	本地化主席先生,現在我想談談公務員的問題.
Reordered Google	本地化主席先生,現在我想談談公務員的問題.
Original English	Firstly, we are engaged in a programme of consultancy studies of the main manufacturing industries.
Reordered English	Firstly, we are engaged in of of the main manufacturing industries consultancy studies a program.

Original Chinese	第一個方式是對主要的製造工業進行顧問研究計劃。
Original Pharaoh	首先，我們正進行一項顧問研究的主要製造行業，政府在計劃。
Reordered Pharaoh	首先，我們正進行的主要製造行業顧問研究一項計劃，在之一。
Original Google	首先,我們正在從事一項方案的顧問研究的主要製造業.
Reordered Google	首先,我們所從事的各主要製造行業顧問研究的一個方案.
Original English	I turn now to some public finance and monetary matters.
Reordered English	I turn now to some public finance and monetary matters.
Original Chinese	我現在轉談一些公共財政及金融事項。
Original Pharaoh	我現在要談談一些公共財政及金融事務。
Reordered Pharaoh	我現在要談談一些公共財政及金融事務。
Original Google	我現在談一些公共財政和金融事務.
Reordered Google	我現在談一些公共財政和金融事務.
Original English	These are sufficient at present.
Reordered English	These are at present sufficient.
Original Chinese	目前，這些設施是足夠的。
Original Pharaoh	這些是足夠的意見。
Reordered Pharaoh	這些都是在現時足夠。
Original Google	這是足夠的.
Reordered Google	這些都是目前已經足夠.

Original English	Indeed, many members of the media have made this point to me personally.
Reordered English	Indeed, of the media many members have to me this point personally made.
Original Chinese	許多傳媒從業員亦曾私下向我表達這個觀點。
Original Pharaoh	其實，很多議員已提出這一點，我認為，傳媒的。
Reordered Pharaoh	其實，很多議員已向我這個問題，提出我的修正案。
Original Google	事實上,許多媒體成員也指出了這一點對我個人.
Reordered Google	事實上,對媒體的許多議員都向我這點,親自作了.
Original English	The answer is as simple as that.
Reordered English	The answer is as simple as that.
Original Chinese	答案便是這麼簡單。
Original Pharaoh	答案很簡單的說，如的
Reordered Pharaoh	答案很簡單的說，如的
Original Google	答案就是這麼簡單.
Reordered Google	答案就是這麼簡單.
Original English	In my view, the tax should be collected by equal half yearly instalments payable in arrears.
Reordered English	In my view, the tax should be by equal half yearly instalments payable in arrears collected.
Original Chinese	我認為入息稅應在課稅年度後繳交，分兩期繳付，每半年繳交一次。
Original Pharaoh	我認為，應該是平等的供款的欠薪。半數的收取的稅款。

Reordered Pharaoh	我認為，應該是平等的供款的意見。欠款由一半的稅項。
Original Google	在我看來,稅制應收取的平等半年分期繳納欠款.
Reordered Google	在我看來,稅制應通過平等半年分期繳納拖欠會費的收集.
Original English	This issue really deserves our deep consideration.
Reordered English	This issue really deserves our deep consideration.
Original Chinese	這個問題實在值得我們深思。
Original Pharaoh	這個問題是值得我們深入考慮。
Reordered Pharaoh	這個問題是值得我們深入考慮。
Original Google	這個問題很值得我們深思.
Reordered Google	這個問題很值得我們深思.
Original English	However, we must act and react to our real situation.
Reordered English	However, we must act and react to our real situation.
Original Chinese	然而，我們必須採取行動，並對實際情況作出反應。
Original Pharaoh	不過，我們必須審慎及迅速地的實際情況。
Reordered Pharaoh	不過，我們必須審慎及迅速地的實際情況。
Original Google	但是,我們必須採取行動,並實際情況作出反應.
Reordered Google	但是,我們必須採取行動,並實際情況作出反應.
Original English	These actions will continue.
Reordered English	These actions will continue.
Original Chinese	上述措施將繼續實行。

Original Pharaoh	這些行動會持續。
Reordered Pharaoh	這些行動會持續。
Original Google	這些行動將繼續進行.
Reordered Google	這些行動將繼續進行.
Original English	Any compromise will mean an infringement of one function upon the other.
Reordered English	Any compromise will mean an infringement of one function upon the other.
Original Chinese	任何妥協都會侵犯二者個別的功能。
Original Pharaoh	任何影響會是一個侵犯的人，在該等。
Reordered Pharaoh	任何影響會是一個侵犯的人，在該等。
Original Google	任何妥協都將意味著侵權人的功能.
Reordered Google	任何妥協都將意味著侵權人的功能.
Original English	Sir, I support the motion.
Reordered English	Sir, I support the motion.
Original Chinese	主席先生，本人支持動議。
Original Pharaoh	先生，我支持這項動議。
Reordered Pharaoh	先生，我支持這項動議。
Original Google	主席先生,我支持這項議案.
Reordered Google	主席先生,我支持這項議案.

