

## Chapter Three

# Treebanking: The Compilation of a Sample PFR Corpus of Skeleton-parsed Sentences

### 3.1 Introduction

The PFR Chinese Corpus used in this book is part-of-speech (POS) tagged. For the purposes of this book, however, I decided to add a further level of annotation to the corpus – constituent structure or syntactic annotation. As its name suggests, syntactic annotation makes explicit the constituent structure of sentences, usually in the form of labelled brackets. In the following sections, I will discuss the methods and issues relevant to the syntactic annotation of the PFR Chinese Corpus by building a sample treebank. The goal in undertaking this annotation is to ease the identification of adverbial clauses in Chinese, which is the prime object of my book.

*Treebanks* are simply corpora in which syntactic constituent structure is made explicit by a process of corpus annotation (Leech and Garside, 1991: 15; Abeillé, 2003: xiv). My major concern here is not with software to achieve this annotation automatically (as at the time of writing, there are no effective available parsers designed for the Chinese language), but with the establishment of a parsing scheme and its manual application to written Chinese corpus data. More specifically, the approach taken here is inspired by the *skeleton parsing* approach (Eyes and Leech, 1993; Garside, 1993; Black et al., 1996; Leech and Eyes, 1997). Skeleton parsing seeks to produce simplified constituent-structure annotations. I do not intend to go into a deep or logical annotation which would apply functional labels for constituents (e.g. subject, object, etc.) as traditional reference grammar books do (Quirk et al., 1985). Rather, I will focus on categorial labels such as noun phrase, prepositional phrase, adverb phrase, etc.

The reason why I use categorial labels rather than functional labels is that the former suffice for the identification of my object of

study. Adverbial clauses function only as adjuncts in the sentences in my corpus. However, not all adjuncts are adverbial clauses: apart from the adverbial clause, an adjunct can take various syntactic forms such as noun phrase, prepositional phrase, adverb phrase, etc., which are all grammatical categories. As will be shown in the description of my parsing scheme (section 3.4), a separate label is set up for a grammatical category to function as an adjunct (e.g. adverbial noun phrase, adverbial prepositional phrase, etc.). As the only grammatical function that is relevant here is adjunct, which can be adequately identified by the use of categorial labels, other grammatical functions such as subject, object, etc. that are not directly related to the goal of this book can in principle be put aside for the sake of simplicity in skeleton parsing. I will return, however, in the conclusion of my book to the possible advantages of future research considering functional labelling.

## 3.2 An overview of past parsing projects

Both Bateman et al. (1997) and Leech and Eyes (1997) provide a detailed survey of a variety of approaches taken to syntactic annotation by different projects, namely the Lancaster/IBM Treebank (Barnett et al., 1996), the SUSANNE Corpus (Sampson, 1995), the Nijmegen Treebanks comprising both the Nijmegen Corpus (Keulen, 1986) and the TOSCA Corpus (Oostdijk, 1991), the Constraint Grammar parser of the Helsinki group (Karlsson et al., 1995) and the Penn Treebank (Marcus et al., 1993; Marcus et al., 1994). These parsing projects were chiefly concerned with the syntactic annotation of the English language.<sup>1</sup> Recently, however, syntactic annotation has been extended to languages other than English. The Penn Chinese Treebank Project<sup>2</sup> (Xia et al.,

- 1 Though, the Constraint Grammar system is currently being developed for Finnish (Karlsson, 1990; Koskenniemi, 1990), Swedish (Voutilainen, 2001), Danish (Bilgram and Keson, 1998), German (Nießen and Ney, 2001), Basque (Aduriz et al., 1997) and French (Chanod and Tapanainen, 1995).
- 2 The first version of the treebank was released via LDC in December 2000. See the project's web page: <<http://www.cis.upenn.edu/~chinese/>> (accessed 15 May 2012).

2000; Chiou et al., 2001; Xue et al., 2002) is a case in point. While this Chinese treebank initially seemed to be a promising corpus on which to carry out my research, it proved to be unsuitable for my purpose as well, as will be outlined shortly.

The TOSCA and Constraint Grammar based work was also unsuitable; the Nijmegen Treebanks are produced by a semi-automatic parser which provides an interactive environment allowing human intervention to complete and correct the parsing. The Penn Treebank<sup>3</sup> and the Bank of English Corpus were annotated by the Constraint Grammar (CG) parser<sup>4</sup>, a fully automatic parser. The TOSCA and CG parsers have one thing in common: they adopt a considerable number of syntactic tags. The CG Framework marks 29 syntactic functions, while the TOSCA annotation scheme contains 16 categorial labels, over 90 functional labels and over 100 attribute labels (e.g. active/passive voice, raised direct object, etc.). My work entails manual parsing and it is difficult for a sole human analyst to annotate consistently and accurately with such complex annotation systems. Hence I rejected the use of the CG and TOSCA schemes and I will not go into details of these automatic/semi-automatic parsing models further in this book. Nevertheless, this is not to ignore the annotation schemes totally. The categorial labels were carefully reviewed in working out my final parsing scheme.

Of the parsing schemes under consideration, only the Lancaster/IBM Treebank, SUSANNE Corpus and Penn Chinese Treebank were manually parsed. At the level of surface grammar, which identifies constituents in a clause and assigns labels to these constituents, the annotation schemes used in both the SUSANNE Corpus and Penn Chinese Treebank are very similar to that of the Lancaster Treebank, although more detail is included in both the SUSANNE and Penn corpora. The SUSANNE Corpus and Penn Chinese Treebank give detailed analyses of elements of syntactic functions and deep structure like traces, syntactic ambiguities, semantic roles, etc. that are of no use to this study.

- 3 The Penn Treebank has undergone an extensive manual post-editing and proved that correcting the parser's output is significantly faster than manually parsing a corpus (Marcus et al. 1993: 317).
- 4 The Constraint Grammar formalism is notable for incorporating a dependency grammar framework, in contrast to the phrase structure grammar models employed for most other treebanks.

A further major drawback, for my book, of the Penn Chinese Treebank is that it is not theory neutral. The influence of Government and Binding Theory (GB)<sup>5</sup> is clear in the corpus. Though it is claimed that only those assumptions of GB that are least controversial are adopted in the annotation scheme (Xue et al., 2000: 4–5), this inevitably called for a number of abstract covert phrasal categories such as IP<sup>6</sup>, CP<sup>7</sup>, etc. Those phrasal categories may be well-motivated within the theory but their significance cannot be maintained across other theoretical frameworks. Hence, the treebank is far from theory neutral. Since a primary goal of corpus annotation is to provide empirical data for linguistic investigations, it is desirable to annotate categories/features that are the most commonly shared among theories, so that others can use the treebank for their own purposes, without being encumbered with unwanted theoretical assumptions; the minimal basic level syntactic categories are the ones that are shared by all theories. Thus I rejected the use of the Penn Chinese Treebank in part because it jeopardised my goal of theory neutrality.

A final, crucial, point that led to me not using the Penn Chinese Treebank is that its parsing scheme does not contain a separate label for the adverbial clause and thus no attempt was made to give a definition of adverbial clause in this scheme. Rather, it tends to subsume the adverbial clause under the entry of subordinate clause (including relative clause and appositive clause) and VP adjunct. If I were to use the Penn Chinese Treebank, I would have had to restart the annotation from scratch: first, I would have had to introduce a clear definition of what an adverbial clause should be; second, with this clear working definition, I would have to have taken pains to screen all of the subordinate clauses and VP adjuncts, labelling and sorting out only those adverbial clauses of interest to me, on the basis of the assumption that the subor-

- 5 For details of the Government and Binding Theory, see Haegeman (1994).
- 6 IP refers to both finite and infinitival clauses, depending on the content of the node INFL, which is a terminal node signifying the inflectional morphology of the verb, affixes and infinitival *to* (Haegeman, 1994: 114 ff.).
- 7 Complementisers, represented as C, such as *whether*, *if*, *that* and *for* in English introduce a sentence (IP): C selects an IP-complement. The choice of the type of IP depends on the choice of C. The complementisers *that* and *if* select a finite clause as their complement; *for* selects an infinitival clause and *whether* selects either type of clause. Thus CP refers to either a finite or non-finite clause introduced by a complementiser. See Haegeman (1994: 116 ff.).

dinate clauses and VP adjuncts include all of the adverbial clauses I want. In other words, there is not much difference between using the data produced by the Penn Chinese Treebank and conducting my own manual parsing. I therefore decided to start afresh and conduct my own annotation. Having decided not to use the Penn treebank annotation, I considered whether to use the raw corpus texts. I decided not to use the raw corpus data from the Penn Chinese Treebank firstly because the newspaper articles available in the treebank were taken from the Xinhua newswire between 1994 and 1998, which were comparatively less recent than other data I had available to me (Yu, 1999: 20); secondly, the Penn Chinese Treebank contains about 100K words, which is ten times smaller in size than the data available to me, which amounted to approximately 1 million Chinese words.

Having decided that I needed to develop a theory neutral annotation of Chinese, and having also rejected schemes such as CG, TOSCA and SUSANNE as they were too complex and in some cases required partial automation, I decided to follow the approach to syntactic annotation used in the Lancaster/IBM Treebank. With reference to skeleton parsing, the Lancaster/IBM Treebank (compiled between 1987 and 1991), consisting of approximately 3 million words, is the largest of all of the manually annotated treebanks. It was produced by Leech and Garside (1991), in collaboration with IBM. The Lancaster/IBM treebank built upon the work carried out on its precursors, the Lancaster-Leeds Treebank (Lang, 1989; Edwards, 1993: 298) and the LOB Corpus Treebank (Sampson, 1987: 83–84). One of the major innovations of the Lancaster/IBM treebank was the skeleton parsing approach (see Garside and McEnery, 1993). The simplicity of skeleton parsing not only expedites the process of treebank compilation, it also reduces the intellectual complexity of the task of parsing for human analysts, thereby minimising inconsistencies and inaccuracies in treebanking practice. This makes it an attractive approach for the lone scholar. Given that I had no suitable treebanked Chinese corpus available, I proceeded to construct my own using the skeleton parsing approach.

In the remainder of this chapter, I will discuss at length the treebanking method I have developed for Chinese. In doing so, I will analyse in detail the processing of a sample text. I discuss first, in section 3.3, the selection of the text sample. In section 3.4, I describe the key components of a standard parsing scheme as exemplified by the

UCREL skeleton parsing tagset and then give a detailed description of my own parsing scheme. In section 3.5, I explain the elaborate guidelines developed for the purposes of consistency and accuracy in my annotation. I then explain the process of assigning labelled brackets to the part-of-speech-tagged sentences and the difficulties encountered in the application of the annotation practices proposed (section 3.6). Following from this, the parses of adverbials and, most importantly, adverbial clauses will be examined (section 3.7). The conclusion will give an evaluation of the success of my manual skeleton parsing (section 3.8).

### 3.3 PFR Sample Skeleton Treebank: Text selection

From the PFR Chinese Corpus, a sample text of some 100,000 word tokens, yielding approximately 2,500 sentences, was chosen for the production of my treebank. I contend that a standard block of about one hundred thousand words is a unit of about the right size for skeleton parsing and handling by a human treebanker. Furthermore, my choice of text was governed by the need to produce a parsed sample corpus of reasonable length which could not only be manageable for hand-parsing but also represent a typical chunk of the data that would eventually be treebanked by me.

### 3.4 PFR Sample Skeleton Treebank: Parsing scheme

As Sampson (1995: 2ff.) puts it, the process of parsing refers to the ability to extract from a linear sequence of words the underlying hierarchical grammatical structure, and thus a *parsing scheme* “is a set of categories and notational conventions allowing the grammatical properties of a text to be made explicit”. In other words, it is a guideline document which helps the human analyst parse sentences (Leech and Garside, 1991: 15–16). A clearly defined parsing scheme is essential for the production of a satisfactorily parsed text.

### 3.4.1 UCREL skeleton parsing annotation scheme

As most of the existing treebanks are primarily based upon English texts, it does not come as a surprise that the annotation schemes used on those treebanks chiefly reflect the syntactic categories which are directly relevant to English grammar. A case in point is the UCREL skeleton parsing scheme, as illustrated in the following table.<sup>8</sup>

UCREL	<i>Skeleton Parsing Annotation Scheme</i>
<b>Fa</b>	Adverbial Clause
<b>Fc</b>	Comparative Clause
<b>Fn</b>	Noun Clause
<b>Fr</b>	Relative Clause
<b>G</b>	Genitive
<b>J</b>	Adjective Phrase (predicative)
<b>N</b>	Noun Phrase
<b>Nr</b>	Adverbial Noun Phrase (temporal)
<b>Nv</b>	Adverbial Noun Phrase (non-temporal) (not in AP or SEC corpora)
<b>P</b>	Prepositional Phrase
<b>S</b>	Sentence (used eg in quoted speech, also with + and & as co-ordinates)
<b>Tg</b>	-ing Clause
<b>Ti</b>	Infinitive Clause
<b>Tn</b>	Past Participle Clause
<b>V</b>	Verb Phrase
<b>(null)</b>	Unlabelled Constituent

Table 2: The UCREL skeleton parsing annotation scheme.

As stated, some of the syntactic constituent labels in the UCREL skeleton parsing scheme are specially designed to suit English grammar. When I attempted to adopt this scheme wholesale for my own research, I found that some modifications were needed to accommodate the syntactic properties of the Chinese language.

First of all, since there is no conclusive morphological evidence that motivates the postulation of infinitival clauses in Chinese (Xue et al., 2000: 32), the non-finite clauses, including the *-ing* clause (Tg),

8 The table was adapted from UCREL (University Centre for Computer Corpus Research on Language)'s website <<http://ucrel.lancs.ac.uk/skeletontags.html>> (accessed 15 May 2012).

infinitive clause (Ti) and past participle clause (Tn) were not taken over from the UCREL parsing scheme to my parsing scheme. Similarly, I did not include the noun clause (Fn) or the relative clause (Fr) as they are also not compatible with Chinese syntax. Relative clauses have long been a source of controversy in Chinese linguistics. Some scholars (e.g. Li and Thompson, 1989: 579ff.; Aoun and Li, 1993; Chiu, 1993; Wu, 2000) believe that a nominalisation (whereby a verb, verb phrase, or sentence, followed by the particle *de*, functions as a noun phrase) can be called a relative clause in Chinese if the head noun that it modifies refers to an unspecified element involved in the situation described by the nominalisation. For instance, in such a noun phrase as “卖报纸的老人 *mai baozhi de laoren* ‘the old man who sells newspapers’”, the relative clause refers to the verb phrase marked by the following particle *de*, i.e. 卖报纸的 *maibaozhi de* “sell newspapers”. In this relative clause, the subject, namely the person who sells newspapers, is missing. Accordingly, the head noun 老人 *laoren* refers to this missing subject. The particle *de* is functionally equivalent to the English relative pronoun *who* which indicates a gap or trace that is co-referential with the head noun.

On the other hand, Chao (1968), among others, does not adopt the notion of relative clause in his descriptive grammar of Chinese.<sup>9</sup> Relative clauses may not exist in Chinese for at least two reasons. First, rather than a relativiser, the particle 的 *de*, as Chao (1968: 285) points out, is a marker of explicit modification. It is inserted after the premodifier of a noun phrase to make explicit the modifier-modified structure. Zhu (1982 and 2000) also notes that this particle is “multi-functional” in that it can be a genitive marker (e.g. 他的书 *ta de shu* “his book”), a marker of nominalisation (e.g. 买书的是个学生 *mai shu de shi ge xuesheng* “The person who bought the book is a student”), and an adjectival marker (e.g. 勤劳的学生 *qinlao de xuesheng* “a hardworking student”). It is therefore not a proper “counterpart” of the English relative pronoun which is “unifunctional” (except *that*).<sup>10</sup> The second rea-

9 He does not use the term relative clause; rather, he refers to those constructions marked by *de* as *adjectival clauses* which can modify any nominal expression (Chao, 1968: 111).

10 The function word, *that*, is a relative pronoun, a demonstrative pronoun and a conjunction (Quirk et al., 1985).

son is that some relative clauses do not actually include a gap that the head noun can co-refer to. So far scholars have been considering relative clauses to be clauses in which either the subject or object is missing, thereby producing a gap that co-refers to the head noun the relative clauses modify. Such gaps are used as evidence to prove the existence of relative clauses in Chinese. However, such a gap does not always exist as in 我写信的房子 *wo xie xin de fangzi* “the room in which I write letters”. In this case the so-called relative clause marked by the particle *de* (i.e. 我写信的 *wo xie xin de* “I write letters”) is a well-formed clause without any missing subject or object. In cases such as this, two conflicting explanations were proposed. Chu and Chi (1999: 226) claim that the construction is no longer a relative clause but an appositive clause on the grounds that no gap can be identified in the noun phrase and they use this as a definitive test for distinguishing a relative clause from an appositive clause. Yet, Xue et al. (2000: 80ff.) and Li and Thompson (1989: 582ff.) still perceive it as a relative clause though with an PP trace rather than an NP trace; it differs from the relative clause with an NP trace in that the gap is an adjunct position. Since this is a highly controversial and unresolved issue, I chose not to use this grammatical category in my parsing scheme. Instead, constructions marked by the particle *de* were left unspecified by using unlabelled brackets (see section 3.5.2.1).

