CHAPTER 1 INTRODUCTION

> Many grammars (even relatively modern ones) are not very sophisticated or accurate when it comes to verb aspect. This is one of my spot checks — one of several — that I use to gauge the quality of a new grammar [...] I check what they have to say about aspect. (Mark B. O'Brien 1997)

This book is about 'aspect', and more specifically aspect in Chinese. The book outlines a new model of aspect, based upon our investigation of the Chinese language using a fusion of native speaker intuition and evidence from corpora of Mandarin Chinese. The overall purpose of the book in part is to demonstrate how corpora and linguistic theory can interact. But this secondary goal should not distract readers from the main purpose of this book – to outline a new model of aspect which is generalisable beyond Chinese. As part of our development of this model, we will use a corpus of modern British English to test the generalisability of our findings for Chinese on English. In order to begin the process of presenting this model, we need to answer two basic questions – "what is aspect" and "why study aspect in Chinese."

1.1. What is aspect and why study aspect in Chinese?

To begin with, let us develop a working definition of aspect which will suffice as an answer to the question "what is aspect?" for now: aspect is an important linguistic category which relates to the study of linguistic devices that enable a speaker to direct the hearer's attention to the temporality of a situation, either intrinsic or viewed from a certain perspective. Such knowledge is required for interpreting event sequences in discourse (Dowty 1986; Moens & Steedman 1988; Passonneau 1988), processing temporal modifiers (Antonisse 1994), and describing allowable alterations and their semantic effects (Resnik 1996; Tenny 1994). The perspectives from which situations are presented "sometimes uniquely determine, and sometimes just strongly constrain" the underlying temporal structures of these situations (Nakhimovsky 1988:33; cf. also Bickel 1997:115). Aspect plays an important role in the interpretation of the temporal information contained in a sentence. Unsurprisingly, considering its importance, aspect is a common linguistic feature which has been the subject of research in a number of areas such as linguistic theory (e.g. Comrie 1976), language philosophy (e.g. Galton 1984), language typology (e.g. Dahl 1985), language acquisition (e.g. Salaberry & Shirai 2002) and computational linguistics (e.g. Passonneau 1988).

While aspect and tense both provide temporal information, they are two different concepts. On the one hand, tense is deictic in that it indicates the *temporal location* of a situation, i.e. its occurrence in relation to a specific reference time. On the other hand, aspect is non-deictic in that it is related to the *temporal shape* of a situation, i.e. its internal temporal structure and ways of presentation, independent of its temporal location (cf. Lyons 1977:705).

Having given a brief description of aspect, we can now turn to our second question – why study aspect via the Chinese language? Chinese has a different relation to aspect by comparison to a language such as English. Languages can broadly be classified as tense languages and aspect languages depending upon how they denote time relations. In a tense language, such as English, tense and grammatical aspect are often combined morphologically. For example, in English the simple past not only presents a situation as perfective, but also locates it prior to the

speech time; similarly, the French *imparfait* is both past and perfective. However, grammatical aspect and tense can also be encoded distinctly, as demonstrated in Polish (Weist et al. 1984). In contrast with a tense language like English, Mandarin Chinese does not have the grammatical category of tense (cf. Smith 1997; Kang 1999; *inter alia*), because the concept denoted by tense is lexicalised, i.e. indicated by content words like adverbs of time. Aspectual meanings, however, are conveyed systematically by aspect markers — grammaticalised function words that are semantically encoded to convey aspectual meanings. In other words, Chinese grammatically marks aspect but does not grammatically mark tense. As such, Chinese is exclusively an aspect language (cf. Wang 1943; Gao 1948:189; Gong 1991:252; Norman 1988). As an aspect language, Mandarin Chinese has played an important role in the development of aspect theory. Nearly all of the major works on aspect theory make reference to Chinese (e.g. Comrie 1976; Smith 1997). It is on these grounds that we decided to focus on Chinese in investigating aspect.

Aspect languages are "concerned with whether the action is completed or not, whether the action is in progress or not. The plotting of action, so important in tense languages, is not important in Chinese" (Norman 1988:163). This observation is in line with the claim made nearly four decades earlier by Wang Li:

> Broadly speaking, an event and time can be related by 1) focusing on when it happens while ignoring its temporal distance and length; and 2) focusing on its time duration and whether it starts or finishes while ignoring when it happens. The first approach is taken by Romance languages (like French, Italian and Spanish, etc.) while the second is typical of Chinese. (Wang 1943:151, our translation)

Aspect markers are pervasive in Chinese. As such it is necessary to take account of aspect markers in Chinese when interpreting almost any Chinese utterance. The main focus of this book therefore, will be on grammatical aspect where the tense/aspect dichotomy is relevant. However, before proceeding to outline the research presented in this book, it is appropriate to outline previous research undertaken on aspect in Chinese.

1.2. Previous research on aspect in Chinese

Unfortunately, previous research on aspect in Chinese is deficient in a number of ways. With one or two exceptions, previous research on aspect in Chinese has been limited to a few aspect markers like *-le*, *-zhe* and *-guo*.[1] Little attention has been paid to date to the question of systematically describing the linguistic devices that the language employs to express aspectual meanings. Still less attention has been paid to the inherent temporality of situations denoted by utterances in Chinese. But aspect markers that signal different perspectives from which a situation can be presented are only one component of aspect, which interplays with the inherent temporal features of a situation to determine the aspectual meaning of an utterance (see chapter 2).

While Chinese is recognised as an aspect language, and aspect marking has been studied intensively in Chinese linguistics in the last three decades,[2] there is no generally agreed account of the aspect system of this language, as different researchers define aspect in their own ways. As a consequence there is much

---

[1] Chinese aspect marker and examples are given in Romanised form using Pinyin symbols.