Original English	In addition, further exemptions may be allowed by regulations made under clause 37.
Reordered English	In addition, further exemptions may be by under clause 37 made regulations allowed.
Original Chinese	此外，根據條例第 37 條制定的規例，亦容許進一步給予豁免。
Original Pharaoh	此外，更可獲豁免令的規定，根據第百分之三十七。在
Reordered Pharaoh	此外，亦可能是由根據第３７作出規定，可豁免。
Original Google	此外,還可能豁免准許規例第 37 條.
Reordered Google	此外,還可能豁免受第 37 條規定作出的允許.
Original English	Even the Tourist Association is largely financed from the hotel accommodation tax.
Reordered English	Even the Tourist Association is largely financed from the hotel accommodation tax.
Original Chinese	即使是旅遊協會，其經費主要也是來自酒店房租稅。
Original Pharaoh	旅遊協會是主要的酒店房租稅，從財務的。
Reordered Pharaoh	旅遊協會是主要的酒店房租稅，從財務的。
Original Google	即使是旅遊協會,其經費主要來自酒店房租稅.
Reordered Google	即使是旅遊協會,其經費主要來自酒店房租稅.
Original English	We have already this afternoon touched on the question of child abuse.
Reordered English	We have already this afternoon touched on the question of abuse child.
Original Chinese	這個下午我們已談過虐待兒童的問題。
Original Pharaoh	我們亦已就此問題提出討論的問題的兒童濫用。

Reordered Pharaoh	我們亦已就此問題提出討論的問題虐待的兒童。
Original Google	我們已經在今天下午談到虐待兒童的問題.
Reordered Google	這個下午我們已觸及有關虐待兒童.
Original English	Commercial crime Sir, a related subject is the prevalence of commercial crimes in recent years.
Reordered English	Commercial crime Sir, a related subject is of commercial crimes in recent years the prevalence.
Original Chinese	主席先生，另一個有關的問題，是近年本港的商業罪案十分猖獗。
Original Pharaoh	商業罪案先生，一個是一般的商業罪案在過去三年，有關問題。
Reordered Pharaoh	商業罪案先生，關於問題的商業罪案在過去兩年的蔚然成風，是一項
Original Google	商業罪案先生,與此相關的問題,是普遍存在的商業罪案近年.
Reordered Google	商業罪案先生,一個相關的主題是商業犯罪近年來盛行.
Original English	They may be few, but deadly.
Reordered English	They may be few, but deadly.
Original Chinese	這類存心不良的經營者可能為數不多，但卻能致命。
Original Pharaoh	他們可能是少數，但致命傷。
Reordered Pharaoh	他們可能是少數，但致命傷。
Original Google	他們可能是少數,但致命.
Reordered Google	他們可能是少數,但致命.
Original English	With my remarks, Sir, I support the motion.
Reordered English	With Sir, my remarks, I support the motion.

Original Chinese	主席先生，我謹此陳辭，支持動議。
Original Pharaoh	與此陳辭，支持我的議案。先生，
Reordered Pharaoh	與先生，我發言支持的，我動議。
Original Google	我的陳詞,我支持這項議案.
Reordered Google	與先生,我的話,我支持這項議案.
Original English	Sir, the new duty structure comes into effect this afternoon by virtue of a Public Revenue Protection Order.
Reordered English	Sir, the new duty structure comes into effect this afternoon by virtue of a Public Revenue Protection Order.
Original Chinese	主席先生，根據保障公共收入令的規定，新稅制在今天下午起生效。
Original Pharaoh	因此，新架構來考慮這問題的一個公共收入保障令，根據所影響的。
Reordered Pharaoh	因此，新架構來考慮這問題的一個公共收入保障令，根據所影響的。
Original Google	主席先生,新稅制在今天下午起生效憑藉一項公共收入保障令.
Reordered Google	主席先生,新稅制在今天下午起生效憑藉一項公共收入保障令.
Original English	Trade unions posed an even greater problem.
Reordered English	Trade unions an even greater problem posed.
Original Chinese	工會問題更大。
Original Pharaoh	職工盟提出一個更大的問題。
Reordered Pharaoh	工會聯盟作出更大的問題存在。
Original Google	工會構成一個更大的問題.