Nominal clauses are also difficult in Chinese linguistics. The notion of the nominal clause may be useful in English as the clauses used as sentential subject or object are formally distinctive from those used independently: they are typically introduced by the conjunction *that* as in “That John stole my book was totally incredible” (Quirk and Greenbaum, 1973: 316–322). Chinese clauses or sentences, however, do not vary morphosyntactically when they are used as subject or direct object (Liu et al., 1996: 245, 253). Hence, there is no justification for adopting this grammatical category in a parsing scheme designed for the Chinese language. The lack of previous accounts of nominal clauses in Chinese tends to serve as counterevidence to the notion that nominal clauses exist in Chinese.

Furthermore, owing to the fact that different languages tend to employ different strategies in signalling the same grammatical relation, the parsing label of comparative clause (Fc) was also not adopted in my parsing scheme. While English makes use of a clause to give

comparisons, the comparative constructions in Chinese are expressed by means of a prepositional phrase. The comparative marker 比 *bi* “than” and the phrase that immediately follows it form a prepositional phrase which serves as a preverbal adjunct (Zhao, 1989; Liu, 1999: 204ff.). Compare the following contrived sentences from English and Chinese respectively, both of which express the same meaning:

(1a) He does the assignment [Fc better than I do Fc].

(1b) 他 做 的 功 课            做 得            [P 比    我 P]    好 。

*ta    zuo de    gongke            zuo de            [P bi    woP]    hao*

    he do DE assignment do COMP BI        I            well

    “He does the assignment better than I do.”

Lastly, I did not take the genitive (G) as one of the constituent labels in my parsing scheme so as to avoid terminological controversy. The genitive constructions in English *roughly* correspond to those constructions marked by the particle 的 *de* in Chinese. However, this apparent correspondence is complicated by two issues. Firstly, possessive constructions in Chinese do not necessarily take the particle 的 *de* (Li and Thompson, 1989: 115) as in 他爸爸 *ta baba* and 他的爸爸 *ta de baba* which both indicate the same meaning “his father”. Secondly, apart from marking possessions, the particle 的 *de* can be an adjectival marker and a marker of nominalisation and explicit modification, as mentioned earlier. Since the term “genitive” is confusing as far as Chinese linguistics is concerned, I decided to exclude this constituent from my parsing scheme.

In view of the differences between the English and Chinese grammatical systems, new constituent labels that are not used in the UCREL skeleton parsing scheme had to be invented for the purposes of this research. These are: adverb phrase, correlative clause, adverbial idiom/set phrase, adverbial adjective phrase, adverbial prepositional phrase, adverbial verb phrase and verbal object, which are to be discussed below.

### 3.4.2 PFR skeleton parsing labels

As Leech and Eyes (1997: 37) note, Sampson's (1995) annotation scheme demonstrates three key components of a clearly specified parsing scheme:

- (a) A list of symbols used in the annotation: non-terminals, terminals, and other symbols;
- (b) A basic definition of the symbols: e.g. N = noun phrase;
- (c) A description, which is as detailed as possible, of how the symbols are actually applied to text sentences. For example, how do annotators recognise a noun phrase when they see one, and how do they distinguish noun phrase tokens from words or word sequences which are not noun phrases?

With respect to the first and second points, in keeping with Sampson, I intend to provide these 3 sets of data for my treebank: (a) the non-terminal labels and (b) their definitions with illustrative examples<sup>35</sup> are given for the PFR treebank in Table 3.

Nonterminal category	Symbol	Example
Adverbial Clause	Fa	<p>&lt;Fa&gt;只要_c &lt;N&gt;我们_r&lt;N&gt; &lt;R&gt;进一 步_d&lt;/R&gt; &lt;解放思想_i ,_w 实事求 是_i&gt; ,_w &lt;V&gt;抓住_v &lt;N&gt;机遇 _n&lt;N&gt;&lt;/V&gt; ,_w &lt;开拓进取 _l&gt;&lt;/Fa&gt; ,_w</p> <p>&lt;Fa&gt;zhiyao_c &lt;N&gt;women_r&lt;N&gt; &lt;R&gt;jinyibu_d&lt;/R&gt; &lt;jiefangsixiang_i , _w shishiqiushi_i&gt; ,_w &lt;V&gt;zhuazhu_v &lt;N&gt;jiyu_n&lt;N&gt;&lt;/V&gt; , _w &lt;kaituojinqu_l&gt;&lt;/Fa&gt; ,_w</p> <p>“If we become more open-minded and realistic and make every effort to explore new possibilities,”</p>

35 All of the examples in this book are given in the Chinese characters, followed by *pinyin* romanisations, glosses and English translations.

Nonterminal category	Symbol	Example
Correlative Clause	Fc	<p>&lt;F<b>c</b>&gt;越_d 走_v 越_d 宽广_a&lt;/F<b>c</b>&gt;            &lt;F<b>c</b>&gt;yue_d zou_v yue_d            kuanguang_a&lt;/F<b>c</b>&gt;            “the more we walk, the broader (the road)            will be”</p>
Main Clause (to which the adverbial clause is subordinated)	Fm	<p>&lt;F<b>m</b>&gt;&lt;F<b>a</b>&gt;只要_c &lt;N&gt;我们_r&lt;/N&gt; &lt;R&gt;            进一步_d&lt;/R&gt; &lt;解放思想_j ,_w 实            事求是_i&gt; ,_w &lt;V&gt;抓住_v &lt;N&gt;机遇            _n&lt;/N&gt;&lt;/V&gt; ,_w &lt;开拓进取            _l&gt;&lt;/F<b>a</b>&gt; ,_w &lt;N&gt;&lt;V&gt;建设_v &lt;V&gt;有            _v &lt;N&gt;中国_ns 特色_n 社会主义            _n&lt;/N&gt;&lt;/V&gt;&lt;/V&gt; 的_u 道路_n&lt;/N&gt;            就_c &lt;V&gt;会_v &lt;F<b>c</b>&gt;越_d 走_v 越_d            宽广_a&lt;/F<b>c</b>&gt;&lt;/V&gt; 。_w&lt;/F<b>m</b>&gt;            &lt;F<b>m</b>&gt;&lt;F<b>a</b>&gt;zhiyao_c &lt;N&gt;women_r&lt;/N&gt;            &lt;R&gt;jinyibu_d&lt;/R&gt; &lt;jiefangsixiang_i ,            _w shishiqiushi_i&gt; ,_w            &lt;V&gt;zhuazhu_v &lt;N&gt;jiyu_n&lt;/N&gt;&lt;/V&gt; ,            _w &lt;kaituojinqu_l&gt;&lt;/F<b>a</b>&gt; ,_w            &lt;N&gt;&lt;V&gt;jianshe_v &lt;V&gt;you_v            &lt;N&gt;Zhongguo_ns tese_n            shehuizhuyi_n&lt;/N&gt;&lt;/V&gt;&lt;/V&gt; de_u            daolu_n&lt;/N&gt; jiu_c &lt;V&gt;hui_v            &lt;F<b>c</b>&gt;yue_d zou_v yue_d            kuanguang_a&lt;/F<b>c</b>&gt;&lt;/V&gt; 。_w&lt;/F<b>m</b>&gt;            “If we become more open-minded and            realistic and make every effort to explore            new possibilities, we are in a better position            to set up an ideology which can fully            represent Chinese values.”</p>
Adverbial Idiom/Set Phrase	Ia	<p>&lt;I<b>a</b>&gt;坚定不移_i 地_u&lt;/I<b>a</b>&gt;            &lt;I<b>a</b>&gt;jiandingbuyi_i de_u&lt;/I<b>a</b>&gt;            “persistently”</p>

Nonterminal category	Symbol	Example
		<p>&lt;Ia&gt;满怀信心_l 地_u&lt;/Ia&gt;            &lt;Ia&gt;manhuaixinxin_l de_u&lt;/Ia&gt;            “confidently”</p>
Adjective Phrase	J	<p>&lt;J&gt;非常_d 重要_a 的_u&lt;/J&gt; &lt;J&gt;很_d 不_d 平凡_a 的_u&lt;/J&gt; 一_m 年_q            &lt;J&gt;feichang_d zhongyao_a de_u&lt;/J&gt;            &lt;J&gt;hen_d bu_d pingfan_a de_u&lt;/J&gt;            yi_m nian_q            “a very important and extraordinary year”</p>
Adverbial Adjective Phrase	Ja	<p>&lt;Ja&gt;成功_a 地_u&lt;/Ja&gt; &lt;V&gt;召开_v 了_u &lt;N&gt;第十五_m 次_q 全国_n 代表大会_n&lt;/N&gt;&lt;/V&gt;            &lt;Ja&gt;chenggong_a de_u&lt;/Ja&gt;            &lt;V&gt;zhaokai_v le_u &lt;N&gt;dishiwu_m ci_q Quanguo_n Daibiaodahui_n&lt;/N&gt;&lt;/V&gt;            “successfully held the 15<sup>th</sup> National People’s Congress”</p>
Noun Phrase	N	<p>&lt;N&gt;百年_m 历史_n&lt;/N&gt;            &lt;N&gt;bainian_m lishi_n&lt;/N&gt;            “a hundred years’ history”</p>
Adverbial Noun Phrase	Na	<p>&lt;Na&gt;今天_t 上午_t&lt;/Na&gt;            &lt;Na&gt;jintian_t shangwu_t&lt;/Na&gt;            “this morning”            &lt;Na&gt;现在_t&lt;/Na&gt;            &lt;Na&gt;xianzai_t&lt;/Na&gt;            “at present”</p>
Prepositional Phrase	P	<p>&lt;S&amp;&gt;&lt;N&gt;[中国_ns 政府_n]nt&lt;/N&gt; &lt;Ja&gt;顺利_ad&lt;/Ja&gt; &lt;V&gt;恢复_v &lt;P&gt;对_p &lt;N&gt;香港_ns&lt;/N&gt;&lt;/P&gt; &lt;Vo&gt;行使_v &lt;N&gt;主权_n&lt;/N&gt;&lt;/Vo&gt;&lt;/V&gt;&lt;/S&amp;&gt;            &lt;S&amp;&gt;&lt;N&gt;[Zhongguo_ns Zhengfu_n]nt&lt;/N&gt; &lt;Ja&gt;shunli_ad&lt;/Ja&gt;</p>

Nonterminal category	Symbol	Example
		<p>&lt;V&gt;huifu_v &lt;P&gt;dui_p            &lt;N&gt;Xianggang_ns&lt;/N&gt;&lt;/P&gt;            &lt;Vo&gt;xingshi_v            &lt;N&gt;zhuquan_n&lt;/N&gt;&lt;/Vo&gt;&lt;/V&gt;&lt;/S&amp;&gt;</p> <p>“The Chinese government has succeeded in resuming its sovereignty over Hong Kong”</p>
Adverbial Prepositional Phrase	Pa	<p>&lt;S N="7"&gt;&lt;Pa&gt;在_p &lt;N&gt;这_r 一_m 年_q 中_f&lt;/N&gt;&lt;/Pa&gt; ,_w &lt;N&gt;中国_ns 的_u &lt;N&gt;&lt;N&amp;&gt;改革_vn 开放_vn&lt;/N&amp;&gt; 和_c &lt;N+&gt;现代化_vn 建设_vn&lt;/N+&gt;&lt;/N&gt;&lt;/N&gt; &lt;V&gt;继续_v &lt;Vo&gt;向前_v &lt;Vo&gt;迈进_v&lt;/Vo&gt;&lt;/Vo&gt;&lt;/V&gt; 。_w&lt;/S&gt;</p> <p>&lt;S N="7"&gt;&lt;Pa&gt;zai_p &lt;N&gt;zhe_r yi_m nian_q zhong_f&lt;/N&gt;&lt;/Pa&gt; ,_w &lt;N&gt;Zhongguo_ns de_u &lt;N&gt;&lt;N&amp;&gt;gaige_vn kaifong_vn&lt;/N&amp;&gt; he_c &lt;N+&gt;xiandaihua_vn jianshe_vn&lt;/N+&gt;&lt;/N&gt;&lt;/N&gt; &lt;V&gt;jixu_v &lt;Vo&gt;xiangqian_v &lt;Vo&gt;majin_v&lt;/Vo&gt;&lt;/Vo&gt;&lt;/V&gt; 。_w&lt;/S&gt;</p> <p>“In this year, our nation is progressing in our reform and modernisation.”</p>
Adverb Phrase	R	<p>&lt;R&gt;还_d 不_d&lt;/R&gt;            &lt;R&gt;hai_d bu_d&lt;/R&gt;            “not...though”</p>
Sentence (including direct speech quotation, also with & and + as co-ordinates)	S	<p>&lt;S N="1"&gt;&lt;Pa&gt;在_p &lt;N&gt;这_r 一_m 年_q 中&lt;/N&gt;_f&lt;/Pa&gt; ,_w &lt;N&gt;中国_ns 的_u 外交_n 工作_vn&lt;/N&gt; &lt;V&gt;取得_v 了_u &lt;N&gt;重要_a 成果_n&lt;/N&gt;&lt;/V&gt; 。_w&lt;/S&gt;</p> <p>&lt;S N="1"&gt;&lt;Pa&gt;zai_p &lt;N&gt;zhe_r yi_m nian_q zhong&lt;/N&gt;_f&lt;/Pa&gt; ,_w &lt;N&gt;Zhongguo_ns de_u waijiao_n</p>

Nonterminal category	Symbol	Example
		<p><i>gongzuo_vn</i>&lt;/N&gt; &lt;V&gt;<i>qude_v le_u</i>  &lt;N&gt;<i>zhongyao_a</i>  <i>chengguo_n</i>&lt;/N&gt;&lt;/V&gt; ° <i>_w</i>&lt;/S&gt;</p> <p>“In this year, the Chinese government has gained great success in its diplomacy.”</p>
Verb Phrase	V	<p>&lt;V&gt;<i>zhanwang_v</i> &lt;N&gt;<i>xin_a de_u</i>  <i>_n</i>&lt;/N&gt;&lt;/V&gt;</p> <p>&lt;V&gt;<i>zhanwang_v</i> &lt;N&gt;<i>xin_a de_u</i>  <i>sheji_n</i>&lt;/N&gt;&lt;/V&gt;</p> <p>“have hope in the new era”</p>
Adverbial Verb Phrase	Va	<p>&lt;Va&gt;<i>biaoxiang_v</i> &lt;N&gt;&lt;<i>chongman_v</i> <i>xiwang_n</i> <i>de_u</i>  1998_nian_t 之际_f&lt;/N&gt;&lt;/Va&gt;</p> <p>&lt;Va&gt;<i>maixiang_v</i> &lt;N&gt;&lt;<i>chongman_v</i>  <i>xiwang_n de_u</i>&gt; 1998_nian_t  <i>zhiji_f</i>&lt;/N&gt;&lt;/Va&gt;</p> <p>“at the moment we are looking forward to the prosperous year of 1998”</p> <p>&lt;Va&gt;<i>tuixiu_v</i> 前_f&lt;/Va&gt;</p> <p>&lt;Va&gt;<i>tuixiu_v</i> <i>qian_f</i>&lt;/Va&gt;</p> <p>“before retirement”</p>
Verbal Object	Vo	<p>&lt;V&gt;<i>xiwang_v</i> &lt;Vo&gt;<i>yikao_v</i> &lt;N&gt;<i>dajia</i>  <i>_r</i>&lt;/N&gt;&lt;/Vo&gt;&lt;/V&gt;</p> <p>&lt;V&gt;<i>xiwang_v</i> &lt;Vo&gt;<i>yikao_v</i>  &lt;N&gt;<i>dajia_r</i>&lt;/N&gt;&lt;/Vo&gt;&lt;/V&gt;</p> <p>“wish to rely on you”</p>
Initial Conjunct	&	<p>&lt;N&amp;&gt;<i>zhongguo_ns</i> <i>gaige_v</i>&lt;/N&amp;&gt; 和_c  &lt;N&amp;&gt;<i>Zhongguo_ns</i> <i>gaige_v</i>&lt;/N&amp;&gt;</p> <p><i>he_c</i></p> <p>“China’s revolution”</p>
Non-initial Conjunct	+	<p>和_c &lt;N+&gt;<i>fazhan_v</i> <i>de_u</i> <i>quanju_n</i>  &lt;/N+&gt;</p> <p><i>he_c</i> &lt;N+&gt;<i>fazhan_v</i> <i>de_u</i>  <i>quanju_n</i>&lt;/N+&gt;</p> <p>“and the entire development”</p>

Table 3: The list of constituent labels for the PFR Sample Skeleton Treebank parsing scheme.

Before dealing with point three (see section 3.5), I would first like to discuss some issues relating to the use of this labelling scheme in my treebank.

### 3.4.2.1 Selection and coding of parsing labels

The hand-parsing task was simplified as much as possible by the use of a simple bracketing notation. This was used to annotate 17 non-terminal categories which were considered to be “canonical”, as Leech and Garside (1991: 22) put it, i.e. likely to be uncontroversial, and to remain unaffected by differences of theory. The following table demonstrates the importance of adopting canonical nonterminal categories, by comparing seven different syntactic theories with respect to their phrasal categories, namely Transformational Grammar or TG (Radford, 1988), Government and Binding Theory or GB (Haegeman, 1994), Generalised Phrase Structure Grammar or GPSG (Gazdar et al., 1985), Head-driven Phrase Structure Grammar or HPSG (Pollard and Sag, 1994), Functional Grammar or FG (Dik, 1978, 1989 and 1997), Lexical-Functional Grammar or LFG (Bresnan, 2001), and Relational Grammar or RG (Blake, 1990). It shows that S, NP, VP, AP, PP and ADVP are the six core phrasal categories that exist among these theories. Additionally, the representation of conjuncts is different across theories: some theories (e.g. TG, GB, LFG, etc.) do not overtly mark the relationship of coordination between phrases, while others (i.e. GPSG and HPSG) represent the conjuncts using different abstract categorial labels. In the light of these differences across theories, in my bracketing system, the six core syntactic categories which are common to almost all of the syntactic theories above were adopted, and simple symbols (i.e. “&” and “+”) were used to indicate a coordinate relationship between two or more phrases.

