Chapter Nine

Conclusion

9.1 Summary of findings

In the introductory chapter of this book (section 1.2), I set out *seven* research objectives related to Chinese adverbial clauses. My research brought out the following results, presented in the following sub-subsections.

9.1.1 Identification of adverbial clauses

To identify adverbial clauses in my corpora, I carried out skeleton parsing on a sample text of approximately 100,000 word tokens (or about 2,500 sentences) taken from the PFR Chinese Corpus with a clearly defined parsing scheme of 17 constituent labels. The manually-parsed sample skeleton treebank is one of the very few extant Chinese treebanks (cf. Xue et al., 2000). While Chinese part-of-speech tagging (Liu et al., 1995; Zhang and Sheng, 1997) and word segmentation (Wong and Chan, 1996; Zhang and Sheng, 1996; Sun et al., 1998; Yuan and Kim, 1998; Sun et al., 2000) have been the subject of concerted research for many years, the syntactic annotation of Chinese corpora is a comparatively new field. Although much treebanking of English has occurred, relatively little of such work has been done on Asian languages, Chinese included (Han et al., 2002). As shown in my skeleton treebank, adverbial clauses in Chinese are typically overtly marked by a subordinating conjunction. Moreover, the difficulties that I encountered in the production of this treebank demonstrate some of the peculiarities of Chinese syntax. The particle 的 *de* is typically used to introduce a lengthy premodifier in a noun phrase. These long multi-word noun phrase premodifiers greatly complicate the structure of the noun phrase in which they occur and thus pose enormous problems in parsing a sentence in Chinese. Another noteworthy syntactic property is that some serial verb

constructions tend to be used as if they were compound verbs. When these serial verb constructions occur, the two transitive verbs in series, unlike common transitive verbs, do not take an object separately within the construction. Rather, the serial construction as a whole is able to take the same direct object and the verbal suffix 了 -le. Thus they appear to be used as a compound verb functioning as a single unit.

9.1.2 Adverbial subordinators in Chinese

As no clear distinction has been made between coordinating conjunctions and subordinating conjunctions in Chinese grammars, I proposed in this book two working criteria for a given item to be regarded as an adverbial subordinator: first, it introduces a clause which must be linked to another clause in the same sentence; second, the relation of these two clauses is not of coordinate nature. There are in total 57 subordinators identified in the PFR corpus, most of which are correlative subordinators having either an adverb or a conjunction as their correlative. The correlatives are not obligatory and an adverbial subordinator can take more than one correlative.

9.1.3 Semantic classes of adverbial clauses

I adopted a problem-oriented tagging approach to identify adverbial clauses in my corpus and to analyse their semantic roles. This meant that only relevant parts of the corpus, not the entire corpus, were annotated to meet my specific research purposes: only the sentences in which the adverbial clause is overtly marked by a subordinating conjunction were annotated. I identified a total of 2,417 instances of adverbial clauses in the PFR corpus and subsumed them into eleven semantically based classes which include, in descending order of frequency, clauses of condition, clauses of concession, clauses of purpose, clauses of cause/ reason, clauses of contrast, clauses of result, clauses of exception, clauses of addition, clauses of inference, clauses of preference and clauses of time. Some observations about these adverbial clauses were made. For instance, 而是 *ershi* "rather" interacts with negated verb forms and negative adverbs that occur in the main clause. In addition, the appar-

302

ent overlap between purpose clauses and result clauses is marked by the use of the same adverbial subordinator 从而 *conger* to introduce these two kinds of adverbial clause. While time, inference and preference clauses occur relatively infrequently and are marked by no more than two adverbial subordinators, conditional and concessive clauses make up approximately half of the adverbial clauses studied and exhibit a vast diversity of subordinating conjunctions used to introduce them. Clauses of inference were considered as a distinct adverbial semantic class in this book as they can be distinguished from apparently functionally analogous conditional and concessive clauses.

9.1.4 A government and binding approach to the distribution of subjects of adverbial clauses

The non-overt subjects (PRO) in Chinese adverbial clauses (CACs) were proven to support the PRO theorem of Government and Binding (GB) Theory in that they occur only in ungoverned positions and are properly licensed. According to control theory, they can be referentially dependent on, or controlled by, another NP or an implicit argument (e.g. the agent of an action) in the main clause, or they are not controlled at all and have an arbitrary interpretation. As a feature specific to PRO in CACs, it can be controlled by an NP occurring in the context prior to the main clause. As this type of control cannot be explained by control theory, the Principle of the Separation of Reference and Role (PSRR) was used to account for it. Furthermore, it was shown that there is a high statistical significance for the hypothesis that preference for a particular kind of subject (either overt NP or PRO) depends on the interclausal semantic relations of adverbial clauses. I found that in the PFR corpus, overt NP subjects are preponderant in concessive clauses while null subjects mostly occur in conditional clauses. This distribution pattern is attributed to the fact that the subject of the adverbial clause may be different from that of the main clause, resulting in the occurrence of overt NP subject in favour of PRO, which is always co-referential with either the main clause subject/object or an NP in the context. The distribution of overt and non-overt subjects also differs significantly in clauses of purpose, contrast and result which favour the use of PRO and in causal clauses which favour overt NP subjects. Other types of adverbial