Reordered Google	工會一個更大的問題所造成.
Original English	Blocks are regularly patrolled by uniformed estate caretakers.
Reordered English	Blocks are regularly patrolled by uniformed estate caretakers.
Original Chinese	此外，穿著制服的屋管理員亦定時巡邏各座大廈。
Original Pharaoh	我們是經常巡邏屋的軍裝管理員。
Reordered Pharaoh	我們是經常巡邏屋的軍裝管理員。
Original Google	磚定時巡邏軍警屋管理員.
Reordered Google	磚定時巡邏軍警屋管理員.
Original English	Government, subvented and private hospitals also provide facilities for testing.
Reordered English	Government, subvented and private hospitals also provide for testing facilities.
Original Chinese	各政府醫院、補助醫院和私家醫院亦有提供檢驗設施。
Original Pharaoh	政府補助和私家醫院提供設施為測試。此外，
Reordered Pharaoh	政府補助和私家醫院亦提供的設施。測試，
Original Google	政府補助和私立醫院也提供檢驗設施.
Reordered Google	政府補助和私立醫院也提供測試設施.
Original English	The authority's greater operational freedom will not reduce its public accountability.
Reordered English	The authority's greater operational freedom will not reduce its public accountability.
Original Chinese	房委會在運作方面有更大的自由，並不表示其對公眾的責任會減少。

Original Pharaoh	該委員會的運作會否減少對市民負責。自由。
Reordered Pharaoh	該委員會的運作會否減少對市民負責。自由。
Original Google	管理局的運作方面有更大的自由,也不會減少對市民的問責性.
Reordered Google	管理局的運作方面有更大的自由,也不會減少對市民的問責性.
Original English	I sympathise with their frustration and anger.
Reordered English	I sympathise with their frustration and anger.
Original Chinese	我對該局的失望及憤怒，亦深表同情。
Original Pharaoh	我同意他們的不滿和憤怒。
Reordered Pharaoh	我同意他們的不滿和憤怒。
Original Google	我非常同情他們的挫折感和憤怒.
Reordered Google	我非常同情他們的挫折感和憤怒.
Original English	This is not just a matter of academic interest.
Reordered English	This is not just a matter of academic interest.
Original Chinese	這提議並不是憑空想像出來的。
Original Pharaoh	這並不是一個問題的關注。學者。
Reordered Pharaoh	這並不是一個問題的關注。學者。
Original Google	這不僅僅是一個學術興趣.
Reordered Google	這不僅僅是一個學術興趣.
Original English	We speak often on prosperity and stability.
Reordered English	We often speak on prosperity and stability.

Original Chinese	我們時常談及繁榮和安定問題。
Original Pharaoh	我們也經常就繁榮和穩定。
Reordered Pharaoh	我們經常說的繁榮和穩定。
Original Google	我們經常談到對繁榮和穩定.
Reordered Google	我們常常就繁榮和穩定.
Original English	Meanwhile Hong Kong once again faces the prospect of increasing numbers of boat people arriving on our shores.
Reordered English	Meanwhile Hong Kong once again faces the prospect of increasing numbers of arriving on our shores boat people.
Original Chinese	在此期間，香港又再面對抵港越南難民數目日增的問題。
Original Pharaoh	而香港又再面對的就業人數的人來港，我們對船民的增加。
Reordered Pharaoh	此外，又面對的經濟增長的旅客對本港的船民，香港市民的數目
Original Google	與此同時,香港再次面臨前景的人越來越多,乘船抵達我們的海岸.
Reordered Google	與此同時,香港再次面臨的前景,越來越多的來到我們海岸船民.
Original English	These units are regularly checked by Housing Department management staff and security guards.
Reordered English	These units are regularly by Housing Department management staff and security guards checked.
Original Chinese	該署的管理人員和護衛員經常都有巡視這些單位。
Original Pharaoh	這些單位是定期檢查的房屋署管理人員及保安人員。
Reordered Pharaoh	這些單位是由房屋署管理人員及保安人員檢查。定期