	S	NP	VP	AP	PP	ADVP	DP	IP	CP	X[CONJ NIL], X[CONJ and]	RP	[SUBCAT <N>]
<b>TG</b>	+	+	+	+	+	+	+	+	+	-	-	-
<b>GB</b>	+	+	+	+	+	+	+	+	+	-	-	-
<b>GPSG</b>	+	+	+	+	+	+	-	-	-	+	-	-
<b>HPSG</b>	+	+	+	+	+	+	+	-	-	-	+	+
<b>FG</b>	-	+	-	-	+	-	-	-	-	-	-	-
<b>LFG</b>	+	+	+	+	+	+	+	+	+	-	-	-
<b>RG</b>	+	+	+	-	+	-	-	-	-	-	-	-

Table 4: A comparison of phrasal categories in seven syntactic theories.<sup>12</sup>

Apart from these eight canonical nonterminal categories (i.e. six core syntactic categories plus two symbols for initial and non-initial conjuncts), I added six other parsing labels which focussed on the adverbial, i.e. *Fa* (adverbial clause), *Ia* (adverbial idiom/set phrase), *Ja* (adverbial adjective phrase), *Na* (adverbial noun phrase), *Pa* (adverbial prepositional phrase), and *Va* (adverbial verb phrase), because I was particularly interested in adverbials in this book in which adverbial clauses were investigated. To explicitly indicate the clause(s) to which the adverbial clause is subordinated, the constituent label *Fm* which stands for the main clause, was also included in the bracketing system. The last two parsing labels I added were correlative clause, *Fc*, and verbal object, *V<sub>o</sub>*, which are specific to the Chinese language (see section 3.5.1.2 and 3.5.1.15).

Like the parsing symbols used for the Lancaster/IMB Treebank, the distinction between constituent tags and word tags is reflected in their coding: nonterminal categories (or syntactic constituents) consist of one upper case letter, optionally followed by one lower case letter

12 “+” indicates that the syntactic theory has the phrasal category, while “-” indicates that it does not. S stands for sentence; NP stands for noun phrase; VP stands for verb phrase; AP stands for adjective phrase; PP stands for prepositional phrase; ADVP stands for adverb phrase; DP stands for determiner phrase; IP stands for a clause with inflection (INFL) as its head; CP stands for a clause with complementiser (C) as its head; X[CONJ NIL] stands for an initial conjunct of any phrasal category (X), while X[CONJ and] stands for a final conjunct of any phrasal category (X); RP stands for any phrase projected from a relativiser; [SUBCAT <N>] stands for both the initial and final conjuncts which belong to the same phrasal category and thus must be marked with the same number N.

(representing feature values), e.g. “N” stands for noun phrase, “Na” refers to adverbial noun phrase, etc.; and terminal categories (or word classes), on the other hand, include just one or two lower case letters or one capital letter followed by one lower case letter, e.g. “n” for noun, “ns” for place name, “Ng” for “nominal morpheme”.<sup>13</sup>

As we have seen, upper case letters were used to represent the constituent labels. However, the reverse happens in the annotation standard established by the Text Encoding Initiative (TEI): parsing labels are represented by lower case letters (rather than upper case ones as I used here), and POS tags are denoted by upper case letters. Since the compilers of the PFR Chinese Corpus used small letters to represent POS tags (Yu, 1999), in order to distinguish from the originally marked POS annotation, I use capital letters for the parsing labels, though I am aware that this goes against the TEI guidelines. It does, however, improve the readability of my treebank.

### 3.4.2.2 Set of annotation devices used

It is important to note that in my annotation scheme, the labels used to mark the grammatical structure of a sentence are not enclosed in a pair of square brackets as in the UCREL format (Leech and Garside, 1991); rather they are circumscribed by angled brackets as I wished to provide my treebank in a TEI-conformant interchange format, i.e. the syntactic information is enclosed using XML tags delimited by `<...>` and `</...>`. The reason for adopting this TEI-conformant style rather than the UCREL labelled-bracketing format is that the resulting skeleton parsed sample corpus can then be more readily re-used by many different users developing or testing natural language processing (NLP) software of varied kinds, for example, then would be possible using a “bespoke” format.

13 As we have seen, both the adverbials (e.g. adverbial noun phrase (*Na*)) and morphemes (e.g. nominal morpheme (*Ng*)) have the same coding format i.e. a capital letter followed by a lower case letter. However, as the latter is rare, syntactic tags (my annotation) and word tags (original annotation) are contrasting visually in most parts of my treebank.

### 3.4.2.3 Insertion of textual annotations

As the original PFR Chinese Corpus does not provide any textual markup such as the title of a text, this information was inserted in my sample treebank using TEI markup elements.<sup>14</sup> The following gives the various kinds of information added:

- (a) `<CAPTION>` `</CAPTION>` contains the text of a caption of an illustration or a photograph printed in the original newspaper
- (b) `<DATE>` `</DATE>` gives the date of an event described in the body of a text
- (c) `<ILLUSTRATION>` `</ILLUSTRATION>` indicates an illustration or a photograph in the original text
- (d) `<JOURNALIST>` `</JOURNALIST>` encloses a journalist's name
- (e) `<NAME>` `</NAME>` indicates a proper noun referring to a person, a place or an organisation mentioned in the header
- (f) `<ORIGINALMP>` `</ORIGINALMP>` encloses the original markup of the PFR Chinese Corpus, which includes the date, page no., article no. and paragraph no. of the newspaper text
- (g) `<SOURCEDESC>` `</SOURCEDESC>` identifies the source of a text
- (h) `<SUBTITLE>` `</SUBTITLE>` denotes a subtitle in a text
- (i) `<TITLE>` `</TITLE>` encloses the title or headline of a text
- (j) `<UNCERTAIN>` `</UNCERTAIN>` indicates an unknown unit in the text
- (k) `N="..."` provides a sentence number

### 3.4.2.4 Layers of annotation undertaken

The EAGLES documentation (Kahrel et al., 1997: 237ff.) specifies a number of different layers of annotation, the first and second of which were adopted in my skeleton parsing scheme. The first layer, bracketing of segments, involves the delimitation of sentence segments which have a "syntactic integrity" (e.g. sentences, clauses, phrases, words) by nested pairs of angled brackets. The second layer, labelling of segments, indi-

14 These textual markers are based on the elements proposed in the guidelines of the Text Encoding Initiative (Sperberg-McQueen and Burnard (Eds.), 2001), except `<ILLUSTRATION>`, `<JOURNALIST>`, `<ORIGINALMP>` and `<SUBTITLE>` which have been invented for the purposes of this research.

cates the formal category of the non-terminal syntactic constituents identified by bracketing, such as noun phrase, verb phrase, adverbial clause, etc.

The other six layers, due to the complexity of analysis which they entail, were ignored for the purposes of this research. For the record, these unannotated layers would show: the relation of a head to its dependent(s) in a phrase or clause, syntactic function labels (e.g. subject, object, adjunct, etc.), subcategorisation or feature values of sentences (e.g. declarative, imperative, interrogative) and phrases (e.g. marking noun phrase as singular), deep/logical relations (e.g. co-referentiality, ellipsis, traces, etc.), ranking of syntactic categories (e.g. phrases being of lower rank than clauses, clauses being of lower rank than sentences), non-fluency phenomena (e.g. false starts, reiterations, etc.).

### 3.5. Guidelines of skeleton parsing

Having reviewed my annotation scheme in some depth, I can now present my guidelines for annotation, in keeping with my desire to match the advantages of Sampson's SUSANNE scheme (see section 3.4.2). The parsing scheme matches features (a) and (b) of a clear and explicit parsing scheme (Leech and Eyes, 1997: 37). Nonetheless, an annotation scheme is more than (a) and (b) above. Feature (c), a set of *parsing guidelines* should also be provided in order to explain how the parsing symbols are actually applied to text sentences to avoid undesirable inconsistency. Ideally, complete guidelines should be available to the annotators before annotation begins. However, linguistic problems posed by corpora are much more diverse and complicated than those discussed in theoretical linguistics or grammar books, and new problems arise as more data is annotated. Hence, the guidelines must be revised, updated and enriched incrementally as the annotation process progresses. Such guidelines are likened to case law as it represents a continuously refined system which changes as new cases occur, rather than a set of unchanging hard and fast rules.

During the course of annotation, as more data was analysed, the guidelines took shape and were recorded and updated. Hence, after I finished the task of parsing a sample text taken from the PFR Chinese

Corpus, I had a set of guidelines ready to be consolidated into a document to be available to users of the treebank and to future annotators who might want to adopt the same scheme. It is advisable, as Kahrel et al. (1997: 241 ff.) note, to document explicitly all of the decisions taken in the development of an annotation scheme, as well as its application so that future users can apply the scheme in a manner consistent with that of the originators of the scheme.

In section 3.5.1, I will give a clear account of the 17 non-terminal constituents that are defined and instantiated in my texts, with illustrative examples taken from the parsed sample treebank as well as an explanation of problem cases. Apart from a detailed treatment of parsing symbols<sup>15</sup>, my documented parsing guidelines include practical issues related to map any parses on to sentences in the application of the parsing scheme, which will be discussed in section 3.5.2.

### 3.5.1 Detailed description of parsing symbols

#### 3.5.1.1 Adverbial clause (Fa)

An adverbial clause functions as an adjunct. The occurrence of an adverbial clause, which is a type of subordinate clause, is dependent on that of a main clause which it modifies. However, omission of the adverbial clause does not affect the grammaticality of the main clause (Hengeveld, 1998; Quintero, 2002).

(2) <Fm><Fa>只有\_c <解放思想\_i 、\_w 实事求是\_i></Fa> ，\_w 才\_c <V>能\_v 冲破\_v <N>各种\_r <<不\_d 切合\_v 实际\_n 的\_u 或者\_c 过时\_a 的\_u> 观念\_n 的\_u> 束缚\_vn<N></V> ，\_w <R>真正\_d</R> <V>做到\_v <V><尊重\_v 、\_w 认识\_v 和\_c 掌握\_v> <N>客观\_a 规律\_n<N></V></V> ，\_w <R>勇于\_d</R> <V>突破\_v</V> ，\_w <R>勇于\_d</R> <V>创新\_v</V> ，\_w <R>不断\_d</R> <V>开创\_v <N><社会主义\_n 现代化\_vn 建设\_vn 的\_u> 新\_a 局面\_n<N></V></Fm>

15 When writing this detailed description, I was informed by the EAGLES Recommendations for the Syntactic Annotation of Corpora, Version of 11<sup>th</sup> March 1996. See Leech, Barnett and Kahrel (1996) available at EAGLES's web page: <<http://www.ilc.cnr.it/EAGLES96/segsasg1/segsasg1.html>> (accessed 15 May 2012).

<Fm><Fa>zhiyou\_c <jiefangsixiang\_i \\_w shishiqiushi\_i></Fa> \\_w  
 only.if be.open-minded realistic  
 cai\_c <V>neng\_v chongpo\_v <N>gezhong\_r <<bu\_d qiehe\_v  
 then can get.rid.of different.kinds.of not in.line.with  
 shiji\_n de\_u> huozhe\_c <guoshi\_a de\_u guannian\_n  
 pragmatism DE or obsolete DE values  
 de\_u>> shufu\_vn</N></V> \\_w <R>zhenzheng\_d</R> <V>zuodao\_v  
 DE restriction really realise  
 <V><zhuangzhong\_v \\_w renshi\_v he\_c zhangwo\_v> <N>keguan\_a  
 respect understand and master practical  
 guiliu\_n</N></V></V> \\_w <R>yongyu\_d</R> <V>tupo\_v</V> \\_w  
 principles have.the.guts.to breakthrough  
 <R>yongyu\_d</R> <V>chuangxin\_v</V> \\_w <R>buduan\_d</R>  
 have.the.guts.to break.fresh.ground continuously  
 <V>kaichuang\_v <N><shehuizhuyi\_n xiandaihua\_vn jianshe\_vn de\_u>  
 create social.ideology modern construction DE  
 xin\_a jiumian\_n</N></V></Fm>  
 new situation

“Only if we become more open-minded and realistic, can we overcome the restrictions imposed by outdated practices, and know how to respect, understand and take advantage of what we learn in everyday life, thereby breaking through and continuously building an efficient modern society.”

- (3) <S N="21"><Fm><Fa>只要\_c <N>我们\_r</N> <R>进一步\_d</R> <解放思想\_i \\_w 实事求是\_i> \\_w <V>抓住\_v <N>机遇\_n</N></V> \\_w <开拓进取\_]></Fa> \\_w <N><建设\_v 有\_v 中国\_ns 特色\_n 社会主义\_n 的\_u> 道路\_n</N> 就\_c <V>会\_v <Fc>越\_d 走\_v 越\_d 宽广\_a</Fc></V> 。\_w</Fm></S>
- <S N="21"><Fm><Fa>zhiyao\_c <N>women\_r</N> <R>jinyibu\_d</R>  
 provided.that we further  
 <jiefangsixiang\_i \\_w shishiqiushi\_i> \\_w <V>zhuazhu\_v  
 be.open-minded realistic grasp  
 <N>jiyu\_n</N></V> \\_w <kaituojinqu\_]></Fa> \\_w <N><jianshe\_v  
 opportunity work.hard.and.explore establish  
 you\_v Zhongguo\_ns tese\_n shehuizhuyi\_n de\_u>  
 have China characteristics social.ideology DE

daolu\_n</N> jiu\_c <V>hui\_v <Fc>yue\_d zou\_v yue\_d  
road then will the.more walk the.more  
kuanguang\_a</Fc></V> ° \_w</Fm></S>

broad

“If we become more open-minded and realistic, and grasp every chance to explore new possibilities, we can come closer to building a society fully representing our national characteristics.”

The first clauses of the following two sentences and sentences like them, however, were not accepted by me as adverbial clauses proper, though some scholars may accept them as such (e.g. Rai-Chu, 1982: 21–30; Liu et al., 1996: 505–513; Chu and Chi, 1999: 248f.). I did not consider them as adverbial clauses because these clauses do not show any explicit sign of their dependence on a main clause, i.e. the following clause. Chao (1968: 104–106) treats such kind of sentences as coordinate sentences or even separate sentences. As will be discussed in the next chapter (Chapter Four, section 4.1), English adverbial clauses, if not overtly marked by a subordinating conjunction to be subordinate clauses, are always non-finite clauses. Chinese adverbial clauses or clauses in general, however, do not have this finite/non-finite distinction (Norman, 1988: 163). Since it is not clear whether the first clauses of the following examples are subordinate clauses, they were not taken as adverbial clauses in my treebank.

- (4) <S N="57"><R>进一步\_d</R> <Ja>认真\_ad</Ja> <V><学习\_v 和\_c 掌握\_v> <N>十五大\_j 精神\_n</N></V> ,\_w <解放思想\_i 、\_w 实事求是\_i> ,\_w <N>我们\_r 的\_u 各项\_r 事业\_n</N> <R>就\_d</R> <V>能\_v 结\_v 出\_v <N><更加\_d 丰硕\_a 的\_u> 成果\_n</N></V> °  
\_w</S>
- <S N="57"><R>jinyibu\_d</R> <Ja>renzhen\_ad</Ja> <V><xuexi\_v he\_c zhangwo\_v> <N>shiwuda\_j jingshen\_n</N></V> ,\_w  
master 15.primary principles  
<jiefangsixiang\_i 、\_w shishiqiushi\_i> ,\_w <N>women\_r de\_u  
be.open.minded realistic our GEN  
gexiang\_r shiye\_n</N> <R>jiu\_d</R> <V>neng\_v jie\_v chu\_v  
various.kinds.of undertaking then can result.in out

<N><gengjia\_d fengshuo\_a de\_u> chengguo\_n</N></V> °\_w</S>  
 even.more promising result

“If we can better understand and master the Communist Party’s 15 principles and become more open-minded and realistic, we can gain greater success in every venture we make.”

(5) <S N="114"><Na>现在\_t</Na> <N>条件\_n</N> <J>好\_a 了\_y</J> ,\_w  
 但\_c <N><艰苦奋斗\_i 、\_w 无私奉献\_i 的\_u> 精神\_n</N> <R>可  
 \_d</R> <V>不能\_v 丢\_v</V> °\_w ”\_w</S>

<S N="114"><Na>xianzai\_t</Na> <N>tiaojian\_n</N> <J>hao\_a  
 at.present condition better  
 le\_y</J> ,\_w dan\_c <N><jiankufengdou\_i 、\_w  
 PERF but work.hard  
 wusifengxian\_i de\_u jingshen\_n</N> <R>ke\_d</R>  
 contribute.unconditionally DE spirit yet  
 <V>buneng\_v diao\_v</V> °\_w ”\_w</S>  
 cannot lose

“Although the situation is improved, we must continue to strive hard and contribute to our society unconditionally.”

### 3.5.1.2 Correlative clause (Fc)

The term “correlative clause” is proposed by Chao (1968: 121ff.) to refer to two coordinated clauses which are in parallel, in contrast, or in some other close relation, making a pair of dependent clauses typically marked by the correlative adverbs 越 yue ... 越 yue “the more..., the more...” and repeated interrogative indefinite pronouns. The last two examples below are taken from Chao’s analysis (1968: 121–122) as no examples of this sort can be found in my sample treebank.