303

clause, however, do not show a marked preference for either subject type such as clauses of exception, addition, inference, preference and time. An integrated approach which combines a theoretically-informed corpus-based approach with an information-structure-based approach was taken to explain the distribution of PROs across the semantic domains of adverbial clauses. An NP referential controller occurring in the main clause or previous context can be seen as given information and thus the subject of the adverbial clause which co-refers to it can be dropped as in clauses of condition, purpose, contrast and result. On the other hand, the subject of the main clause which is an NP with a brand-new referent and is different from that of the adverbial clause can be seen as new information. Thus the subject of the adverbial clause cannot be dropped as in clauses of concession and cause/reason.

9.1.5 Distribution of adverbial clauses across text types in written Chinese

The results obtained from the PFR corpus were compared with the results obtained from the Lancaster Corpus of Mandarin Chinese (LCMC), a balanced corpus with 500 samples of fifteen distinct text types, to explore the influence of text type on both the distribution of semantic classes of adverbial clauses and the distribution of subjects in the adverbial clause. While in both the PFR and LCMC corpora conditional and concessive clauses dominate all other semantic types of adverbial clause, a comparison of their occurrence in the journalistic texts of the LCMC corpus (i.e. press reportage, editorials and reviews) and that in the non-journalistic texts (i.e. other text categories remaining in the corpus) indicated that they are not a marked feature of journalistic writing. Rather, conditional and concessive clauses are most frequently used in category E (skills/trades/hobbies) and category N (adventure and martial arts fiction) of the corpus respectively. Other semantic classes of adverbial clauses also have interaction with different text types of the corpus. Reason and result clauses are closely related to category D (religion) and category J (scientific academic prose) in that they are most commonly used to describe the origin and development of a religion and to discuss methods and findings of a piece of research, whereas purpose clauses interact with category E (skills/trades/hobbies) and

304

category F (popular lore) to give instructions for mastering a particular skill and hobby and to offer advice on issues related to local culture and everyday life such as fashion, diet, interpersonal relationship, family and careers.

9.1.6 Distribution of subjects of adverbial clauses across text types and semantic domains in written Chinese

It was shown that the distribution of subjects of adverbial clauses varies significantly across text types in the LCMC corpus. Yet this disguises the fact that the significant difference in the distribution of overt and non-overt subjects is merely appreciable in six out of fifteen text categories of the corpus. This does not, however, preclude text type from exerting an influence on the distribution of subjects in the adverbial clause. Results obtained from the LCMC corpus demonstrated that the distribution of subjects varies significantly across semantic domains of adverbial clauses as in the PFR corpus: while non-overt subjects are used overwhelmingly in clauses of contrast and purpose, overt subjects are mostly used in clauses of concession, reason and exception. In contrast to the results obtained from the PFR corpus, conditional clauses do not show a significant preponderance of non-overt subjects in the LCMC corpus. This unexpected finding gives some credence to the claim that text type influences the distribution of subjects in the adverbial clause. A closer examination of the distribution of subjects across text types in the five adverbial clause types giving statistically significant results on their choice of subject further confirms it: the contrast between the distribution of overt and non-overt subjects is marked in certain text categories only. Hence, the effect of the semantic domains of adverbial clauses on the distribution of PROs as shown in the PFR corpus indeed depends on text type. It was also shown that text type influences the choice of the type of control of PRO. Non-overt subjects with an arbitrary interpretation (PRO_{arb}) occur twice as frequently in the LCMC corpus as in the PFR corpus. They are most frequently used in text category E which is composed of articles on teaching particular skills and hobbies to general public. Since no specific referent is intended to be the reader in these articles, category E favours the use of PRO_{arb} more strongly than other text types.