Original Google	這些單位的定期檢查房屋署管理人員和警衛.
Reordered Google	這些單位常常是由房屋署的管理人員和保安人員檢查.
Original English	In serious cases, this could cause permanent disability or even endanger the life of the patient.
Reordered English	In serious cases, this could cause permanent disability or even endanger of the patient the life.
Original Chinese	在嚴重的情況下，更可使病人終身殘廢，甚或性命垂危。
Original Pharaoh	在嚴重個案，這可能導致永久性傷殘或甚至危害的生活的病人的。
Reordered Pharaoh	在嚴重個案，這可能導致永久性傷殘或甚至影響的病人的生命的。
Original Google	在嚴重的情況下,這可能會造成終身殘疾,甚至危及生命的病人.
Reordered Google	在嚴重的情況下,這可能會造成終身殘疾,甚至危害病人的生命.
Original English	The support given to the group by the OMELCO secretariat and the legal adviser is also greatly appreciated.
Reordered English	given to by the OMELCO secretariat and the legal adviser the group The support is also greatly appreciated.
Original Chinese	此外，我亦非常感謝兩局議員辦事處及法律顧問對小組的支持。
Original Pharaoh	有組織的秘書處的法律顧問也很明白，是法律及兩局的支持的。
Reordered Pharaoh	在秘書處的法律顧問小組的支持亦十分明白，要提出的方案和法律的。
Original Google	支持該集團由兩秘書處和法律顧問,也大加讚賞.
Reordered Google	給了兩秘書處和法律顧問小組的支持更是大加讚賞.
Original English	Nevertheless, the point will be explained again by the Financial Secretary.

Reordered English	Nevertheless, the point will be again explained by the Financial Secretary.
Original Chinese	無論如何，財政司將會再次闡釋這點。
Original Pharaoh	不過，問題是，由財政司司長的解釋會的。
Reordered Pharaoh	不過，問題是，由財政司司長的解釋會的。
Original Google	不過,一點也可以解釋,再由財政司司長.
Reordered Google	不過,一點也可以解釋,再由財政司司長.
Original English	138. Meanwhile, the Administration has been examining the working conditions and terms of service of Government doctors.
Reordered English	138. Meanwhile, the Administration has been examining the working conditions and terms of of Government doctors service.
Original Chinese	138.同時，當局一直在研究政府醫生的工作環境和服務條件。
Original Pharaoh	此外,政府亦正研究的工作條件和條款的服務的政府醫生。１３８的。
Reordered Pharaoh	此外,政府亦正研究的工作條件和條款的政府醫生的服務。１３８的。
Original Google	138.同時,當局一直在研究的工作條件和服務條件,政府醫生.
Reordered Google	138.與此同時,政府已就他們的工作條件和條款中,政府醫生服務.
Original English	With these remarks, Sir, I support the motion.
Reordered English	With Sir, these remarks, I support the motion.
Original Chinese	主席先生，我謹此陳辭，支持動議。
Original Pharaoh	與此陳辭，先生，我支持這項動議。
Reordered Pharaoh	與此陳辭，支持我的議案。先生，

Original Google	我謹此陳詞,我支持這項議案.
Reordered Google	與先生,這些話,我支持這項議案.
Original English	In the event, the final report did not differ, in substance, from the interim report.
Reordered English	In the event, the final report did not differ, in substance, from the interim report.
Original Chinese	因此，最後報告和中期報告的內容，實質上並無不同。
Original Pharaoh	在該活動的最後報告並沒有不同，在物質，由目前的報告。，
Reordered Pharaoh	在該活動的最後報告並沒有不同，在物質，由目前的報告。，
Original Google	在這一事件中,最後的報告,並沒有不同,但在實質上,從中期報告.
Reordered Google	在這一事件中,最後的報告,並沒有不同,但在實質上,從中期報告.
Original English	For the whole scene of change is the living and working environment of Hong Kong.
Reordered English	For the whole scene of change is of Hong Kong the living and working environment.
Original Chinese	本港的生活環境和工作環境都在全面改變中。
Original Pharaoh	為整個情況的改變是的生活和工作環境的香港，香港的
Reordered Pharaoh	為整個情況的改變是香港的生活和工作環境，香港的。
Original Google	為整個場景的變化是生活和工作環境的香港.
Reordered Google	為整個場景的變化是,香港的生活和工作環境.
Original English	Manpower shortage is sometimes a sign of continuing success rather than a symptom of failure.
Reordered English	Manpower shortage is sometimes a sign of continuing success rather than