(6) <S N="21"><Fm><Fa>只要\_c <N>我们\_r</N> <R>进一步\_d</R> <解放思  
 想\_i ,\_w 实事求是\_i> ,\_w <V>抓住\_v <N>机遇\_n</N></V> ,\_w  
 <开拓进取\_1></Fa> ,\_w <N><建设\_v 有\_v 中国\_ns 特色\_n 社会主  
 义\_n 的\_u> 道路\_n</N> 就\_c <V>会\_v <Fc>越\_d 走\_v 越\_d 宽  
 广\_a</Fc></V> °\_w</Fm></S>

<S N="21"><Fm><Fa>zhiyao\_c <N>women\_r</N> <R>jinyibu\_d</R>  
 provided.that we further

<jiefangsixiang\_i ' \_w shishiqiushi\_i> ' \_w <V>zhuazhu\_v  
 be.open-minded realistic grasp  
 <N>jiyu\_n</N></V> ' \_w <kaituojinqin\_l></Fa> ' \_w <N><jianshe\_v  
 opportunity work.hard.and.explore establish  
 you\_v Zhongguo\_ns tese\_n shehuizhuyi\_n de\_u> daolu\_n</N>  
 have China characteristics social.ideology DE road  
 jiu\_c <V>hui\_v <Fc>yue\_d zou\_v yue\_d kuanguang\_a</Fc></V> 。  
 then will the.more walk the.more broad  
 \_w</Fm></S>

“If we become more open-minded and realistic, and grasp every chance to explore new possibilities, we can come closer to building a society fully representing our national characteristics.”

- (7) <Fc>谁 先 来 谁 先 吃。 </Fc>  
 <Fc>Shui xian lai shui xian chi</Fc>  
 who early come who early eat

“Whoever comes first eats first.”

- (8) <Fc>哪儿 舒服 哪儿 睡。 </Fc>  
 <Fc>Naer shufu naer shui</Fc>  
 wherever comfortable wherever sleep

“Sleep wherever it is comfortable.”

### 3.5.1.3 Main clause (Fm)

A main clause is the clause(s) which the adverbial clause is subordinated to and modifies.<sup>16</sup>

- (9) <S N="20"><Fm><Fa>尽管\_c <N>我们\_r</N> <P>在\_p <N>经济\_n 社  
 会\_n 发展\_v 中\_f</N></P> <R>还\_d</R> <V>面临\_v <N>不少\_m  
 困难\_an</N></V></Fa> ' \_w 但\_c <N>我们\_r</N> <V>有\_v <N>邓小

16 Although in examples (3) and (9) the main clause is the same as the sentence, this does not necessarily happen as in example (2), where the main clause is part of the whole sentence only. I distinguished main clause from sentence because in some cases like example (2), some clauses of the sentence (i.e. those clauses which are not included in the main clause) are hardly semantically relevant to the meaning of the adverbial clause.

平理论\_n 的\_u 指引\_vn</N></V> ，\_w <V>有\_v <N><改革\_v 开放\_v 近\_a 20\_m 年\_q 来\_f> <N><N&>取得\_v 的\_u 伟大\_a 成就\_n</N&> 和\_c <N+>积累\_v 的\_u 丰富\_a 经验\_n</N+></N></N></V> ，\_w <R>还\_d</R> <V>有\_v <N>其他\_r 的\_u 各种\_r 有利\_a 条件\_n</N></V> ，\_w <N>我们\_r</N> <R>一定\_d</R> <V>能够\_v 克服\_v <N>这些\_r 困难\_an</N></V> ，\_w <V>继续\_v <稳步前进\_l></V> 。\_w</Fm></S>

<S N="20"><Fm><Fa>jinguan\_c <N>women\_r</N> <P>zai\_p  
even.though we in  
<N>jingji\_n shehui\_n fazhan\_v zhong\_f</N></P> <R>hai\_d</R>  
economic society development within still  
<V>mianlin\_v <N>bushao\_m kunnan\_an</N></V></Fa> ，\_w  
face quite.a.number.of difficulties  
dan\_c <N>women\_r</N> <V>you\_v <N>Dengxiaopinglilun\_n  
but we have the.principles.of.Deng.Xiaoping  
de\_u zhiyin\_vn</N></V> ，\_w <V>you\_v <N><gaige\_v kaifang\_v  
DE instruction have reform open-up  
jin\_a 20\_m nian\_q lai\_f> <N><N&>qude\_v de\_u weida\_a  
near 20 years so.far obtain DE great  
chengjiu\_n</N&> he\_c <N+>jilei\_v de\_u fengfu\_a  
achievements and accumulate DE rich  
jingyan\_n</N+></N></N></V> ，\_w <R>hai\_d</R> <V>you\_v  
experience additionally have  
<N>qita\_r de\_u gezhong\_r youli\_a tiaojian\_n</N></V> ，  
other DE various.kinds.of advantageous conditions  
\_w <N>women\_r</N> <R>yiding\_d</R> <V>nenggou\_v kefu\_v  
we certainly can overcome  
<N>zhexie\_r kunnan\_an</N></V> ，\_w <V>jixu\_v  
these difficulties continue  
<wenbuqianjin\_l></V> 。\_w</Fm></S>

proceed.steadily

“Although we are still encountering enormous difficulties in developing our economy, we can seek help from the set of principles established by our former president Deng Xiaoping, and benefit from the success of reform and open policy throughout the past 20 years, our own rich experience and many other advantageous factors, we must definitely get over all these difficulties and proceed ahead steadily.”

### 3.5.1.4 Adverbial idiom/set phrase (Ia)

An adverbial idiom or set phrase functions as a modifier of a verb phrase.

- (10) <S N="19"><Na> 1 9 9 8 年\_t</Na> ,\_w <N>中国\_ns 人民\_n</N> <R> 将\_d</R> <Ia>满怀信心\_l 地\_u</Ia> <V>开创\_v <N>新\_a 的\_u 业 绩\_n</N></V> 。\_w</S>  
 <S N="19"><Na> 1 9 9 8 nian\_t</Na> ,\_w <N>Zhongguo\_ns  
 year.of.1998 China  
 renmin\_n</N> <R>jiang\_d</R> <Ia>manhuaixinxin\_l de\_u</Ia>  
 citizens soon confidently ADVL  
 <V>kaichuang\_v <N>xin\_a de\_u yeji\_n</N></V> 。\_w</S>  
 create new DE accomplishment

"In 1998, the Chinese people will have confidence in achieving great success."

- (11) <S N="42"><Pa>在\_p <N>这\_r 一\_m 年\_q</N></Pa> ,\_w <P>以\_p  
 <N>江\_nr 泽民\_nr 同志\_n</N></P> <V>为\_v <N>核心\_n 的\_u 党中 央\_nt</N></V> ,\_w <V>继承\_v <N><邓\_nr 小平\_nr 同志\_n 的\_u> 遗志\_n</N></V> ,\_w <V>高举\_v <N>邓小平理论\_n 的\_u 伟大\_a 旗帜\_n</N></V> ,\_w <V><V&>领导\_v <N><N&>全党\_n</N&> 和\_c <N+>全国\_n 各族\_r 人民\_n<N+></N></V&> <V+><Ia>坚定不移\_i 地\_u</Ia> <P>沿着\_p <N><V>建设\_v <N><有\_v 中国\_ns 特色\_n> 社会主义\_n</N></V> 道路\_n</N></P> <R>阔步\_d</R> 前进\_v</V+></V> ,\_w <V>写\_v 下\_v 了\_u <N><改革\_v 开放\_v 和\_c 社会主义\_n 现代化\_vn 建设\_vn 的\_u> 辉煌\_a 篇章\_n</N></V> 。\_w</S>  
 <S N="42"><Pa>zai\_p <N>zhe\_r yi\_m nian\_q</N></Pa> ,\_w  
 in this one year  
 <P>yi\_p <N>Jiang\_nr Zemin\_nr tongzhi\_n</N></P> <V>wei\_v  
 with Jiang Zemin comrade to.be  
 <N>hexin\_n de\_u dangzhongyang\_nt</N></V> ,\_w <V>jicheng\_v  
 core DE central.government take.over  
 <N><Deng\_nr Xiaoping\_nr tongzhi\_n de\_u> yizhe\_n</N></V> ,\_w  
 Deng Xiaoping comrade's GEN wish  
 <V>gaoju\_v <N>Dengxiaopinglilun\_n de\_u weida\_a  
 hold the.principles.of.Deng.Xiaoping DE great

qizhi\_n</N></V> ' \_w <V><V&>lingdao\_v  
 flag lead  
 <N><N&>quandang\_n</N&> he\_c <N+>quanguo\_n  
 the.whole.party and across.the.country  
 gezu\_r renmin\_n</N+></N></V&>  
 various.ethnic.communities citizens  
 <V+><Ia>jiandingbuyi\_i de\_u</Ia> <P>yuanzhe\_p <N><V>jianshe\_v  
 persistently ADVL along build  
 <N><you\_v Zhongguo\_ns tese\_n> shehuizhuyi\_n</N></V>  
 have China characteristics social.ideology  
 daolu\_n</N></P> <R>kuobu\_d</R> qianjin\_v</V+></V> ' \_w  
 road take.a.big.step proceed  
 <V>xie\_v xia\_v le\_u <N><gaige\_v kaifang\_v he\_c shehuizhuyi\_n  
 write down PERF reform open-up and social.ideology  
 xiandaihua\_vn jianshe\_vn de\_u> huihuang\_a pianzhang\_n</N></V> 。  
 modern construction DE magnificent page  
 \_w</S>

"In this year, led by our President Jiang Zemin, the Communist Party followed the principles established by our former president Deng Xiaoping, persistently leading its members and our people towards a better society and thus erecting a milestone of reform and modernisation."

### 3.5.1.5 Adjective phrase (J)

An adjective phrase consists of an adjective, optionally followed by the adjectival marker 的 *de* (Zhu, 2000), or a group of words (including adverbs) in which an adjective is the head of the phrase.

- (12) <S N="31"><N><和平\_n 与\_c 发展\_vn 的\_u> 前景\_n</N> <V>是\_v  
 <J>光明\_a 的\_u</J></V> ' \_w <N> 2 1\_m 世纪\_n</N> <R>将\_d</R>  
 <V>是\_v <N><充满\_v 希望\_n 的\_u> 世纪\_n</N></V> 。\_w</S>  
 <S N="31"><N><heping\_n yu\_c fazhan\_vn de\_u> qianjing\_n</N>  
 peace and development DE prospects  
 <V>shi\_v <J>guangming\_a de\_u</J></V> ' \_w <N> 2 1\_m  
 be bright PART 21

shiji\_n</N> <R>jiang\_d</R> <V>shi\_v <N><chongman\_v xiwang\_n  
 century soon be full.of hope  
 de\_u> shiji\_n</N></V> °\_w</S>  
 DE century

“The prospects of maintaining peace and development are encouraging, making the 21<sup>st</sup> century a promising era.”

(13) <S N="32">但\_c <N>前进\_v 的\_u 道路\_n</N> <V><V><V&>不会  
 \_v</V&> <R>也\_d</R> <V+><R>不\_d</R> 可能\_v</V+></V> <—帆风  
 顺\_i></V> ,\_w <N>关键\_n</N> <V>是\_v <S><N>世界\_n 各国\_r 人  
 民\_n</N> <V>要\_v <J>进一步\_d 团结\_a 起来\_v</J></V></S></V> ,  
 \_w <V><V><V&><R>共同\_d</R> 推动\_v</V&> <V+><R>早日\_d</R>  
 建立\_v</V+></V> <N><公正\_a 合理\_a 的\_u 国际\_n 政治\_n 经济  
 \_n 新\_a 秩序\_n</N></V> °\_w</S>

<S N="32">dan\_c <N>qianjin\_v de\_u daolu\_n</N>  
 however proceed DE road  
 <V><V><V&>buhui\_v</V&> <R>ye\_d</R> <V+><R>bu\_d</R>  
 will.not as.well.as not  
 keneng\_v</V+></V> <yifanfengshun\_i></V> ,\_w <N>guanjian\_n</N>  
 can smooth the.key.to.success  
 <V>shi\_v <S><N>shijie\_n geguo\_r renmin\_n</N> <V>yao\_v  
 be world nations people have.to  
 <J>jinyibu\_d tuanjie\_a qilai\_v</J></V></S></V> ,\_w  
 further unite stand.up  
 <V><V><V&><R>gongtong\_d</R> tuidong\_v</V&>  
 together uphold  
 <V+><R>zaori\_d</R> jianli\_v</V+></V> <N><gongzheng\_a  
 as.soon.as.possible build just  
 heli\_a de\_u> guoji\_n zhengzhi\_n jingji\_n xin\_a  
 reasonable DE international political economic new  
 zhixu\_n</N></V> °\_w</S>  
 order

“Yet progress will not come easily without people from all over the world combining efforts in building a fair and just global political and economic climate.”

- (14) <S N="71"><N>这\_r</N> <V>是\_v <J>非常\_d 重要\_a 的  
 \_u</J></V> °\_w</S>  
 <S N="71"><N>zhe\_r</N> <V>shi\_v <J>feichang\_d zhongyao\_a  
 this be extremely important  
 de\_u</J></V> °\_w</S>  
 PART  
 "This is very important indeed."

### 3.5.1.6 Adverbial adjective phrase (Ja)

An adverbial adjective phrase functions as a modifier of a verb phrase, and may optionally be followed by the particle 地 *de* which is the adverbial marker in Chinese (Zhu, 1982).

- (15) <S N="6"><N>[中国\_ns 共产党\_n]nt</N> <Ja>成功\_a 地\_u</Ja> <V>  
 召开\_v 了\_u <N>第十五\_m 次\_q 全国\_n 代表大会\_n</N></V> ,\_w  
 <V>高举\_v <N>邓小平理论\_n 伟大\_a 旗帜\_n</N></V> ,\_w <V>总结  
 \_v <N>百年\_m 历史\_n</N></V> ,\_w <V>展望\_v <N>新\_a 的\_u  
 世纪\_n</N></V> ,\_w <V>制定\_v 了\_u <N><中国\_ns 跨\_v 世纪\_n  
 发展\_v 的\_u> 行动\_vn 纲领\_n</N></V> °\_w</S>  
 <S N="6"><N>[Zhongguo\_ns Gongchandang\_n]nt</N> <Ja>chenggong\_a  
 China the.Communist.Party successfully  
 de\_u</Ja> <V>zhaokai\_v le\_u <N>dishiwu\_m ci\_q quanguo\_n  
 ADVL hold PERF the.15<sup>th</sup> CL national  
 daibiaodahui\_n</N></V> ,\_w <V>gaoju\_v  
 congress hold  
 <N>Dengxiaopinglilun\_n weida\_a qizhi\_n</N></V> ,\_w  
 the.principles.of.Deng.Xiaoping great flag  
 <V>zongjie\_v <N>bainian\_m lishi\_n</N></V> ,\_w <V>zhanwang\_v  
 conclude hundred.years.history look.forward.to  
 <N>xin\_a de\_u shiji\_n</N></V> ,\_w <V>zhiding\_v le\_u  
 new DE century stipulate PERF  
 <N><Zhongguo\_ns kua\_v shiji\_n fazhan\_v de\_u> xingdong\_vn  
 China beyond century develop DE action  
 gangling\_n</N></V> °\_w</S>  
 scheme

“The Communist Party of the People’s Republic of China successfully held the 15<sup>th</sup> Annual General Meeting, reiterating the principles established by our former president Deng Xiaoping, summarising its success in the past hundred years, igniting hope in the new era, and laying down plans for future development of the country.”

(16) <S N="17"><N>中国\_ns</N> <Ja>积极\_ad</Ja> <V>参与\_v <N>[亚\_j 太\_j 经合\_j 组织\_n]nt 的\_u 活动\_vn</N></V> ,\_w <V>参加\_v 了\_u <N><N><N&>东盟\_ns 一\_w 中\_j 日\_j 韩\_j </N&>和\_c <N+>中国\_ns 一\_w 东盟\_ns 首脑\_n</N+></N> 非正式\_b 会晤\_vn</N></V> 。\_w</S>

<S N="17"><N>Zhongguo\_ns</N> <Ja>jiji\_ad</Ja> <V>canyu\_v  
 China actively take.part.in  
 <N>[Ya\_j Tai\_j Jinghe\_j Zuzhi\_n]nt de\_u  
 Asia Pacific Economic.Co-operation organisation DE  
 huodong\_vn</N></V> ,\_w <V>canjia\_v le\_u  
 activity take.part.in PERF  
 <N><N><N&>Dongmeng\_ns 一\_w Zhong\_j Ri\_j Han\_j </N&>he\_c  
 eastern.coalition China Japan Korea and  
 <N+>Zhongguo\_ns 一\_w Dongmeng\_ns shounao\_n</N+></N>  
 China eastern.coalition leading.role  
 feizhengshi\_b huiwu\_vn</N></V> 。\_w</S>  
 informal dialogue

“China took an active role in the activities organised by the Asia Pacific Economic Co-operation (APEC) and participated in the informal meeting with leaders of Japan and Korea.”

(17) <S N="53"><N>我们\_r</N> <V>要\_v <Ja>更\_d 好\_a 地\_u</Ja> <V>坚持\_v <N><解放思想\_i 、\_w 实事求是\_i 的\_u> 思想\_n 路线\_n</N></V></V> 。\_w</S>

<S N="53"><N>women\_r</N> <V>yao\_v <Ja>geng\_d hao\_a  
 we have.to more good  
 de\_u</Ja> <V>jianchi\_v <N><jiefangsixiang\_i 、\_w shishiqiushi\_i  
 ADVL insist.on be.open-minded realistic  
 de\_u> sixiang\_n luxian\_n</N></V></V> 。\_w</S>  
 DE ideology route

“We must closely follow the ideology of opening our mind and being more realistic.”

