words, our phrase reordering model also integrates more significant syntactic knowledge such as POS information and dependencies from the syntax tree, which can avoid some intractable phrase reordering errors.

A hierarchical phrase-based model was proposed by (Chiang, 2005). In his method, a synchronous CFG is used to reorganize the phrases into hierarchical ones and grammar rules are automatically learned from corpus. Different from his work, foreign syntactic knowledge is introduced into the synchronous grammar rules in our method to restrict the arbitrary phrase reordering.

Syntax-based SMT systems (Yamada et al., 2001; Quirk et al., 2005; Liu et al., 2006) totally depend on syntax structures to complete phrase translation. They can capture global reordering by simply swapping the children nodes of a parse tree. However, there are also reordering cases which do not agree with syntactic structures. Furthermore, their model is usually much more complex than a phrase-based system. Our method exactly attempts to integrate the advantages of phrase-based SMT system and syntax-based SMT system to improve the phrase reordering model. Phrase translation in our system is independent of syntactic structures.

3 The Model

In our work, we focus on building a better reordering model with the help of source parsing information. Although we borrow some fundamental elements from a phrase-based SMT system such as the use of bilingual phrases as basic translation unit, we are more interested in introducing syntactic knowledge to strengthen the ability to handle global reordering phenomena in translation.

3.1 Definitions

Given a foreign sentence f and its syntactic parse tree T, each leaf in T corresponds to a single word in f and each sub-tree of T exactly covers a phrase f_i in f which is called as *linguistic phrase*. Except linguistic phrases, any other phrase is regarded as *non-linguistic phrase*. The *height* of phrase f_i is defined as the distance between the root node of T and the root node of the maximum sub-tree which exactly covers f_i. For example, in Figure 1 the phrase “大幅” has the maximum sub-tree rooting at ADJP and its height is 3. The height of phrase “的” is 4 since its maximum sub-tree roots at ADBP instead of AD. If two adjacent phrases have the same height, we regard them as *peer phrases*.

In our model, we make use of *bilingual phrases* as well, which refer to source-target aligned phrase pairs extracted using the same criterion as most phrase-based systems (Och and Ney, 2004).

3.2 Model

Similar to the work in Chiang (2005), our translation model can be formulated as a weighted synchronous context free grammar derivation process. Let D be a derivation that generates a bilingual sentence pair $\langle f, e\rangle$, in which f is the given source sentence, the statistical model that is used to predict the translation probability $p(e|f)$ is defined over Ds as follows:

$$p(e|f) \propto p(D) \propto p_{lm}(e)^{\lambda_{lm}} \times \prod_i \prod_{X\rightarrow\langle\gamma,\alpha\rangle\in D} \phi_i(X \rightarrow \langle\gamma,\alpha\rangle)^{\lambda_i}$$

where $p_{lm}(e)$ is the language model, $\Phi_i(X \rightarrow\langle\gamma,\alpha\rangle)$ is a feature function defined over the derivation rule $X\rightarrow\langle\gamma,\alpha\rangle$, and λ_i is its weight.

Although theoretically it is ideal for translation reorder modeling by constructing a synchronous context free grammar based on bilingual linguistic parsing trees, it is generally a very difficult task in practice. In this work we propose to use a small synchronous grammar constructed on the basis of bilingual phrases to model translation reorder probability and constraints by referring to the source syntactic parse trees. In the grammar, the source / target words serve as terminals, and the bilingual phrases and combination of bilingual phrases are presented with non-terminals. There are two non-terminals in the grammar except the start symbol S: Y and Z. The general derivation rules are defined as follows:

a) Derivations from non-terminal to non-terminals are restricted to binary branching forms;
b) Any non-terminals that derives a list of terminals, or any combination of two non-terminals, if the resulting source string won’t cause any cross-bracketing problems in the source parse tree (it exactly corresponds to a linguistic phrase in binary parse trees), are reduced to Y;
c) Otherwise, they are reduced to Z.

Table 1 shows a complete list of derivation rules in our synchronous context grammar. The first nine grammar rules are used to constrain phrase reor-