	a symptom of failure.
Original Chinese	人手短缺有時是社會繼續興旺的現象，而不是衰落的象徵。
Original Pharaoh	人手短缺是一個簽署的繼續成功而不是一個症狀的原因，有時
Reordered Pharaoh	人手短缺是一個簽署的繼續成功而不是一個症狀的原因，有時
Original Google	人手不足,有時是持續的成功,而不是失敗的症狀.
Reordered Google	人手不足,有時是持續的成功,而不是失敗的症狀.
Original English	Questions can also be raised in the Public Accounts Committee, and in this Council during question time.
Reordered English	Questions can also be during question time in the Public Accounts Committee, and in this Council raised.
Original Chinese	議員亦可在政府帳目委員會以及在本局會議提問時間中提出問題。
Original Pharaoh	問題是政府帳目委員會，並在本局在質詢時間，亦可在舉手。
Reordered Pharaoh	問題是當時的政府帳目委員會，並在本局提出，亦可在問題。
Original Google	問題也可以提出帳目委員會的建議,並在此期間,議會質詢時間.
Reordered Google	問題也可在質詢時間,在公共帳目委員會,並在本會提出.
Original English	Clearly that would fit into existing immigration policy.
Reordered English	Clearly that would fit into existing immigration policy.
Original Chinese	這項政策很明顯是符合目前的入境政策規定。
Original Pharaoh	當然，亦可以考慮的入境政策。
Reordered Pharaoh	當然，亦可以考慮的入境政策。

Original Google	顯然是符合目前的移民政策.
Reordered Google	顯然是符合目前的移民政策.
Original English	131. I have dwelt at some length on the subject of inflation.
Reordered English	131. I have dwelt at on the subject of inflation some length.
Original Chinese	131.我說了頗多關於通貨膨脹的問題。
Original Pharaoh	百分之一百三十一。我已說了很多有關的問題。脹的。
Reordered Pharaoh	一百三十一。我亦打算在有關的問題脹的一些時間。
Original Google	131.我有不少篇幅關於通貨膨脹.
Reordered Google	131.我已經花了關於通貨膨脹的部分長度.
Original English	Council went into Committee.
Reordered English	Council went into Committee.
Original Chinese	本局進入委員會審議階段。
Original Pharaoh	立法會提交委員會。
Reordered Pharaoh	立法會提交委員會。
Original Google	會進入委員會.
Reordered Google	會進入委員會.
Original English	Talks with the nurses will continue.
Reordered English	with the nurses Talks will continue.
Original Chinese	至於與護士的會談，亦會繼續下去。
Original Pharaoh	就處理的護士會繼續。

Reordered Pharaoh	與護士的談判會繼續。
Original Google	與護士的會談將繼續下去.
Reordered Google	與護士的會談將繼續下去.
Original English	Our simple and flexible tax system is one of the major attractions to many investors in Hong Kong.
Reordered English	Our simple and flexible tax system is of to in Hong Kong many investors the major attractions one.
Original Chinese	本港能吸引大量投資者，稅制簡單而富彈性是其中一個主要因素。
Original Pharaoh	我們簡單及低稅率制度是其中的主要景點，很多投資者在香港，香港的。
Reordered Pharaoh	我們簡單及低稅率制度是要在香港投資的主要特色是，香港很多的。
Original Google	我們簡單和靈活的稅收制度是一大吸引力,很多投資者對香港的信心.
Reordered Google	我們簡單和靈活的稅收制度,是為了在香港許多投資者主要景點之一.
Original English	Land, capital and manpower are the three indispensable elements in an economic structure.
Reordered English	Land, capital and manpower are in an economic structure the three indispensable elements.
Original Chinese	土地、資金、人力，是經濟結構中的三元素，缺一不成。
Original Pharaoh	土地建設和人力的三項因素在經濟結構，一個是，
Reordered Pharaoh	土地建設和人力是一個經濟體系的三項因素。此外，
Original Google	土地,資本和人力都是不可或缺的三個要素,在經濟結構.

Reordered Google	土地,資本和人力資源正處於經濟結構的三個不可或缺的因素.
Original English	All areas of activity of the corporation saw strong performance in the past year.
Reordered English	All areas of of the corporation activity saw strong performance in the past year.
Original Chinese	過去一年，該公司各方面的業務都有出色的表現。
Original Pharaoh	這些活動的公司認為，在過去三年，該服務的範圍。
Reordered Pharaoh	各方面的活動是否在過去兩年的表現強勁的公司的。
Original Google	一切活動領域的公司都有出色表現,在過去的一年.
Reordered Google	所有地區的企業活動都有出色表現,在過去的一年.
Original English	For this reason, I have participated actively in the meetings of the OMELCO Constitutional Development Panel.
Reordered English	For this reason, I have in of the OMELCO Constitutional Development Panel the meetings participated actively.
Original Chinese	所以我非常積極參與兩局議員憲制發展小組的會議。
Original Pharaoh	基於這個原因，我沒有參與會議的兩政制發展小組，積極參與的。
Reordered Pharaoh	基於這個原因，我已在該兩政制發展的積極參與委員會的會議。
Original Google	基於這個原因,我都積極地參加了本次會議的兩憲制發展小組.
Reordered Google	為此,我已在兩憲制發展小組會議,積極參與.
Original English	But they have also produced a strong sense of unity in our community.
Reordered English	But they have also in our community produced a strong sense of unity.
Original Chinese	但這些事件亦使本港社會產生一種強烈的團結意識。