### 3.5.1.7 Noun phrase (N)

A noun phrase generally has a noun or a pronoun as its head which may be preceded by determinative elements and/or premodifiers typically marked by the particle 的 *de*.

- (18) <S N="29">但是\_c , \_w <N>世界\_n</N> <R>还\_d 不\_d</R> <J>安宁\_a</J> °\_w</S>  
 <S N="29">danshi\_c , \_w <N>shijie\_n</N> <R>hai\_d bu\_d</R>  
 however world still not  
 <J>anning\_a</J> °\_w</S>  
 tranquil  
 “Nevertheless, the world is not in peace.”

- (19) <S N="53"><N>我们\_r</N> <V>要\_v <Ja>更\_d 好\_a 地\_u</Ja> <V>坚持\_v <N><解放思想\_i , \_w 实事求是\_i 的\_u> 思想\_n 路线\_n</N></V></V> °\_w</S>  
 <S N="53"><N>women\_r</N> <V>yao\_v <Ja>geng\_d hao\_a de\_u</Ja>  
 we have.to more good ADVL  
 <V>jianchi\_v <N><jiefangsixiang\_i , \_w shishiqiushi\_i de\_u>  
 insist.on be.open-minded realistic DE  
 sixiang\_n luxian\_n</N></V></V> °\_w</S>  
 ideology route  
 “We must closely follow the ideology of opening our mind and being more realistic.”

- (20) <S N="22"><N><实现\_v 祖国\_n 的\_u> 完全\_a 统一\_vn</N> , \_w <V>是\_v <N><海内外\_s 全体\_n 中国\_ns 人\_n 的\_u> 共同\_b 心愿\_n</N></V> °\_w</S>  
 <S N="22"><N><shixian\_v zuguo\_n de\_u> wanquan\_a  
 realise motherland DE complete  
 tongyi\_vn</N> , \_w <V>shi\_v <N><haineiwai\_s quanti\_n  
 unification be local.and.overseas the.whole.group  
 Zhongguo\_ns ren\_n de\_u> gongtong\_b xinyuan\_n</N></V> °\_w</S>  
 China people DE shared wish  
 “Uniting the country entirely has long been the goal of both local and overseas Chinese people.”

### 3.5.1.8 Adverbial noun phrase (Na)

An adverbial noun phrase functions as an adjunct in a sentence.

- (21) <S N="19"><Na> 1 9 9 8 年\_t</Na> , \_w <N>中国\_ns 人民\_n</N>  
 <R>将\_d</R> <Ia>满怀信心\_l 地\_u</Ia> <V>开创\_v <N>新\_a 的\_u  
 业绩\_n</N></V> 。\_w</S>  
 <S N="19"><Na> 1 9 9 8 nian\_t</Na> , \_w <N>Zhongguo\_ns  
 year.of.1998 China  
 renmin\_n</N> <R>jiang\_d</R> <Ia>manhuaixinxin\_l de\_u</Ia>  
 citizens soon confidently ADVL  
 <V>kaichuang\_v <N>xin\_a de\_u yeji\_n</N></V> 。\_w</S>  
 create new DE success  
 "In 1998, the Chinese people will have confidence in achieving great success."

- (22) <S N="104"><Na>现在\_t</Na> , \_w <N>我们\_r 的\_u 首都\_n</N>  
 <R>已经\_d</R> <V>结束\_v 了\_u <N><拉\_v 闸\_n 限\_v 电\_n 的\_u>  
 历史\_n</N></V> , \_w <V>希望\_v <Vo>依靠\_v <N>大家\_r</N></Vo></V> , \_w <V>使\_v <N><拉\_v 闸\_n 限\_v 电\_n 的\_u>  
 历史\_n</N> <V><R>永远\_d 不再\_d</R> 重演\_v</V></V> 。\_w</S>  
 <S N="104"><Na>xianzai\_t</Na> , \_w <N>women\_r de\_u  
 at.present our GEN  
 shoudu\_n</N> <R>yijing\_d</R> <V>jieshu\_v le\_u <N><la\_v zha\_n  
 capital already end PERF pull barrier  
 xian\_v dian\_n de\_u lishi\_n</N></V> , \_w <V>xiwang\_v  
 limit electricity DE history hope  
 <Vo>yikao\_v <N>dajia\_r</N></Vo></V> , \_w <V>shi\_v <N><la\_v  
 count.on you make pull  
 zha\_n xian\_v dian\_n de\_u lishi\_n</N> <V><R>yongyuan\_d  
 barrier limit electricity DE history forever  
 buzai\_d</R> chongyan\_v</V></V> 。\_w</S>  
 never repeat  
 "Nowadays, a limited supply of electricity no longer occurs. However, I hope everybody can do their best to avoid a repetition of this event."

### 3.5.1.9 Prepositional phrase (P)

A prepositional phrase is made up of a preposition and its complement. The complement of a preposition is usually a noun phrase, but it can also be a verb phrase or a clause.

- (23) <S N="39"><N>我们\_r <N> <R>即将\_d</R> <P>以\_p <N>丰收\_vn 的\_u  
喜悦\_an</N></P> <V>送\_v 走\_v <N>牛年\_t</N></V> ,\_w <P>以\_p  
<N>昂扬\_a 的\_u 斗志\_n</N></P> <V>迎来\_v <N>虎年\_t</N></V> 。  
\_w</S>  
<S N="39"><N>women\_r <N> <R>jijiang\_d</R> <P>yi\_p  
we imminently with  
<N>fengshou\_vn de\_u xiyue\_an</N></P> <V>song\_v zou\_v  
fruitful DE joy farewell walk  
<N>Niunian\_t</N></V> ,\_w <P>yi\_p <N>angyang\_a de\_u  
Year.of.Ox in high DE  
douzhi\_n</N></P> <V>yinglai\_v <N>Hunian\_t</N></V> 。\_w</S>  
spirits welcome Year.of.Tiger  
“We are going to bid farewell to the Year of Ox joyfully and welcome the arrival  
of the Year of Tiger in high spirits.”

- (24) <S N="18"><N>这些\_r 外交\_n 活动\_vn</N> ,\_w <V>符合\_v <N><和  
平\_n 与\_c 发展\_v 的\_u> 时代\_n 主题\_n</N></V> ,\_w <V>顺应\_v  
<N><世界\_n 走向\_v 多极化\_v 的\_u> 趋势\_n</N></V> ,\_w <P>对  
于\_p <V>促进\_v <N><N&>国际\_n 社会\_n 的\_u 友好\_a 合作  
\_vn</N&> 和\_c <N+>共同\_b 发展\_vn</N+></N></V></P> <V>作出\_v  
了\_u <N>积极\_a 的\_u 贡献\_n</N></V> 。\_w</S>  
<S N="18"><N>zhexie\_r waijiao\_n huodong\_vn</N> ,\_w <V>fuhe\_v  
these diplomatic activities be.in.line.with  
<N><heping\_n yu\_c fazhan\_v de\_u> shidai\_n zhuti\_n</N></V> ,\_w  
peace and develop DE era theme  
<V>shunying\_v <N><shijie\_n zouxiang\_v duojihua\_v de\_u>  
follow peace walk.towards diversify DE  
qushi\_n</N></V> ,\_w <P>duiyu\_p <V>cujin\_v <N><N&>guoji\_n  
trend as.for step.up international  
shehui\_n de\_u youhao\_a hezuo\_vn</N&> he\_c <N+>gongtong\_b  
community DE friendly co-operation and shared

fazhan\_vn</N+></N></V></P> <V>zuochu\_v le\_u <N>jiji\_a de\_u  
 development make PERF constructive DE  
 gongxian\_n</N></V> °\_w</S>  
 contribution

“These diplomatic activities suit our goal of maintaining peace and development, meet the needs of global diversification, and contribute to enhancing our co-operation with other countries.”

- (25) <S N="109"><N>李\_nr 鹏\_nr</N> <P>对\_p <S><N>他们\_r</N> <Ja>倾  
 心\_ad</Ja> <V>照顾\_v <N>下一代\_n</N></V></S></P> <V>表示\_v  
 <N>肯定\_an</N></V> °\_w</S>  
 <S N="109"><N>Li\_nr Peng\_nr</N> <P>dui\_p <S><N>tamen\_r</N>  
 Li Peng as.for they  
 <Ja>qingxin\_ad</Ja> <V>zhaogu\_v <N>xiayidai\_n</N></V></S></P>  
 devotedly nurture next.generation  
 <V>biaoshi\_v <N>kending\_an</N></V> °\_w</S>  
 express praise  
 “Li Peng complimented them on their dedication in taking good care of their children.”

### 3.5.1.10 Adverbial prepositional phrase (Pa)

An adverbial prepositional phrase functions as an adjunct in a sentence. Semantically, it is typically used as an adverbial of time, reason, circumstances, place, etc.

- (26) <S N="37"><Pa>在\_p <N>这\_r 辞旧迎新\_l 的\_u 美好\_a 时刻  
 \_n</N></Pa> ,\_w <N>我\_r</N> <V>祝\_v <N>大家\_r</N>  
 <S><S&><N>新年\_t</N> <J>快乐\_a</J></S&> ,\_w <S+><N>家庭\_n</N>  
 <J>幸福\_a</J></S+></S></V> !\_w</S>  
 <S N="37"><Pa>zai\_p <N>zhe\_r cijiuyingxin\_l  
 on this farewell.the.past.and.welcome.the.future  
 de\_u meihao\_a shike\_n</N></Pa> ,\_w <N>wo\_r</N> <V>zhu\_v  
 DE joyful occasion I wish  
 <N>dajia\_r</N> <S><S&><N>xinnian\_t</N> <J>kuai\_a</J></S&> ,  
 both.of.us New.Year happy

\_w <S+><N>jiating\_n</N> <J>xingfu\_a</J></S+></S></V> !\_w</S>  
 family blessed

“On this joyful occasion of celebrating the New Year, I wish everyone a happy New Year and a blessed family.”

(27) <S N="15"><Pa>通过\_p <S><N>高层\_n</N> <V>互访  
 \_v</V></S></Pa> ,\_w <N>中国\_ns 与\_p 美国\_ns 、\_w 俄罗斯  
 \_ns 、\_w 法国\_ns 、\_w 日本\_ns 等\_u 大国\_n</N> <V>确定\_v 了  
 \_u <N><N&><双方\_n 关系\_n 未来\_t 发展\_v 的\_u 目标\_n</N&>  
 和\_c <N+>指导\_vn 方针\_n</N+></N></V> 。\_w</S>

<S N="15"><Pa>tongguo\_p <S><N>gaoceng\_n</N>  
 through senior.government.officials  
 <V>hufang\_v</V></S></Pa> ,\_w <N>Zhongguo\_ns yu\_p  
 visit.each.other China with  
 Meiguo\_ns 、\_w Eluosi\_ns 、\_w Faguo\_ns 、\_w Riben\_ns deng\_u  
 America Russia France Japan etc.  
 daguo\_n</N> <V>queding\_v le\_u <N><N&><shuangfang\_n guanxi\_n  
 giant.nations ensure PERF one.another relationship  
 weilai\_t fazhan\_v de\_u> bubiao\_n</N&> he\_c <N+>zhidao\_vn  
 future develop DE goal and guidance  
 fangzhen\_n</N+></N></V> 。\_w</S>  
 scheme

“Through frequent visits of Chinese officials to these countries, the United States, Russia, France and Japan have laid down definite goals and concrete plans to enhance co-operation with China.”

(28) <S N="45"><Pa>在\_p <N><<国际\_n 金融\_n 危机\_n 的\_u> 风浪\_n  
 波及\_v 许多\_m 国家\_n 的\_u 情况\_n 下\_f</N></Pa> ,\_w <N>我  
 国\_n</N> <V>保持\_v 了\_u <N><金融\_n 形势\_n 和\_c 整个\_m 经  
 济\_n 形势\_n 的\_u> 稳定\_a 发展\_vn</N></V> 。\_w</S>

<S N="45"><Pa>zai\_p <N><<guoji\_n jinyong\_n weiji\_n de\_u>  
 in international financial crisis DE  
 fenglang\_n boji\_v xuduo\_m guojia\_n de\_u qingkuang\_n  
 storm affect plenty.of countries DE circumstances  
 xia\_f</N></Pa> ,\_w <N>woguo\_n</N> <V>baochi\_v le\_u  
 under our.country maintain PERF

<N><jinyong\_n xingshi\_n he\_c zhengge\_m jingji\_n xingshi\_n de\_u>  
 financial situation and whole economic situation DE  
 wending\_a fazhan\_vn</N></V> °\_w</S>  
 stable development  
 “In the situation where many countries suffered from the global financial crisis,  
 China still maintained a steady development in both finance and economy.”

- (29) <S N="99"><Pa>在\_p <N>控制室\_n</N></Pa> ,\_w <N><N>&李\_nr 鹏  
 \_nr</N>& 与\_p <N+>职工\_n 们\_k</N+></N> <R>一一\_d</R> <V>握  
 手\_v</V> ,\_w <P>向\_p <N>大家\_r</N></P> <V>表示\_v <Vo>慰问  
 \_v</Vo></V> °\_w</S>  
 <S N="99"><Pa>zai\_p <N>kongzhishi\_n</N></Pa> ,\_w  
 in control.station  
 <N><N>&Li\_nr Peng\_nr</N>& yu\_p <N+>zhigong\_n  
 Li Peng with co-workers  
 men\_k</N+></N> <R>yiyi\_d</R> <V>woshou\_v</V> ,\_w <P>xiang\_p  
 PL one.by.one hold.hands to  
 <N>dajia\_r</N></P> <V>biaoshi\_v <Vo>weiwen\_v</Vo></V> °\_w</S>  
 them express compliments  
 “In the control station, Li Peng held hands with every member of staff and  
 expressed his compliments to them.”

### 3.5.1.11 Adverb phrase (R)

An adverb phrase consists of an adverb or a group of words in which an adverb is the head of the phrase.

- (30) <S N="16"><N><中国\_ns 与\_p 周边\_n 国家\_n 和\_c 广大\_b 发展中  
 国家\_l 的\_u> 友好\_a 合作\_vn</N> <R>进一步\_d</R> <V>加强  
 \_v</V> °\_w</S>  
 <S N="16"><N><Zhongguo\_ns yu\_p zhoubian\_n guojia\_n he\_c  
 China with neighbouring countries and  
 guangda\_b fazhanzhongguojia\_l de\_u youhao\_a  
 all.over.the.world developing.countries DE friendly  
 hezuo\_vn</N> <R>jinyibu\_d</R> <V>jiaqiang\_v</V> °\_w</S>  
 co-operation further strengthen  
 “The co-operation between China and her surrounding countries and developing  
 countries was further strengthened.”

- (31) <S N="35"><N>中国\_ns</N> <R>永远\_d</R> <V>是\_v <V>维护\_v  
 <N><世界\_n 和平\_n 与\_c 稳定\_an 的\_u> 重要\_a 力量  
 \_n</N></V></V> °\_w</S>  
 <S N="35"><N>Zhongguo\_ns</N> <R>yongyuan\_d</R> <V>shi\_v  
 China forever be  
 <V>weihu\_v <N><shijie\_n heping\_n yu\_c wending\_an de\_u>  
 safeguard world peace and stability DE  
 zhongyao\_a lilian\_g\_n</N></V></V> °\_w</S>  
 important power  
 "China has long been playing a vital role in maintaining world peace and  
 stability."
- (32) <S N="29">但是\_c ,\_w <N>世界\_n</N> <R>还\_d 不\_d</R> <J>安宁  
 \_a</J> °\_w</S>  
 <S N="29">danshi\_c ,\_w <N>shijie\_n</N> <R>hai\_d bu\_d</R>  
 however peace still not  
 <J>anning\_a</J> °\_w</S>  
 tranquil  
 "Yet, the world is not in peace."

### 3.5.1.12 Sentence (S)

A sentence is the maximal syntactic constituent into which a text is subdivided (Leech et al., 1996: §Sentence). By orthographic definition, in both English and Chinese, it is typically marked by a final full stop or some other terminal punctuation. For parsing purposes in this research, an item is labelled as a sentence if it (a) has an independent syntactic status; (b) is included in another sentence as in a transcript of directly quoted speech; (c) is embedded in a phrase, i.e. it is a clause-like unit with a subject and a verb.

- (33) <S N="8"><N>国民经济\_n</N> <V>保持\_v 了\_u <N><“\_w 高\_a 增  
 长\_vn 、\_w 低\_a 通胀\_j ”\_w 的\_u> 良好\_a 发展\_vn 态势  
 \_n</N></V> °\_w</S>  
 <S N="8"><N>guominjingji\_n</N> <V>baochi\_v le\_u <N><“\_w  
 national.economy maintain PERF

gao\_a zengzhang\_vn 、\_w di\_a tongzhang\_j ”\_w de\_u> lianghao\_a  
 high growth low inflation DE good  
 fazhan\_vn taishi\_n</N></V> 。\_w</S>  
 development trend  
 “Our country maintains a high growth rate and low inflation rate in the  
 economy.”

