



May L-Y Wong

Adverbial Clauses in Mandarin Chinese

A Corpus-based Study



Peter Lang

Chapter One

Introduction

1.1 Corpus-based approach to studying adverbial clauses

This book is motivated by the fact that none of the previous analyses of adverbial clauses in Chinese have based their illustrative examples and exposition on extensive corpus evidence. Rather, researchers have typically relied on their own intuitions about language (e.g. Liu et al., 1996; Chu and Chi, 1999), sometimes supplemented by adapting example sentences from influential novels (e.g. Ding et al., 1979).

Recent work by Wang (1995, 1998, 1999 and 2002) breaks fresh ground in studying adverbial clauses by adopting a corpus-based approach. She quantitatively analyses the distribution and information structure of four main types of adverbial clause (*viz* temporal, conditional, concessive and causal clauses) in spoken Chinese on the basis of a corpus of six hours worth of naturally occurring face-to-face, two party, and multi-party conversations and call-in broadcasts on local radio and television in Taiwan. However, her studies focus solely on a limited range of adverbial clauses and are largely based on the spoken register of Chinese. Also her spoken corpus is rather small and yields just some 700 adverbial clauses in total. Hence, the novelty of a corpus-based study on adverbial clauses in written Chinese as well as an in-depth analysis on the typology of adverbial clauses in Chinese argue for a more thorough quantitative and qualitative account of them in order to discover new insights into their use in written data. Furthermore, as far as adverbial clauses are concerned, theoretically informed corpus-based research is rare (cf. Quintero, 2002). More work can therefore be done on marrying corpus linguistics with linguistic theory in this area. This book aims to achieve such a marriage.

To investigate the use and structure of a grammatical construction, most researchers have found it profitable to investigate constructions that occur relatively frequently, since if a construction occurs too in-

frequently, it is often hard to make strong generalisations about its form and usage (Meyer, 1991). For this reason, to study infrequent linguistic constructions, it is often necessary to study reasonably large corpora, like the two corpora of written Chinese used in this book, both of which contain one million words, namely the PFR Chinese Corpus and the Lancaster Corpus of Mandarin Chinese (henceforth LCMC).

1.2 Research objectives and the organisation of the book

Given the need for a corpus-based approach to linguistic theory and the need for a more extensive corpus-based account of a wider spectrum of adverbial clauses in written Chinese, this book uses a skeleton treebank (i.e. a corpus annotated with basic level syntactic constituents) to explore the syntactic structure of the Chinese language in order to shed light on the following research questions.

- (1) How does the sample skeleton treebank help in the identification of adverbial clauses and revealing the peculiarities of Chinese syntactic properties?
- (2) What are the adverbial subordinators in Chinese that are responsible for overtly marking adverbial clauses?
- (3) Which semantic roles do these adverbial clauses play in relation to the main clause they modify?
- (4) Does the PRO theorem in the Government and Binding (GB) Theory apply in written Chinese?
- (5) How do the semantic types of adverbial clauses vary across genres/text types in written Chinese?
- (6) How does the distribution of PROs vary across both semantic domains and text types in written Chinese?
- (7) Do research findings based on written Chinese hold for spoken Chinese?

In the course of exploring the adverbial subordinators in the PFR corpus, a critique will be provided of the catch-all term 连词 *lianci* “conjunction” as it has been used in Chinese grammars to refer to both a

coordinating conjunction and a subordinating conjunction (Lu and Ma, 1990; Hou, 1998). As will be demonstrated in Chapter Six (section 6.1), Chinese is a pro-drop language (Huang, 1989) i.e. a language which allows the omission of a subject in a clause. While there is an immense literature on null subjects in Chinese (see, for example, Huang, 1987 and Chen, 1990), the focus of the previous literature was on pro-drop in complement clauses and not on pro-drop in adverbial clauses. Hence, my book contributes by investigating null subjects in Chinese adverbial clauses in order to fill the gap in the literature of the pro-drop phenomenon. In particular, this book focusses upon the distribution of non-overt subjects across various semantic types of adverbial clauses because certain of these adverbial clause types (e.g. purpose and contrast clauses) may show a stronger tendency for dropping the subject than other types. By *a priori* reasoning, if a person performs an action (as described in the main clause), s/he must intend to do it for a particular purpose (as described in the adverbial clause of purpose). Thus the subject of the purpose clause is likely to be omitted, which is always the same as the subject of the associated main clause. Clauses of contrast make a contrast between two situations described in the main clause and the adverbial clause. The two situations are closely related to each other as they are in fact two contrasting descriptions regarding the same subject; the situation of the main clause is taken to be wrong and the situation of the adverbial contrastive clause is what is right about the subject of the main clause. It is therefore hardly surprising that the subject of the contrastive clause, which is co-referential with the subject of the main clause, can be dropped. Yet all of these predictions stem purely from intuitions about the behaviour of adverbial clauses. To test these introspective assumptions, a corpus-based analysis is conducted in this book into how null subjects distribute across adverbial semantic classes.

In pursuing these research objectives, my book is organised into three major parts. The first part, Chapter Two, deals primarily with issues relating to the PFR Chinese Corpus, including a brief history of the construction of this corpus and the annotation of sentence boundary markers in the part-of-speech (POS) tagged corpus; the LCMC corpus will also be briefly described. Though it is a corpus-based study, my book is not atheoretical. In my book, a corpus-based approach to theory is advocated. The approach taken to the investigating of my research questions is as follows: rather than set out to use corpus data to testify the validity

of theoretical assumptions, I start my research by examining my corpus data closely, looking for any systematic patterns in the behaviour of the adverbial clause; those patterns or properties of adverbial clauses are then explained in a theoretical framework that lends itself well to the analysis of similar phenomena. In other words, my work does not presuppose the use of nor the rejection of a particular theoretical framework; rather, when it becomes relevant to my discussion of the corpus data, a theory is selected on its merits and adopted to explain my findings. Hence in the second part of my book (Chapters Three to Five), as a prelude to the theory based approach of the third part, initial results are presented in relatively theory-neutral terms, to make the emerging patterns of adverbial clauses in Chinese as accessible to linguists as possible. In the third part (Chapters Six to Eight), the same data are analysed within the Government and Binding Theory Framework in order to understand the description of the adverbial clause developed in this book within a theoretical framework in which a theorem (PRO theorem) is important to the explanation of the occurrence of non-overt subjects in the adverbial clause. The findings concerning the distribution of non-overt subjects are then put to the test in the LCMC corpus which, unlike the PFR corpus, is a balanced corpus with fifteen text types of written Chinese and can therefore provide a sound basis for making reliable generalisations of the properties of adverbial clauses in written Chinese across a range of genres. A contrastive study of the distribution of non-overt subjects in the adverbial clauses in spoken and written Chinese is also conducted on the basis of the CALLHOME Mandarin Chinese Transcripts Corpus, which is a spoken Chinese corpus developed in 1996.

1.2.1 Brief chapter summaries

Chapter One, the present chapter, expressly states the rationale and research objectives of this book.

Chapter Two will present a brief review of the development of written Chinese corpora and their use in linguistics and beyond. The two written corpora used in this book, the PFR Chinese Corpus and the LCMC corpus, will also be described. A literature review of previous studies of Chinese adverbial clauses will be presented, most of which is concerned with the discourse functions of adverbial clauses in different