Original Pharaoh	但他們亦有提供一個更統一的社會，我們的。
Reordered Pharaoh	但他們已在本港有一個很強烈的團結。」
Original Google	但也產生一種強烈的團結,我們的社會.
Reordered Google	但也給我們的社會產生一種強烈的團結.
Original English	The British Government has accepted its obligations to citizens elsewhere.
Reordered English	The British Government has accepted its obligations to citizens elsewhere.
Original Chinese	英國政府亦曾對其他一些屬土的公
Original Pharaoh	英國政府已接納其責任向市民地方。」
Reordered Pharaoh	英國政府已接納其責任向市民地方。」
Original Google	英國政府已接受其義務的公民一樣.
Reordered Google	英國政府已接受其義務的公民一樣.
Original English	Miss Maria TAM as a landlord.
Reordered English	Miss Maria TAM as a landlord.
Original Chinese	譚惠珠議員宣稱為業主。
Original Pharaoh	譚惠珠為一業主。
Reordered Pharaoh	譚惠珠為一業主。
Original Google	譚惠珠為業主.
Reordered Google	譚惠珠為業主.

Original English	Farmers actually receive a subsidy for the installation of treatment facilities.
Reordered English	Farmers actually receive a subsidy for the installation of treatment facilities.
Original Chinese	農民實際上獲得一筆裝置廢物處理設施的津貼。
Original Pharaoh	他們所獲得的資助申請的處理裝置的設施。
Reordered Pharaoh	他們所獲得的資助申請的處理裝置的設施。
Original Google	農民實際得到的補貼用於安裝污水處理設施.
Reordered Google	農民實際得到的補貼用於安裝污水處理設施.
Original English	Many potentially useful technologies are not yet being applied in our industries.
Reordered English	Many potentially useful technologies are not yet being applied in our industries.
Original Chinese	有很多可以發揮效用的科技項目，目前仍未為本港工業界採用。
Original Pharaoh	很多可能是沒有幫助，但在我們工業，申請。
Reordered Pharaoh	很多可能是沒有幫助，但在我們工業，申請。
Original Google	許多潛在有用的技術尚未得到應用的行業.
Reordered Google	許多潛在有用的技術尚未得到應用的行業.
Original English	I welcome the comprehensive scope of your outline of a planned future for Hong Kong.
Reordered English	I welcome of your outline of for Hong Kong a planned future the comprehensive scope.
Original Chinese	閣下就本港未來所描述的建設計劃，不但全面週詳，而且高瞻遠矚。

Original Pharaoh	我歡迎各位提出的一個計劃來港，對香港整體的範圍。
Reordered Pharaoh	我歡迎閣下的施政報告為香港的未來計劃的範圍。全面的。
Original Google	我歡迎全面性你綱要規劃香港的未來.
Reordered Google	我歡迎你們的輪廓,為香港的未來計劃的範圍全面.
Original English	With these words, Sir, I support the motion.
Reordered English	With Sir, these words, I support the motion.
Original Chinese	主席先生，我謹此陳辭，支持動議。
Original Pharaoh	與此陳辭，先生，我支持這項動議。
Reordered Pharaoh	與此陳辭，支持我的議案。先生，
Original Google	這些話,先生,我支持這項議案.
Reordered Google	與先生,這些話,我支持這項議案.
Original English	These are fully accepted.
Reordered English	These are fully accepted.
Original Chinese	對於這些意見，我完全接受。
Original Pharaoh	這些是完全接受。
Reordered Pharaoh	這些是完全接受。
Original Google	這些都是完全接受.
Reordered Google	這些都是完全接受.
Original English	And that, Sir, is not the reason for this Convention.
Reordered English	And that, Sir, is not for this Convention the reason.