- (34) <S N="110"><N>他\_r</N> <V>说\_v :\_w <S> “\_w <N>人\_n</N> <J>  
 老\_a 了\_y</J> ,\_w <V>照顾\_v 照顾\_v <N>后代\_n</N></V> <R>也  
 \_d</R> <V>是\_v <N>一\_m 件\_q <可以\_v 带\_v 来\_v 快乐\_a 的  
 \_u> 事\_n</N></V> ,\_w <R>当然\_d</R> ,\_w <P>对\_p <N>孩子\_n  
 们\_k</N></P> <V>不能\_v 溺爱\_v</V> ,\_w <V>要\_v <Vo>让\_v  
 <S><N>他们\_r</N> <Ja>健康\_ad</Ja> <V>成长\_v</V></S></Vo></V> 。  
 \_w ”\_w</S></V></S>  
 <S N="110"><N>ta\_r</N> <V>shuo\_v :\_w <S>“\_w <N>ren\_n</N>  
 he said people  
 <J>lao\_a le\_y</J> ,\_w <V>zhaogu\_v zhaogu\_v  
 become.old PERF nurture nurture  
 <N>houdai\_n</N></V> <R>ye\_d</R> <V>shi\_v <N>yi\_m jian\_q  
 next.generation also be one CL  
 <keyi\_v dai\_v lai\_v kuaile\_a de\_u> shi\_n</N></V> ,\_w  
 can bring come happiness DE matter  
 <R>dangran\_d</R> ,\_w <P>dui\_p <N>haizi\_n men\_k</N></P>  
 of.course as.for children PL  
 <V>buneng\_v niai\_v</V> ,\_w <V>yao\_v <Vo>rang\_v  
 cannot spoil have.to let  
 <S><N>tamen\_r</N> <Ja>jiankang\_ad</Ja>  
 them healthily  
 <V>chengzhang\_v</V></S></Vo></V> 。\_w ”\_w</S></V></S>  
 grow

“He said, ‘when we grow older, we take delight in rearing our next generation.  
 But we must bear in mind that we cannot spoil them.’”

(35) <S N="105">同时\_c , \_w <R>也\_d</R> <V>希望\_v <S><N>你们\_r</N>  
 <Ja>安全\_ad</Ja> <V>生产\_v</V> 、\_w <R>经济\_ad</R> <V>调度  
 \_v</V> , \_w <V>实现\_v <N><经济\_n 增长\_vn 方式\_n 的\_u> 转变  
 \_vn</N></V></S></V> 。\_w ” \_w</S>  
 <S N="105">tongshi\_c , \_w <R>ye\_d</R> <V>xiwang\_v  
 in.the.meantime also hope  
 <S><N>nimen\_r</N> <Ja>anquan\_ad</Ja> <V>shengchan\_v</V> 、\_w  
 you safely produce  
 <R>jingji\_ad</R> <V>tiaodu\_v</V> , \_w <V>shixian\_v <N><jingji\_n  
 economically adjust realise economy  
 zengzhang\_vn fangshi\_n de\_u> zhuanbian\_vn</N></V></S></V> 。  
 growth mode DE change  
 \_w ” \_w</S>

“In the meantime, I hope you could take note of safety measures in manufacturing, improve productivity and keep up with any changes in our economy.”

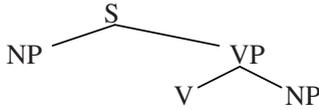
In this scheme, since embedded clauses in a phrase are marked by <S> elements, the “clause” category is not used except adverbial clause and correlative clause, both of which are formally distinguished from simple sentences. Adverbial clauses are generally identified by some sign of their dependence on the superordinate clause or sentence such as an introductory subordinating conjunction whereas simple sentences are not. As for correlative clauses, as Chao (1968: 121) points out, the correlative adverbs 越 *yue*... 越 *yue*... “the more..., the more...” make it impossible for the coordinated parts of correlative clauses to be independent sentences, and to this extent the sentence label would be misleading.

### 3.5.1.13 Verb phrase (V)

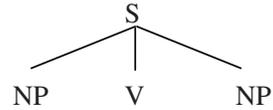
There is disagreement over the scope of the verb phrase, focussed upon whether or not it should include only its verb or its verb and complement (Leech et al., 1996: §Verb phrase; Kahrel et al., 1997: 240). Including the verb complement in the verb phrase evidently has an advantage as this will make the structure of the verb phrase conform to the basic assumption of most of the existing generative syntactic theories, including the theoretical frameworks I compared in section 3.4.2.1,

e.g. GB, LFG, HPSG, etc., in which binary branching of phrasal structures is the norm, as shown in (36a). Excluding the verb complement from the verb phrase will result in ternary branching nodes as in (36b).

(36a)



(36b)



In (36a) and (36b), branching nodes are of different types: in (36a) there are binary branching nodes such as S and VP, which dominate two elements, while in (36b) S is a ternary branching node, which dominates three constituents, in contrast to its counterpart in (36a). As Haegeman (1994: 138–144) notes, there are further advantages to adopting a grammar which allows only binary branching: such a grammar is more aesthetically satisfying, and is more constrained in the sense that lots of imaginable representations are ruled out. I therefore decided in my annotation scheme to include the complements of the verb in the verb phrase. In other words, a verb phrase in this parsing scheme includes the words that are verbs and verb complements. A merit of this approach is that the relative levels of the constituents can be shown more clearly, while adjuncts and peripheral adverbials (mostly adverb phrases in my treebank) are left at the sentence level.

- (37) <S N="11"><N><N&>民主\_a 法制\_n 建设\_vn</N&> 、\_w <N+>精神文明\_n 建设\_vn</N+> 和\_c <N+>其他\_r 各项\_r 事业\_n</N+></N> <R>都\_d</R> <V>有\_v <N>新\_a 的\_u 进展\_vn</N></V> °\_w</S>  
 <S N="11"><N><N&>minzhu\_a fazhi\_n jianshe\_vn</N&> 、\_w  
 democratic system establishment  
 <N+>jingshenwenming\_n jianshe\_vn</N+> he\_c <N+>qita\_r  
 moral.standards establishment and other  
 gexiang\_r shiye\_n</N+></N> <R>dou\_d</R> <V>you\_v  
 various.kinds.of undertaking all have  
 <N>xin\_a de\_u jinzhan\_vn</N></V> °\_w</S>  
 new DE progress  
 “We have made good progress in developing democracy and moral standards and other undertakings.”

### 3.5.1.14 Adverbial verb phrase (Va)

An adverbial verb phrase functions as a modifier to a verb or an adjunct to a sentence.

(38) <S N="70"><N>江\_nr 泽民\_nr 同志\_n</N> <Na>最近\_t</Na> <Va>强调\_vd</Va> <V>指出\_v</V> ,\_w <V>要\_v <R>大力\_d</R> <V>倡导\_v <N><说实话\_l 、\_w 办\_v 实事\_n 、\_w 鼓\_v 实劲\_n 、\_w 讲\_v 实效\_n 的\_u> 作风\_n</N></V></V> ,\_w <Ja>坚决\_ad</Ja> <V>制止\_v <N><追求\_v 表面文章\_l> ,\_w <搞\_v 花架子\_n> 等\_u 形式主义\_n</N></V> ,\_w <Ja>坚决\_ad</Ja> <V>杜绝\_v <N><脱离\_v 群众\_n> 、\_w <脱离\_v 实际\_n> 、\_w <浮躁\_a 虚夸\_v> 等\_u 官僚主义\_n</N></V> 。\_w</S>

<S N="70"><N>Jiang\_nr Zemin\_nr tongzhi\_n</N> <Na>zuijin\_t</Na>  
 Jiang Zemin comrade recently  
 <Va>qiangdiao\_vd</Va> <V>zhichu\_v</V> ,\_w <V>yao\_v  
 strongly point.out have.to  
 <R>dali\_d</R> <V>changdao\_v <N><shuoshihua\_l 、\_w ban\_v  
 highly encourage tell.the.truth do  
 shishi\_n 、\_w gu\_v shijing\_n 、\_w jiang\_v shixiao\_n de\_u>  
 useful.work gather energy focus.on effectiveness DE  
 zuofeng\_n</N></V></V> ,\_w <Ja>jianjue\_ad</Ja> <V>zhizhi\_v  
 attitude persistently prohibit  
 <N><zuiqui\_v biaomianwenzhang\_l> ,\_w <gao\_v huajiazi\_n> deng\_u  
 follow superficiality make flower.rack etc.  
 xingshizhuyi\_n</N></V> ,\_w <Ja>jianjue\_ad</Ja> <V>dujue\_v  
 formalities persistently eliminate  
 <N><tuoli\_v qunzhong\_n> 、\_w <tuoli\_v  
 be.out.of.tune.with general.public be.out.of.tune.with  
 shiji\_n> 、\_w <fuzao\_a xukua\_v> deng\_u  
 pragmatism impatient exaggerated etc.  
 guanliaozhuyi\_n</N></V> 。\_w</S>

red.tape

“Jiang Zemin has recently emphasised that we should encourage an effective and efficient working atmosphere and get rid of unnecessary formalities and red tape.”

(39) <S N="89"><Va>演出\_v 结束\_v 后\_f</Va> ,\_w <N><N&>江\_nr 泽民\_nr 等\_u 党\_n</N&> 和\_c <N+>国家\_n 领导人\_n</N+></N> <V>走\_v 上\_v <N>舞台\_n</N></V> ,\_w <Ja>亲切\_ad</Ja> <V>会见\_v 了\_u <N><参加\_v 演出\_v 的\_u> 全体\_n 人员\_n</N></V> ,\_w <V><V&>祝贺\_v <Vo>演出\_v 成功\_a</Vo></V&> ,\_w 并\_c <V+><P>与\_p <N>他们\_r</N></P> 合影\_v <Vo>留念\_v</Vo></V+></V> 。\_w</S>

<S N="89"><Va>yanchu\_v jieshu\_v hou\_f</Va> ,\_w  
 performance finish after  
 <N><N&>Jiang\_nr Zemin\_nr deng\_u dang\_n</N&> he\_c  
 Jiang Zemin etc. the.Party and  
 <N+>guojia\_n lingdaoren\_n</N+></N> <V>zou\_v shang\_v  
 country leader step on  
 <N>wutai\_n</N></V> ,\_w <Ja>qinqie\_ad</Ja> <V>huijian\_v le\_u  
 stage sincerely meet PERF  
 <N><canjia\_v yanchu\_v de\_u> quanti\_n  
 participate.in performance DE the.whole.group.of  
 renyuan\_n</N></V> ,\_w <V><V&>:huhe\_v <Vo>yanchu\_v  
 performers compliment performance  
 chenggong\_a</Vo></V&> ,\_w bing\_c <V+><P>yu\_p  
 successful and with  
 <N>tamen\_r</N></P> heyingshu\_v <Vo>liunian\_v</Vo></V+></V> 。  
 them take.photo as.a.memento  
 \_w</S>

“After the performance finished, Jiang Zemin’s Communist Party and other leaders stepped onto the stage, met the performers and complimented them on their success and took photographs with them.”

### 3.5.1.15 Verbal object (Vo)

This parsing label is specific to Chinese. As Norman (1988: 163) notes, “Chinese, like a number of other East Asian languages, is an aspect and not a tense language”. This means that unlike English, Chinese verbs in themselves lack any distinction of finiteness and nonfiniteness (Rai-Chu, 1982: 78ff.). In cases where a main verb takes another verb as direct object, English and Chinese verbs behave differently. While in



(42) <S N="104"><Na>现在\_t</Na> ,\_w <N>我们\_r 的\_u 首都\_n</N> <R>已经\_d</R> <V>结束\_v 了\_u <N><拉\_v 闸\_n 限\_v 电\_n 的\_u>历史\_n</N></V> ,\_w <V>希望\_v <Vo>依靠\_v <N>大家\_r</N></Vo></V> ,\_w <V>使\_v <N><拉\_v 闸\_n 限\_v 电\_n 的\_u>历史\_n</N> <V><R>永远\_d 不再\_d</R> 重演\_v</V></V> 。\_w </S>  
 <S N="104"><Na>xianzai\_t</Na> ,\_w <N>women\_r de\_u  
 at.present our GEN  
 shoudu\_n</N> <R>yijing\_d</R> <V>jieshu\_v le\_u <N><la\_v zha\_n  
 capital.city already end PERF pull barrier  
 xian\_v dian\_n de\_u> lishi\_n</N></V> ,\_w <V>xiwang\_v  
 limit electricity DE history hope  
 <Vo>yikao\_v <N>dajia\_r</N></Vo></V> ,\_w <V>shi\_v <N><la\_v  
 count.on you make pull  
 zha\_n xian\_v dian\_n de\_u> lishi\_n</N> <V><R>yongyuan\_d  
 barrier limit electricity DE history forever  
 buzai\_d</R> chongyan\_v</V></V> 。\_w </S>  
 never repeat  
 “Nowadays, a limited supply of electricity no longer occurs. However, I hope everybody can do their best to avoid a repetition of this event.”

However, verb phrases with a verbal object are distinct from verb phrases with an auxiliary and a main verb. The main verb is not taken to be the verbal object of the auxiliary verb and thus no <Vo> element is used as in (43) and (44) where the auxiliaries are highlighted in grey colour, as opposed to the examples of verb phrases having a verbal object as shown in (41) and (42) above.

(43) <S N="12"><N>我们\_r</N> <R>十分\_m<sup>18</sup></R> <V>关注\_v <N><最近\_t 一个\_m 时期\_n 一些\_m 国家\_n 和\_c 地区\_n 发生\_v 的\_u>金融\_n 风波\_n</N></V> ,\_w <N>我们\_r</N> <V>相信\_v <S><Pa>通过\_p <N><N&><这些\_r 国家\_n 和\_c 地区\_n 的\_u> 努力\_an</N&> 以及\_c <N+>有关\_v 的\_u 国际\_n 合作\_vn</N+></N></Pa> ,\_w <N>情况\_n</N> <V>会\_v <R>逐步\_d</R> 得到\_v <N>缓解\_vn</N></V></S></V> 。\_w </S>

18 The tag “m” for numeral is erroneously assigned in the PFR Chinese Corpus; it should be “d”, an adverb.





### 3.5.1.16 Initial (&) and non-initial conjunct (+)

Any of the above parsing tags may be used with “&” and “+” to indicate coordination. The first coordinate element will carry “&” sign. Subsequent ones will have “+”, placed after both medial punctuations and coordinating conjunctions (if any).

- (46) <S N="13">总的来说\_c ,\_w <N><N&>中国\_ns 改革\_v<N&> 和\_c  
 <N+>发展\_v 的\_u 全局\_n<N+></N> <V>继续\_v <Vo>保持\_v 了\_u  
 <N>稳定\_an<N></Vo></V> 。\_w</S>  
 <S N="13">zongdelaishuo\_c ,\_w <N><N&>Zhongguo\_ns  
 on.the.whole China  
 gaige\_v<N&> he\_c <N+>fazhan\_v de\_u quanju\_n<N+></N>  
 reform and development DE situation  
 <V>jixu\_v <Vo>baochi\_v le\_u <N>wending\_an<N></Vo></V> 。  
 continue maintain PERF stability  
 \_w</S>

“On the whole, China’s reform and development keep on progressing steadily.”

- (47) <S N="67"><N>这\_r<N> <V>是\_v <P>把\_p <N>党\_n 的\_u 方针\_n 、  
 \_w 政策\_n<N></P> <落到实处\_l></V> ,\_w <V>使\_v <Vo><Vo&>改  
 革\_v</Vo&> 和\_c <Vo+>建设\_v</Vo+></Vo> <Vo>取得\_v <N>胜利  
 \_vn 的\_u 重要\_a 保证\_vn<N></Vo></V> 。\_w</S>  
 <S N="67"><N>zhe\_r<N> <V>shi\_v <P>ba\_p <N>dang\_n de\_u  
 this be BA the.Party’s GEN  
 fangzhen\_n 、\_w zhengce\_n<N></P> <luodaoshichu\_l></V> ,\_w  
 objective policy realise  
 <V>shi\_v <Vo><Vo&>gaige\_v</Vo&> he\_c  
 make reform and  
 <Vo+>jianshe\_v</Vo+></Vo> <Vo>qude\_v <N>shengli\_vn de\_u  
 construction obtain success DE  
 zhongyao\_a baozheng\_vn<N></Vo></V> 。\_w</S>  
 important guarantee

“This is to safeguard the implementation of the Communist Party’s policy and success in the country’s reform and development.”

(48) <S N="5"><S&><N>[中国\_ns 政府\_n]nt</N> <Ja>顺利\_ad</Ja> <V>恢复  
 \_v <P>对\_p <N>香港\_ns</N></P> <Vo>行使\_v <N>主权  
 \_n</N></Vo></V></S&> , \_w 并\_c <S+><P>按照\_p <N>< “\_w 一国  
 两制\_j ”\_w 、\_w “\_w 港人治港\_l ”\_w 、\_w 高度\_d 自治\_v 的  
 \_u> 方针\_n</N></P> <V>保持\_v <N>香港\_ns 的\_u 繁荣\_an 稳定  
 \_an</N></V></S+> 。\_w</S>  
 <S N="5"><S&><N>[Zhongguo\_ns Zhengfu\_n]nt</N> <Ja>shunli\_ad</Ja>  
 China government successfully  
 <V>huiifu\_v <P>dui\_p <N>Xianggang\_ns</N></P> <Vo>xingshi\_v  
 resume to Hong.Kong exercise  
 <N>zhuquan\_n</N></Vo></V></S&> , \_w bing\_c <S+><P>anzhao\_p  
 sovereignty and according.to  
 <N><“\_w yiguoliangzhi\_j ”\_w 、\_w “\_w gangrenzhigang\_l ”\_w 、  
 one.country.two.systems people’s.ruling.themselves  
 \_w gaodu\_d zizhi\_v de\_u fangzhen\_n</N></P> <V>baochi\_v  
 highly autonomous DE objective maintain  
 <N>Xianggang\_ns de\_u fanrong\_an wending\_an</N></V></S+> 。  
 Hong.Kong DE prosperity stability  
 \_w</S>

“The Chinese government has succeeded in resuming its sovereignty over Hong Kong and maintaining the prosperity and stability of Hong Kong, according to the objectives of ‘one country two systems’, ‘Hong Kong people ruling themselves’ and ‘high degree of autonomy’.”