9.1.7 Distribution of adverbial clauses and their subjects in spoken Chinese

In comparing spoken and written Chinese corpora, adverbial clauses in spoken Chinese typically demonstrate two features of speech i.e. fragmentation and involvement, as Chafe (1982) identified: (i) an adverbial clause occurring in several different turns from the associated main clause uttered both by the same speaker demonstrates a degree of fragmentation because both the adverbial clause and the main clause are expressed in several idea units or in fragmented grammatical structures; (ii) that the adverbial clause and the main clause are uttered separately by two different speakers who attempt to ensure communication is functioning well demonstrates a quality of involvement. A subcategory of adverbial clause, an adverbial-main clause, was proposed to address the peculiarity of spoken language in that the main clause to which an adverbial clause is subordinated is not always explicitly expressed and it can, however, be inferred from the context. A speaker's interpretation of a non-overt subject also relies heavily on context, as shown in the relatively higher frequency of controllers for PRO identified in the context (i.e. contextual controllers) than subject and object controllers in the main clause. In contrasting the distribution of adverbial clauses and their subjects in CALLHOME with that in the LCMC, I showed that (i) both conversation and narrative texts show a strikingly similar distribution pattern of adverbial clauses in that purpose, result and contrast clauses are relatively less frequent than other adverbial clause types; (ii) clauses of condition and cause/reason, however, have a disproportionately higher frequency of occurrence in conversation than in narrative texts; (iii) clauses of reason and exception take more overt subjects than PROs in both spoken and written Chinese; (iv) unlike their counterparts in written Chinese, clauses of concession in spoken Chinese do not show a marked preference for overt subjects, marking an important difference in the use of adverbial clauses between spoken and written Chinese.

9.2 Limitations of the present study

The spoken Chinese corpus used in this book contains just one type of spoken discourse i.e. telephone conversations. This is very limited in comparison with corpora such as the London-Lund Corpus of Spoken English (Svartvik and Quirk, 1980; Johansson, 1982), which represents six major speech situations: face-to-face conversations, public conversations including debates and interviews, telephone conversations, radio broadcasts, spontaneous speeches and prepared speeches. However, at the time of writing, there is no spoken Chinese corpus available which is a heterogeneous collection of varied kinds of spoken texts to match corpora such as London-Lund. Whilst I am aware that the findings based on a spoken corpus of conversation data cannot be easily transferred to some spoken discourse types, especially those which differ considerably in nature from conversation e.g. prepared public speeches, the CALLHOME corpus used in my work can at least throw some light on the primary differences between spoken and written Chinese.

9.3 Suggestions for future research

9.3.1 Annotation of functions of syntactic constituents

Further research can be conducted into the functional labelling of syntactic constituents in a treebank. That was not undertaken in this book. In my skeleton treebank, functional labels were put aside in order to give a consistent and accurate manual parsing. However, the annotation of syntactic functions may throw up interesting results regarding the range of functions that a phrasal category can take within a sentence; a phrasal category may assume a syntactic function that is not conventionally associated with it. Additionally, a functionally labelled treebank may provide a broader coverage of syntactic phenomena. Take, for example, the case of 把 *ba* constructions in Chinese as mentioned in my book (see Chapter Three, section 3.6.2.1). The annotation of traces

307

for displaced constituents can better explain the peculiarity of those prepositional phrases headed by the preposition *ba*, in which the prepositional complement co-refers to the (displaced) null object of the following verb phrase. It is expected that more large-scale treebanks with expanded size and coverage will be built in the near future (cf. Han et al., 2002).

9.3.2 A contrastive study of adverbial clauses between English and Chinese

As mentioned earlier (see Chapter Two, section 2.2.1; Chapter Seven, section 7.1), the LCMC corpus is a Chinese match for the FLOB corpus (on written British English) in terms of sampling frame. The FLOB corpus has a matching American English corpus, the Freiburg-Brown Corpus of American English or Frown (Hundt et al., 1999). As the LCMC, FLOB and Frown corpora are comparable corpora, it is feasible to study Chinese in contrast with the two major varieties of English (McEnery et al., 2003; McEnery and Xiao, forthcoming). By examining the distribution of non-overt subjects of adverbial clauses in the fifteen text categories of the LCMC and FLOB/Frown corpora, the distribution patterns of non-overt subjects in Chinese and British/American English adverbial clauses could be compared.

9.4 Concluding remarks

This book is aimed at offering a theoretically informed, corpus-based, comprehensive analysis of adverbial clauses in spoken and written Chinese. Two advantages of the corpus-based approach are demonstrated in this book. Firstly, the exploitation of a representative corpus enables quantitative statements to be made on the distribution of linguistic forms across a language variety and avoids idiosyncratic bias present in studies based on a researcher's intuition (Collins, 1991; Fillmore, 1992). Secondly, in the study of a grammatical category, a sizeable corpus can provide researchers with a wealth of language data upon which to base

308

their explanations (Meyers, 1991), and it allows these explanations to include information on authentic language use of the grammatical construction under consideration i.e. to be accountable to interpersonally-observable evidence (Sampson, 2001 and 2005).