Original Chinese	然而，主席先生，那並非締結該公約的原因。
Original Pharaoh	說，是沒有的，因為這公約，而先生，
Reordered Pharaoh	說，是不會對這公約的原因，以及先生，
Original Google	而且,先生,是不是該公約的原因.
Reordered Google	而且,先生,不是本公約的理由.
Original English	"Care in the community" is also the concept adopted in our health programme for the elderly.
Reordered English	"in the community Care " is also adopted in our health program for the elderly the concept.
Original Chinese	政府在推行老人保健計劃時，亦採用「社區照顧」的觀念。
Original Pharaoh	「關懷的社會」的概念，在我們計劃對老人的健康亦是事實。
Reordered Pharaoh	「社區照顧」是用在健康計劃申請的長者的概念，我們亦在該
Original Google	"社區照顧"也是一種觀念,通過對老人保健計劃.
Reordered Google	"在社區照顧'也是通過我們的健康計劃的老觀念.
Original English	Regrettably this particular point on the relationship between the council and the boards has never been properly addressed.
Reordered English	Regrettably on between the council and the boards the relationship this particular point has never been properly addressed.
Original Chinese	可惜這點有關管理局和管理委員會之間關係的意見從未獲妥善處理。
Original Pharaoh	可惜，特別是對的血緣關係的議員，有適當的，是沒有的。
Reordered Pharaoh	可惜有關的立法議會的關係，特別是絕對正確的，是沒有關係的。

Original Google	可惜這點的關係會和議會從來沒有得到妥善處理.
Reordered Google	可惜在之間和議會的關係,這一點從未得到妥善處理.
Original English	The law must be applied fairly.
Reordered English	The law must be fairly applied.
Original Chinese	立法必須公允。
Original Pharaoh	該法例必須是公平的。
Reordered Pharaoh	該法例必須是公平的。
Original Google	法律必須公允.
Reordered Google	法律必須是公平應用.
Original English	This is not without reason.
Reordered English	This is not without reason.
Original Chinese	需時如此長久並非無因。
Original Pharaoh	這並不是沒有理由。
Reordered Pharaoh	這並不是沒有理由。
Original Google	這是沒有道理的.
Reordered Google	這是沒有道理的.
Original English	SECRETARY FOR SECURITY: Sir, this of course is a very difficult area.
Reordered English	FOR SECURITY: Sir, this of course SECRETARY is a very difficult area.
Original Chinese	保安司答（譯文）: 主席先生，這確實是一個非常棘手的問題。
Original Pharaoh	主席先生，這當然是一個很困難的，是為保障。

Reordered Pharaoh	主席先生，這當然是一個很困難的，局長的保障。
Original Google	保安局局長:先生,這當然是個非常棘手的問題.
Reordered Google	答:先生,這當然是庫是一個非常棘手的問題.
Original English	(a) the rapid acceleration of demand in recent years.
Reordered English	(a) the rapid acceleration of in recent years demand.
Original Chinese	(a)這方面的需求近年來急劇增加。
Original Pharaoh	(一)該急劇加速，在最近數年的。
Reordered Pharaoh	(一)加快在最近的需求。急劇的。
Original Google	(一)迅速加快需求近年來.
Reordered Google	(一)迅速加快,近年來需求.
Original English	Only British nationals are entitled to British consular protection.
Reordered English	Only British nationals are entitled to British consular protection.
Original Chinese	只有英國國民才可享有英國領事保護權。
Original Pharaoh	只有英國公民享有英國領事保護。通過。
Reordered Pharaoh	只有英國公民享有英國領事保護。通過。
Original Google	只有英國國民都享有英國的領事保護權.
Reordered Google	只有英國國民都享有英國的領事保護權.
Original English	This misconception is based on China's lack of understanding of the mechanism of local politics.
Reordered English	This misconception is based on China's lack of of of local politics the mechanism understanding.

Original Chinese	這個誤解是起於中國不理解本港政治的機制。
Original Pharaoh	這個錯誤是基於對中國缺乏認識的機制的地方。政黨的議員。
Reordered Pharaoh	這種想法是對中國的政治制度的認識。本地缺乏的是的。
Original Google	這種誤解,是基於對中國缺乏了解的機制,地方政治.
Reordered Google	這種誤解,是基於中國缺乏對當地政治機制的理解.
Original English	However, in spite of this scheme wages have not eased.
Reordered English	However, in spite of this scheme wages have not eased.
Original Chinese	然而，儘管實行了此項計劃，工資卻
Original Pharaoh	然而，在這計劃亦不容易。不過的。
Reordered Pharaoh	然而，在這計劃亦不容易。不過的。
Original Google	不過,儘管有此計劃,工資都不能放鬆.
Reordered Google	不過,儘管有此計劃,工資都不能放鬆.
Original English	Rents have increased considerably since the last revaluation exercise.
Reordered English	Rents have since the last revaluation exercise considerably increased.
Original Chinese	自從上次重估了樓宇應課差餉租值之後，住宅租金大幅度上漲很多。
Original Pharaoh	租金已大幅上升，去年的重估工作。
Reordered Pharaoh	租金已由去年的重估工作大幅增加。
Original Google	租金已大大提高,自去年重估.
Reordered Google	租金已自去年重估大大增加.