### 3.5.2 Issues arising from the application of the scheme

Even if a basic parsing scheme for annotation is made explicit, many issues may arise when the scheme is applied to corpus data (Sampson, 1987: 94ff.). Throughout the entire parsing process, choices had to be made as such issues arose. These choices will be documented and explained in this section. The decisions were then adhered to consistently in the annotation of similar cases thereafter. In the following subsections, I will discuss the issues that arose and illustrate them with examples drawn from the sample treebank.

### 3.5.2.1 Underspecification – Use of unlabelled bracketings

Brackets may be left unlabelled in cases where a particular grouped sequence of words cannot fit in to any of the existing phrase or clause categories. Examples of constituents enclosed in unlabelled brackets are given below from (a) to (e).

(a) Multi-word premodifiers of noun phrases marked by the particle 的 *de* (see also section 3.6.2.3):

e.g. <N><全国\_n 各族\_r 的\_u> 人民\_n</N> <N><quanguo\_n gezu\_r de\_u renmin\_n</N> “people from different ethnic groups throughout the country”;

e.g. <N><同\_p 联合国\_nt 和\_c 其他\_r 国际\_n 组织\_n 的\_u> 协调\_vn</N> <N><tong\_p Lianheguo\_nt he\_c qita\_r guoji\_n zuzhi\_n de\_u xietiao\_vn</N> “the compromise between the United Nations and other international organisations”;

e.g. <N><最近\_t 一个\_m 时期\_n 一些\_m 国家\_n 和\_c 地区\_n 发生\_v 的\_u> 金融\_n 风波\_n</N> <N><zuijin\_t yige\_m shiqi\_n yixie\_m guojia\_n he\_c diqu\_n fasheng\_v de\_u jinyong\_n fengbo\_n</N> “the recent financial crises happened in some countries and districts”;

e.g. <N+><<V><V>>保护\_v 环境\_n</V&> 、\_w <V+>消除\_v 贫困\_an</V+> 、\_w <V+>打击\_v 国际\_n 犯罪\_vn</V+></V> 等\_u 方面\_n 的\_u> 国际\_n 合作\_vn</N+> <N+><<V><V&>baohu\_v huanjing\_n</V&> 、\_w <V+>xiaochu\_v penkun\_an</V+> 、\_w <V+>daji\_v guoji\_n fazui\_vn</V+></V> deng\_u fangmian\_n de\_u guoji\_n hezuo\_vn</N+> “worldwide co-operation on such matters as environmental protection, the elimination of poverty and a crackdown on international crime”.

As stated in section 3.4.1, those grammatical constructions marked by the particle *de* as in (a) above are highly controversial: some scholars refer to them as relative clauses (e.g. Li and Thompson, 1989: 579 ff.; Aoun and Li, 1993; Chiu, 1993; Wu, 2000; Xue et al., 2000) or appositive clauses (Chu and Chi, 1999: 226), while others do not agree with either approach (e.g. Chao, 1968; Zhu, 1982 and 2000; Liu, 2003). Moreover, using some catch-all label, such as “*de* constructions”, does not help

either because the particle *de* in itself is vague in its functions: it can be a genitive marker, a marker of nominalisation and an adjectival marker (see section 3.4.1). It is therefore not easy to agree upon how the *de* constructions are defined and instantiated in texts. Having considered that there is to date no consensus on this issue of how *de* constructions should be analysed, and any invented label that attempts to refer to them will act as a locus of controversy and disagreement, I decided not to set up a new label for them in my parsing scheme and thus these constructions were enclosed by unlabelled brackets in the treebank.

(b) Serial verb constructions which are used as if they were compound verbs (see also section 3.6.2.4):

- e.g. <坚持\_v 奉行\_v> <jianchi\_v fengxing\_v> “insist on following”;
- e.g. <指挥\_v 演奏\_v> <zhihui\_v yanzou\_v> “lead and perform”;
- e.g. <看望\_v 慰问\_v> <kanwang\_v weiwen\_v> “visit and send regards to ...”.

As will be explained in more detail in section 3.6.2.4, some serial verb constructions are problematic in that their status of being serial verbs is dubious. Given no in-depth investigations of whether or not they are actually serial verbs or compound verbs that have been carried out as yet, these constructions were left unlabelled in my treebank.

(c) Serial adjective constructions:

- e.g. <团结\_a 一致\_a> <tuanjie\_a yizhi\_a> “be united together”;
- e.g. <圆满\_a 成功\_a> <yuanman\_a chenggong\_a> “perfectly successful”.

For the same reason as in (b), since it is not clear whether these serial adjective constructions are compound adjectives, I left them unlabelled.

(d) Idioms/set phrases which are used idiosyncratically as if they were single-word nouns or verbs (see also section 3.6.2.2):

- e.g. <大势所趋\_i ,\_w 民心所向\_i> <dashisuoqu\_i ,\_w minxinsuoxiang\_i> “urged by the trend, supported by general public”;
- e.g. <大气磅礴\_i ,\_w 波澜壮阔\_i> <diqibangbo\_i ,\_w bolanzhuangkuo\_i> “powerful wind, fierce waves”;
- e.g. <流光溢彩\_l ,\_w 火树银花\_i> <liuguangyicai\_l ,\_w huoshuyinhua\_i> “filled with colourful lights, magnificent”.

In cases like (d) above, idioms or set phrases can function as predicate in a sentence. However, as will be discussed in section 3.6.2.2, even if context is taken into consideration, it is not obvious whether these idiomatic expressions function as a nominal predicate or a verbal predicate, both of which are allowed in the Chinese syntax (Chao, 1968: 90). They were thus not enclosed in the <V> element nor <N> element in the treebank and left unlabelled instead.

(e) Coordinated verbs with shared direct object:

- e.g. <V><学习\_v 和\_c 掌握\_v> <N>党\_n 的\_u 十大\_j 精神\_n</N></V> <V><xuexi\_v he\_c zhangwo\_v> <N>dang\_n de\_u shiwuda\_j jingshen\_n</N></V> “learn and master the Communist Party’s 15 principles”;
- e.g. <V><尊重\_v 、\_w 认识\_v 和\_c 掌握\_v> <N>客观\_a 规律\_n</N></V> <V><zhuanzhong\_v 、\_w renshi\_v he\_c zhangwo\_v> <N>keguan\_a guilü\_n</N></V> “respect, understand and master what we learn in our daily life”.

In the examples given in (e) above, two or more transitive verbs in coordination share the same direct object. The coordinated verbs (except the last one) are not constituent-like in the sense that they do not constitute a complete verb phrase structure because the following shared object does not come immediately after them. I did not therefore apply the usual practice of marking conjuncts by enclosing them in the <V&> and <V+> elements, which are only used for coordinated verbs or verb phrases with complete verb phrase structure (Eyes and Leech, 1993: 53; Sampson, 1995: 310f.; Leech et al., 1996: §Coordination). Since it is also not worthwhile to set up a new parsing label to mark such a non-frequent phenomenon, I decided to put these verbal segments into unlabelled brackets.

### 3.5.2.2 Bracketing of multi-word constituents

The unlabelled bracketing facility evidently has its uses in skeleton parsing as it allows analysis to proceed where labelling decisions are not obvious or straightforward. Nevertheless, for some multi-word adverb phrases containing two adverbs (e.g. <R>还\_d 不\_d</R> <R>hai\_d

*bu\_d*</R> “not...though”; <R>永远\_d 不再\_d</R> <R>yongyuan\_d buzai\_d</R> “never forever”; <R>一直\_d 都\_d</R> <R>yizhi\_d dou\_d</R> “constantly”), and multi-word attributive adjectival phrases containing an adjective premodified by at least one adverb (e.g. <J>非常\_d 重要\_a 的\_u</J> <J>feichang\_d zhongyao\_a de\_u</J> “very important”; <J>很\_d 不\_d 平凡\_a 的\_u</J> <J>hen\_d bu\_d pingfan\_a de\_u</J> “very extraordinary; <J>十分\_m 高兴\_a</J> <J>shifen\_m gaoxing\_a</J> “very happy”), though Eyes and Leech (1993: 53) chose to put them into unlabelled brackets, they were labelled in my treebank. The reason for this is that their internal structure is clear, having a head (adjective or adverb) being modified by another adverb.

### 3.5.2.3 Bracketing of single-word constituents

As suggested in the EAGLES Recommendations for the Syntactic Annotation of Corpora, Version of 11<sup>th</sup> March 1996 (Leech et al., 1996), it is considered preferable to bracket single-word constituents where they show their phrasal status by the possibility of adding modifiers or replacing them by a multi-word phrase as in (49), or where they are in coordination with other multi-word constituents as in (50).

- (49) <N>人民\_n 生活\_vn</N> <R>进一步\_d</R> <V>改善\_v</V>  
 <N>renmin\_n shenghuo\_vn</N> <R>jinyibu\_d</R> <V>gaishan\_v</V>  
 citizens life further improve  
 “the life of the citizens is further improved”

- (50) <N><N&>全党\_n</N&> 和\_c <N+>全国\_n 各族\_r 人民\_n</N+></N>  
 <N><N&>quandang\_n</N&> he\_c <N+>quanguo\_n  
 the.whole.Communist.Party and the.whole.country  
 gezu\_r renmin\_n</N+></N>  
 various.kinds.of.ethnic.groups citizens  
 “the Communist Party and the citizens of varied ethnic groups throughout the country”

In example (49), the phrase brackets delimiting the adverb 进一步 *jinyibu* “further” and the main verb 改善 *gaishan* “improve” are meant to indicate the potential phrasal status of these words, because of the

possibility of adding another adverb (e.g. 再 *zai* “again”) before the former or replacing the latter with a multi-word verb phrase (e.g. 得到改善 *dedao gaishan* “have got (it) improved”). In example (50), the equivalence of the two conjuncts can be represented in the bracketing, though the initial conjunct is a noun and the final conjunct is a noun phrase.

### 3.5.2.4 Punctuation

Generally speaking, I included punctuation within the bracketing. As for phrase/sentence-initial and phrase/sentence-final punctuations, I enclosed them within the parsing bracketing, as in (51):

- (51) <P>为\_p <N> “\_w 两手抓\_l 、\_w 两手\_m 都\_d 要\_v 硬\_a ”  
 \_w<N><P> <V>提供\_v 了\_u <N>新\_a 的\_u 理论\_n 根据  
 \_n<N><V>  
 <P>wei\_p <N>“\_w liangshouzhua\_l 、\_w liangshou\_m dou\_d yao\_v  
 for both.hands.hold.sth both.hands all have.to  
 ying\_a ”\_w<N><P> <V>tigong\_v le\_u <N>xin\_a de\_u lilun\_n  
 stiff provide PERF new DE theoretical  
 genju\_n<N><V>  
 grounds  
 “provide new theoretical evidence for the principle of perseverance in an  
 undertaking”

As regards medial punctuation marks, typically commas, I attached them to the highest available node in the parse tree, thus these punctuation marks can be used as delimiters of major constituents, as in (52):

- (52) <S N="5"><S&><N>[中国\_ns 政府\_n]nt<N> <Ja>顺利\_ad</Ja> <V>恢复  
 \_v <P>对\_p <N>香港\_ns<N><P> <Vo>行使\_v <N>主权  
 \_n<N><Vo><V></S&> ,\_w 并\_c <S+><P>按照\_p <N>< “\_w 一国  
 两制\_j ”\_w 、\_w “\_w 港人治港\_l ”\_w 、\_w 高度\_d 自治\_v 的  
 \_u 方针\_n<N><P> <V>保持\_v <N>香港\_ns 的\_u 繁荣\_an 稳定  
 \_an<N><V></S+> 。\_w<S>  
 <S N="5"><S&><N>[Zhongguo\_ns zhengfu\_n]nt<N> <Ja>shunli\_ad</Ja>  
 China government successfully

<V>huifu\_v <P>dui\_p <N>Xianggang\_ns</N></P> <Vo>xingshi\_v  
 resume to Hong.Kong exercise  
 <N>zhuquan\_n</N></Vo></V></S&> ’\_w bing\_c <S+><P>anzhao\_p  
 sovereignty and according.to  
 <N><“\_w yiguoliangzhi\_j ”\_w \\_w “\_w gangrenzhigang\_l ”\_w \\_w  
 one.country.two.systems people’s.ruling.themselves  
 \_w gaodu\_d zizhi\_v de\_u> fangzhen\_n</N></P> <V>baochi\_v  
 highly autonomous DE objective maintain  
 <N>Xianggang\_ns de\_u fanrong\_an wending\_an</N></V></S+> °  
 Hong.Kong DE prosperity stability  
 \_w</S>  
 “The Chinese government has succeeded in resuming its sovereignty over Hong  
 Kong and maintaining the prosperity and stability of Hong Kong, according to  
 the objectives on ‘one country two systems’, ‘Hong Kong people ruling  
 themselves’ and ‘high degree of autonomy’.”

### 3.5.2.5 Ambiguity

Linguistic forms are often ambiguous. My annotation scheme, however, did not contain any notation for representing ambiguity explicitly with which the human analyst selects one possible sense for a form and represents it. I decided not to explicitly mark an ambiguous form because even if a given item has more than one reading, the human analyst will not recognise this in the course of parsing and just annotate the item with the interpretation that seems initially most plausible. In fact, similar problems were encountered in the production of the Penn Chinese Treebank and the annotators of the treebank did not annotate ambiguities either (Xue et al., 2000: 173–178). They believed that in each case one of these ambiguous readings was unlikely and thus they annotated assuming the more plausible reading. In this regard, my treebank may appear unsatisfactory in connection with research on different kinds of ambiguity.

With such a detailed and carefully articulated parsing scheme and guidelines, I can now proceed to describe the actual process of skeleton parsing and difficulties encountered in the process of parsing.

## 3.6 The process of skeleton parsing

### 3.6.1 The basic concept of skeleton parsing

The basic idea of skeleton parsing, as Garside and McEnery (1993: 19) demonstrate, is that the treebanker marks only those syntactic structures which seem “intuitively obvious”, rather than keeping track of a particular reference grammar. In the course of skeleton parsing, I inserted a nested set of brackets around a sequence of word tokens which appeared to be intuitively correct to group as a single unit. I then assigned to each of these units (i.e. sentence constituents) a label from the set of categories specified in my parsing scheme. An excerpt of the PFR Skeleton Treebank is provided in the following figure.

```
<ORIGINALMP>19980101-01-001-013_m</ORIGINALMP> <S N="28"><V>环顾_v <N>全球_n</N></V> ,_w <N><日益_d 密切_a 的_u> 世界_n 经济_n 联系_vn</N> ,_w <N>日新月异_j 的_u 科技_n 进步_vn</N> ,_w <R>正在_d</R> <P>为_p <N><各国_r 经济_n 的_u> 发展_vn</N></P> <V>提供_v <N>历史_n 机遇_n</N></V> 。_w</S> <S N="29">但是_c ,_w <N>世界_n</N> <R>还_d 不_d</R> <I>安宁_a</I> 。_w</S> <S N="30"><N><南北_n 之间_f 的_u> 贫富_n 差距_n</N> <V>继续_v <Vo>扩大_v</Vo></V> ;_w <N>局部_b 冲突_vn</N> <I>时有发生_j> ;_w <N><不_d 公正_a 不_d 合理_a 的_u> 旧_a 的_u 国际_n 政治_n 经济_n 秩序_n</N> <R>还_d</R> <V>没有_v <N>根本_a 改变_vn</N></V> ;_w <N>发展中国家_l</N> <P>在_p <N>激烈_a 的_u 国际_n 经济_n 竞争_vn 中_f</N></P> <R>仍_d</R> <V>处于_v <N>弱势_n 地位_n</N></V> ;_w <N>人类_n 的_u <N><N>&生存_vn</N>& 与_c <N>发展_vn</N><N></N> <R>还_d</R> <V>面临_v <N>种种_q <N><N>&威胁_vn</N>& 和_c <N>挑战_vn</N><N></N></V> 。_w</S> <S N="31"><N><和平_n 与_c 发展_vn 的_u> 前景_n</N> <V>是_v <I>光明_a 的_u</I></V> ,_w <N>2 1_m 世纪_n</N> <R>将_d</R> <V>是_v <N><充满_v 希望_n 的_u> 世纪_n</N></V> 。_w</S> <S N="32">但_c <N>前进_v 的_u 道路_n</N> <V><V><V>&不会_v</V>& <R>也_d</R> <V><R>不_d</R> 可能_v</V></V> <I>一帆风顺_j</I></V> ,_w <N>关键_n</N> <V>是_v <S><N>世界_n 各国_r 人民_n</N> <V>要_v <I>进一步_d 团结_a 起来_v</I></V></S></V> ,_w <V><V><V>&共同_d</R> 推动_v</V>& <V><R>早日_d</R> 建立_v</V></V> <N><公正_a 合理_a 的_u> 国际_n 政治_n 经济_n 新_a 秩序_n</N></V> 。_w</S>
```

Figure 4: An excerpt of the PFR Sample Skeleton Treebank.

### 3.6.2 Difficulties in skeleton parsing Chinese text

It is noteworthy here to discuss the major difficulties that I encountered in the course of skeleton parsing a sample text taken from my corpus, as this illuminates some of the peculiarities of the Chinese language.

### 3.6.2.1 *Ba* constructions

Firstly, 把 *ba* constructions make the parse of a verb phrase incomplete. The *ba* construction is a widely discussed topic in the grammar of Chinese (see, for instance, Li and Thompson, 1989: 463–491; Chen, 1990; Kit, 1992; Zou, 1993; Xia and Wu, 1996; Li, 1997; Xue et al., 2000; Li, 2001). In general, the structure of the *ba* construction is expressed and underlined in (53): a *ba* sentence has a subject, followed by *ba* and the *ba* noun phrase (i.e. the NP directly following *ba*) followed by a verb.

(53) subject                    *ba* NP                    verb

In my PFR treebank, 104 instances of the *ba* construction were found, which fall into three types of this construction. The general pattern of *ba* sentences is to place the direct object of the following verb immediately after *ba* as in example (54).

(54) <P>把\_p    <N\_i>电厂\_n</N\_i></P>    <V>建\_v    好\_a</V> Ø<sub>i</sub>  
       <P>*ba*\_p    <N\_i>dianchang\_n</N\_i></P>    <V>jian\_v    hao\_a</V> Ø<sub>i</sub>  
           BA                    electricity.supply.station            build    well  
       “build an electricity supply station”

Typically, a transitive verb should follow the pattern “<V>...<N>...</N></V>”. However, the verb<sup>20</sup>, 建\_v 好\_a, *jian-hao*, “build well, build in good shape” that follows the *ba* construction lacks a direct object, which is equivalent to the *ba* noun phrase, i.e. *ba* NP<sub>i</sub> verb Ø<sub>i</sub>, where Ø<sub>i</sub> denotes the empty position of the preposed verb object that shares the same reference as the prepositional complement of *ba*. In other words, the structure of the verb phrase following the *ba* construction is in the form of <V>...</V> rather than the canonical form mentioned before.

More complicated *ba* constructions involve the occurrence of two following verbs and a passivised verb. In cases like (55) where two

20 建\_v 好\_a, *jian-hao* “build well, build in good shape” is a compound verb. More specifically, it is a verb-complement (V-R) compound (Chao, 1968: 435 ff.). The resultative complement 好 *hao* “good” is bound to and follows the verb 建 *jian* “build” and expresses the result of the action of the verb.

different transitive verbs follow the *ba* construction, it is not immediately obvious whether that the prepositional complement of *ba* co-refers to the object of the first verb (改编 *gaibian* “be adapted for”) or that of the second one (为 *wei* “be changed as”). Since the second verb already takes a direct object (器乐曲 *qiyuequ* “acoustics of musical instruments”), the *ba* noun phrase must be co-referential with the object of the first verb.

- (55) <P>把\_p <N\_i><广大\_b 听众\_n 耳熟能详\_i 的\_u> 歌曲\_n</N\_i></P>  
 <V>改编\_v Ø\_i <V>为\_v <N>器乐曲\_n</N></V></V>  
 <P>*ba\_p* <N\_i><*guangda\_b tingzhong\_n ershunengxiang\_i de\_u*>  
 BA general audience familiar DE  
*gequ\_n*</N\_i></P> <V>*gaibian\_v* Ø\_i <V>*wei\_v*  
 songs change to.be  
 <N>*qiyuequ\_n*</N></V></V>  
 acoustics.of.musical.instruments  
 “change those popular songs into acoustic versions”

In less obvious cases like (56), however, it is impossible to locate any empty position that co-refers to the *ba* complement. The verb 带入 *dairu* “bring to” that follows the *ba* construction is used causatively without any visible passivisation. As Norman (1988: 164) notes, Chinese verbs do not make any distinction between the active (or unaccusative) and passive (or causative). The *ba* noun phrase 人们 *renmen* “people” therefore actually refers to the logical subject of the verb.

- (56) <P>把\_p <N>人们\_n</N></P> <V>带入\_v <N>迷人\_a 的\_u 艺术\_n  
 境地\_n</N></V>  
 <P>*ba\_p* <N>*renmen\_n*</N></P> <V>*dairu\_v* <N>*miren\_a de\_u*  
 BA people bring.into fascinating DE  
*yishu\_n jingdi\_n*</N></V>  
 imaginary world  
 “bring people into a fascinating imaginary world”

### 3.6.2.2 Idioms or set phrases

The use of idioms (tagged “i”) or set phrases (tagged “I”) as if they were nouns and verbs is also problematic. Noun-like idioms and set phrases are illustrated in example (57) and verb-like set phrases in example (58). To my knowledge, the grammatical categories of this kind of idiomatic expressions have not been documented so far.

- (57) <N>今晚\_t 的\_u 长安街\_ns</N> <流光溢彩\_l ,\_w 火树银花\_i>  
<N>jinwan\_t de\_u Changanjie\_ns</N> <liuguangyicai\_l ,\_w  
tonight DE Changan.Street filled.with.colourful.lights  
huoshuyinhua\_i>  
bright.red.trees.with.silver.flowers  
“Tonight the Changan Street was filled with colourful lights and really looked magnificent.”

- (58) <N>国民经济\_n</N> <稳中求进\_l>  
<N>guominjingji\_n</N> <wenzhongqiujin\_l>  
national.economy steadily.progress  
“The national economy is progressing steadily.”

That they can be used rather idiosyncratically as a noun or a verb makes it almost impossible for even a human analyst to determine the phrasal category of a given idiomatic expression: whether it is a noun phrase or a verb phrase. As in the above two examples, it is unclear whether the idiom/set phrase placed after the subject noun phrase is intended to function as a nominal expression or a verbal one. Unlike English, in which the subject must be followed by a verbal predicate, a Chinese predicate can be a verbal predicate, an adjectival predicate or a nominal predicate (Chao, 1968: 90). In the absence of further evidence of the categorial status of such segments, those idioms and set phrases occurring in the predicate position were left unlabelled in my treebank.

### 3.6.2.3 Lengthy premodifiers of a noun phrase

Unlike English, which favours the use of postmodification if a modifier of a noun phrase is long (Quirk and Greenbaum, 1973: 425; de Haan, 1991), Chinese prefers premodification to postmodification, re-

ardless of the length of the modifier (cf. Liu et al., 1996: 265–274). It is thus common in the PFR treebank that a noun is qualified by a grammatical unit of over six words which is marked by the particle 的 *de* at the end, as in example (59). As stated in section 3.4.1, the particle 的 *de* is traditionally treated as a marker of modification.

- (59) <N><中国\_ns 与\_p 周边\_n 国家\_n 和\_c 广大\_b 发展中国家\_l 的\_u> 友好\_a 合作\_vn</N>  
 <N><Zhongguo\_ns yu\_p zhoubian\_n guojia\_n he\_c guangda\_b  
 China with neighbouring countries and all.over.the.world  
 fazhanzhongguojia\_l de\_u> youhao\_a hezuo\_vn</N>  
 developing.countries DE friendly co-operation  
 “the co-operation between China and her surrounding countries and developing countries”

These lengthy premodifiers make the structure of the noun phrase in which they occur extremely difficult to interpret. Some premodifiers of this sort are complicated by the fact that they are further modified by another element marked by *de* in their internal structure, as in example (60).

- (60) <N>党\_n 的\_u <<基本\_a 路线\_n 提出\_v 的\_u> 党\_n 在\_p 社会主义\_n 初级\_b 阶段\_n 经济\_n 、\_w 政治\_n 、\_w 文化\_n 的\_u> 基本\_a 纲领\_n</N>  
 <N>dang\_n de\_u <<jiben\_a luxian\_n tichu\_v de\_u> dang\_n zai\_p  
 the.Party’s GEN primary route mention DE the.Party at  
 shehuizhuyi\_n chuji\_b jieduan\_n jingji\_n 、\_w zhengzhi\_n 、\_w  
 social.ideology initial stage economy politics  
 wenhua\_n de\_u> jiben\_a gangling\_n</N>  
 culture DE basic objectives  
 “the primary principles of the Communist Party on economy, politics and culture, which are also on a par with the Party’s basic directions”

### 3.6.2.4 Serial verb constructions

Serial verb constructions in Chinese also increase the complexity of parsing. There is an immense literature on Chinese serial verb constructions (see, for instance, Li and Thompson, 1989: 594ff.; Lin and Soo, 1994; Liu, 1996). Generally speaking, a serial verb construction

refers to a succession of two or more actions that share the same subject, as illustrated in the following concocted example.

- (61) <N>我<N> <V>去 <N>朋友 家<N></V> <V>吃 <N>晚饭<N></V>  
 <N>wo<N> <V>qu <N>pengyou jia<N></V> <V>chi  
 I go.to friend home eat  
 <N>wanfan<N></V>  
 dinner  
 “I went to my friend’s house to have dinner.”

However, some of the serial verb constructions in my treebank do not conform to this general pattern of two successive verbs, each of which has a different direct object. Unlike ordinary serial verbs, the serial verbs, as shown in (62) and (63), do not take a direct object separately. They are more like compound verbs than serial verbs, though it is not clear that they can be fully assimilated to the former category. Evidence in support of this analysis comes from the fact that these verbs (i.e. 指挥\_v 演奏\_v *zhihui yanzou* “lead and perform” as in (62), and 坚持\_v 奉行\_v *jianchi fengxing* “insist and follow” as in (63)), functioning as if they were a single unit, take the same object, i.e. the following noun phrase.

- (62) <V><指挥\_v 演奏\_v> 了\_u <N>一\_m 批\_q 中外\_j 名曲\_n<N></V>  
 <V><zhihui\_v yanzou\_v> le\_u <N>yi\_m pi\_q zhongwai\_j  
 lead perform PERF one CL Chinese.and.western  
*mingqu\_n<N></V>*  
 popular.songs  
 “led and performed a variety of Chinese and western popular songs”
- (63) <坚持\_v 奉行\_v> <N>独立自主\_l 的\_u 和平\_n 外交\_n 政策\_n<N>  
 <jianchi\_v fengxing\_v> <N>dulizizhu\_l de\_u heping\_n waijiao\_n  
 insist.on follow independent DE peace diplomatic  
*zhengce\_n<N>*  
 policy  
 “insist on adopting an independent diplomatic policy in maintaining peace”

Besides sharing the same direct object, another clue that tends to prove that the two verbs are actually used as a compound verb is the suffixation of the morpheme 了 *-le*, as highlighted in (62). The verbal *-le* has generally been taken as an aspect marker, indicating completion (Norman, 1988: 163; Xiao, 2002), and it is attached to verbs and not to the objects of verbs (Chao, 1968: 247), excluding the possibility that the first verb takes the second verb (and the following noun phrase) as its object. Further research on clarifying their subcategorisation (whether they are serial or compound verbs) ought to be done in order to give a more precise parse.

### 3.7 Parses of adverbials and adverbial clauses

As stated in section 3.1, the annotation scheme that I used to manually parse the PFR treebank includes different sorts of adverbials: adverbial clauses (*Fa*), adverbial idioms/set phrases (*Ia*), adverbial adjective phrases (*Ja*), adverbial noun phrases (*Na*), adverbial prepositional phrases (*Pa*), adverb phrases (*R*) and adverbial verb phrases (*Va*). As adverb phrases, by definition, belong to the class of adverbials, they were not marked as “adverbial adverb phrases” in the scheme. A frequency table of these seven adverbial types occurring in the sample text under examination is given as follows:

Adverbial Type in PFR Sample Skeleton Treebank	Frequency
Adverbial Clause ( <i>Fa</i> )	181
Adverbial Idiom/Set Phrase ( <i>Ia</i> )	53
Adverbial Adjective Phrase ( <i>Ja</i> )	352
Adverbial Noun Phrase ( <i>Na</i> ) – Temporal	713
Adverbial Noun Phrase ( <i>Na</i> ) – Non-temporal	44
Adverbial Prepositional Phrase ( <i>Pa</i> ) – Temporal	369
Adverbial Prepositional Phrase ( <i>Pa</i> ) – Non-temporal	524
Adverb Phrase ( <i>R</i> )	2,208
Adverbial Verb Phrase ( <i>Va</i> )	187
<b>Total</b>	<b>4,631</b>

Table 5: Frequency of seven adverbial types in the sample treebank.



### 3.8 Conclusion: Quality control of the skeleton parsing process

In evaluating the success of an annotation project, Eyes and Leech (1993: 37–42) provide six essential criteria that can be used for evaluating my skeleton parsing scheme.

1. Consensual categories: The linguistic categories that were employed in my parsing scheme have been demonstrated, by comparison to seven syntactic theories, to represent grammatical features largely agreed upon by linguists, rather than features which are theory-specific or deeply controversial.
2. Overall coverage: My sample treebank represents a reasonable length of text (comprising about 100,000 word tokens or 2,500 sentences) to be manually parsed and could be re-used in future research.
3. Productivity: Productivity (i.e. the number of word tokens parsed within a reasonable length of time) was satisfactory with the simplified syntactic analysis provided by skeleton parsing.<sup>23</sup>
4. Accuracy: The output of the parsed sentences was cross-checked by several posteditors with a background in linguistics.<sup>24</sup> While one can never guarantee 100% accuracy, I believe the sample treebank to be highly accurate.
5. Uniformity of analysis: To demonstrate consistency of analysis, a concordance of the verb 要 *yao* “need” was drawn from my skeleton treebank. This verb always takes a verbal object, i.e. a verb functioning as the direct object of another verb, which is represented as *Vo* in my parsing scheme and is distinct from *V*, which stands for an independent verb phrase (see section 3.5.1.15). There are 252 instances of the verb *yao* in my treebank. In each case, it is

23 My skeleton treebank was completed in 6 months. The rate of manual skeleton parsing averages 500 words or 14 sentences a day.

24 An acknowledgment is due to the posteditors who corrected the manual skeleton parses of the sample treebank with skill and dedication: Gloria Lee, Candy Siu, Sara Wong and Sharon Wong.

followed by a verbal object consistently marked as *Vo* not *V*, as highlighted in the following figure.<sup>25</sup>

```

<V>注_定_v <Vo>要_v <Vo>失_败_v</Vo></Vo></V>
<V>要_v <J>进_一_步_d 团_结_a 起_来_v</J></V>
<V>要_v <Ja>更_d 好_a 地_u</Ja> <Vo>坚_持_v <N><解_放_思_想_j 、_w 实_事_求_是_j 的_u> 思_想_n 路_线_n</N></Vo></V>
<V>要_v <Ja>更_d 好_a 地_u</Ja> <Vo>坚_持_v <P>以_p <N>经_济_n 建_设_vn</N></P> <Vo>为_v <N>中_心_n</N></Vo></Vo></V>
<V>要_v <Vo>看_v <Vo>能_否_v <P>把_p <N>经_济_n 工_作_vn</N></P> 搞_v 上_去_v</Vo></Vo></V>
<V>要_v <Ja>切_实_ad</Ja> <P>把_p <N>精_力_n</N></P> <Vo>集_中_v <Vo>到_v <N><贯_彻_v 落_实_v 好_a 中_央_n 关_于_p 今_年_t 经_济_n 工_作_vn 的_u> <N><N&&总_体_n 要_求_n</N&& 和_c <N+>各_项_r 重_要_a 任_务_n</N+> 上_f 来_v</N></N></Vo></Vo></V>
<V>要_v <Ja>更_d 好_a 地_u</Ja> <Vo>坚_持_v <N> “<_w 两_手_抓_l 、_w 两_手_m 都_d 要_v 硬_a ”_w 的_u 方_针_n</N></Vo></V>
<V>要_v <Ja>更_d 好_a 地_u</Ja> <Vo>发_扬_v <N>求_真_务_实_l 、_w <密_切_ad 联_系_v 群_众_n 的_u> 作_风_n</N></Vo></V>
<V>要_v <Vo>发_扬_v <N>这_样_r 的_u 好_a 作_风_n</N></Vo></V>
<V>要_v <Vo>尊_重_v <N>群_众_n 的_u 意_愿_n</N></Vo></V>
<V>要_v <R>大_力_d</R> <Vo>倡_导_v <N><说_实_话_l 、_w 办_v 实_事_n 、_w 鼓_v 实_劲_n 、_w 讲_v 实_效_n 的_u> 作_风_n</N></Vo></V>
<V>要_v <Vo>坚_守_v <N>岗_位_n</N></Vo></V>
<V>要_v <Vo>让_v <S><N>他_们_r</N> <Ja>健_康_ad</Ja> <V>成_长_v</V></S></Vo></V>

```

Figure 6: A concordance of the verb *yao*.

6. Linguistic validity: One of the aims of carrying out a skeleton parsing on a sample text of the PFR Chinese Corpus is to gain a better understanding of how to precisely locate adverbial clauses in a piece of POS tagged text. In written Chinese, adverbial clauses are typically overtly marked by subordinating conjunctions (or subordinators) of various sorts. Hence in the next chapter, I will describe in more detail the role of subordinators that are commonly employed in introducing adverbial clauses.

25 The verb *yao* can take an adjectival direct object, as shown in one instance of the concordance extracted. A vast majority of adjectives may function as verbs in Mandarin Chinese by taking aspect markers (e.g. *-guo* “experiential aspect”, *-le* “perfective aspect”, etc.) or directional complements (e.g. *qilai* “inchoative”). See Li and Thompson (1989: 141–147).