Original English	Prevention is better than cure.
Reordered English	Prevention is better than cure.
Original Chinese	預防勝於治療，所費雖少，效果則甚大。
Original Pharaoh	預防是勝於治療的。
Reordered Pharaoh	預防是勝於治療的。
Original Google	預防勝於治療.
Reordered Google	預防勝於治療.
Original English	Sufficient of these just do not exist.
Reordered English	Sufficient of these just do not exist.
Original Chinese	本港在這數方面的服務並不足夠。
Original Pharaoh	足夠的，是不會沒有的。
Reordered Pharaoh	足夠的，是不會沒有的。
Original Google	充足,這些是不存在的.
Reordered Google	充足,這些是不存在的.
Original English	These fees were last revised in 1989.
Reordered English	These fees were last in 1989 revised.
Original Chinese	這些收費上次調整的時間是一九八九年。
Original Pharaoh	這些費用是一次修訂是在一九八九年。
Reordered Pharaoh	這些費用是上次在一九八九年修訂。
Original Google	這些收費,上一次是在 1989 年.

Reordered Google	這些費用上次於 1989 年修訂的.
Original English	Currently, for example, we have the Provisional Hospital Authority and the Provisional Airport Authority.
Reordered English	Currently, for example, we have the Provisional Hospital Authority and the Provisional Airport Authority.
Original Chinese	現有例子是臨時醫院管理局和臨時機場管理局。
Original Pharaoh	現時，我們已臨時醫院管理局、臨時機場管理局局長例如，政府的。
Reordered Pharaoh	現時，我們已臨時醫院管理局、臨時機場管理局局長例如，政府的。
Original Google	例如,目前,我們已經建立臨時醫院管理局和臨時機場管理局.
Reordered Google	例如,目前,我們已經建立臨時醫院管理局和臨時機場管理局.
Original English	I personally do not agree with the change to the reform in a piecemeal approach.
Reordered English	I personally do not agree to the reform in a piecemeal approach with the change.
Original Chinese	我個人不贊成片面修訂法例的做法。
Original Pharaoh	我個人並不同意，要改革的一個方法。零碎的改變的。
Reordered Pharaoh	我個人並不同意要改革的一個零碎方式處理的。改變的。
Original Google	我個人不同意變更為改革採用化整為零的辦法.
Reordered Google	我個人並不贊成改革的零碎的變化.
Original English	The question has nevertheless been given thorough consideration by the Administration.
Reordered English	. nevertheless has been given thorough consideration by the Administration The question

Original Chinese	不過，當局已對這問題詳加研究。
Original Pharaoh	問題是，在詳細考慮的。政府的決定。
Reordered Pharaoh	不過，要在詳細考慮的有關當局的問題。
Original Google	問題,但是,已經充分考慮了.
Reordered Google	.不過,已經充分考慮到了管理問題
Original English	The Honourable Paul CHENG has highlighted the main issues.
Reordered English	The Honourable Paul CHENG has highlighted the main issues.
Original Chinese	鄭明訓議員已談論過各項主要的問題。
Original Pharaoh	鄭明訓議員所提出的主要問題的。
Reordered Pharaoh	鄭明訓議員所提出的主要問題的。
Original Google	鄭明訓也突出了主要問題.
Reordered Google	鄭明訓也突出了主要問題.
Original English	A White Bill was published for public discussion on 6 March 1990.
Reordered English	A White Bill was March 1990 for public discussion on 6 published.
Original Chinese	白紙條例草案已於一九九零年三月六日發表，以供公眾討論。
Original Pharaoh	一個條例草案是公開讓市民討論有關。三月)。
Reordered Pharaoh	一個條例草案是三月，為市民在討論六刊登。
Original Google	白紙草案供公眾討論,對 1990 年 3 月 6 日.
Reordered Google	白紙條例草案已於 1990 年 3 月進行公開討論 6 月出版.


TSING HUA UNIVERSITY