[2] Klein, Li & Hendriks (2000:723) estimate conservatively that over 200 articles have been published on the linguistic analyses of aspect markers in Chinese.

controversy surrounding the form and function of aspect markers. This leads to the following questions:

- Is it necessary to distinguish the verbal *-le* and the sentential *le*? Does the verbal *-le* function to mark the completiveness, terminativeness or simply the realisation of a situation? Can *-le* interact with stative situations?

- Does the form of marker *-zhe* function to signal resultativeness, progressiveness or durativity?

- Does verb reduplication function semantically as an aspect marker? If so, does it encode tentativeness, casualness, mildness, slightness or delimitativeness?

- Is it necessary to differentiate between the aspect marker *-guo* and the resultative verb complement (RVC) *guo*?[3] How can one account for the interchangeability between *-le* and *guo* as in *mingtian ni chi-guo/-le wanfan lai zhao wo* "Come to see me after you have supper tomorrow"?

- Do the forms of *zai*, *-qilai*, *-xiaqu* and RVCs encode aspect?

These questions, which are addressed and answered in this book, serve a clear purpose for the moment – they clearly show that aspect in Chinese is an area of ongoing debate and research.

1.3. Studying aspect: intuition vs. corpus-based approaches

While "linguistic analysis will benefit if it is based on real language used in real contexts" (Meyer 2002:11), previous studies of aspect have largely been conducted without recourse to attested language data. They have, rather, been based on a handful of confected examples which, if not intuitively unacceptable, are atypical of attested language use. Furthermore, those proposals have not, to date, been tested with corpus data. As far as we are aware, with few exceptions (e.g. Chappell 1988, 1998), to date there has been little research on aspect in Chinese based on corpus data. Yet corpora have a role to play both in developing and testing such theories, as demonstrated in the remaining chapters of this book.

With that said, we do not mean that the corpus-based approach and the intuition-based approach are completely exclusive. The two are complementary (cf. McEnery & Wilson 2001:19). With the intuition-based approach, researchers can invent purer examples instantly for analysis, because intuition is readily available and invented examples are free from language-external influences existing in naturally occurring language. However, intuition should be applied with caution (cf. Seuren 1998:260-262). Firstly, it is possible to be influenced by one's dialect or sociolect. As such, what appears unacceptable to one speaker may be perfectly felicitous to another (cf. Xiao 2002:17). Secondly, when a researcher invents an example to support or disprove an argument, he is consciously monitoring his language production. Therefore, even if his intuition is correct, the utterance may not represent typical language use. Finally, results based on introspection alone are difficult to verify as introspection is not observable. In contrast, all of these disadvantages are circumvented by the corpus-based approach. Additional advantages of the corpus-based approach are that a corpus can find differences that intuition alone cannot

---

[3] RVC is an acronym of a 'resultative verb complement' like *open* in *push the door open* (see sections 3.4.1 and 4.4).

perceive (cf. Francis, Hunston & Manning 1996; Chief, Hung, Chen, Tsai & Chang 2000) and a representative corpus can yield reliable quantitative data. As we wish to both quantify aspect in Chinese and develop a model of aspect, we will use corpus data as our main source of evidence throughout this book. Where appropriate, we do call on native speaker intuition. Nonetheless, attested language data is the principal source of evidence that we use.

The use of corpus data as an input to the semantic analysis of aspect, a methodology to be elaborated in the following chapters, represents something new. Our study seeks to achieve a marriage between theory-driven and corpus-based approaches to linguistics, with the goal of providing an effective and fruitful avenue for the study of aspect.

Having decided to take a corpus-based approach, we can now present the corpus data used in this book. We used five corpora in developing and testing the aspect model presented in this work, two monolingual corpora of Mandarin Chinese, namely Weekly and LCMC, two English corpora (FLOB and Frown) and an English-Chinese parallel corpus. All of the corpora are annotated with part-of-speech information and the parallel corpus is further aligned at the sentence level.

The Weekly corpus. The principal Chinese corpus on which the research presented in this book is based consists of newspaper texts from *Nanfang Zhoumo* "The South Weekly" and is thus named the Weekly corpus. *Nanfang Zhoumo*, with a sales volume of 1.3 million copies, is one of the largest, most influential and comprehensive weekly newspapers published in China. The sampling period for our corpus covers one calendar year, with the data included in the corpus being taken from the CD-ROM edition of the newspaper for the year of 1995. We divided our corpus into a training and a test set of data. The training set provided training material for our model while it was under development. The test set provided data that could serve as an unseen test for our model developed on the basis of the training data. Our test and training data follows best practice, in that a good test corpus is qualitatively similar to the training corpus, but contains data which the learning algorithm or model has never seen before (cf. van Everbroeck 1996). In terms of size, a test corpus with a size of one tenth of the training corpus is normally assumed to be sufficient. The test corpus used in this book follows these guidelines. The training corpus contains 125,825 Chinese characters and the test corpus contains 12,869 characters. In terms of content, the test corpus mirrors the training corpus, covering a range of topics such as society, economic, legal and arts news. It is this corpus that we use in the tests carried out on our model of aspect later in this book, with the training corpus only being used to provide examples and as the basis of the general model development.

The Weekly corpus is small, the training corpus being merely 125,825 Chinese characters in size. Our defence of the use of this corpus is that, while small, the corpus contains sufficient examples of the linguistic feature we are interested in, i.e. aspect markers. As shown in Table 1.1, there are plenty of examples of these markers in the corpus. The high frequency and rich variety of aspect markers in the corpus not only furnishes further evidence that Chinese is an aspect language, but also justifies our choice of this corpus for the study of aspect in Chinese. Furthermore, the corpus

achieved a sufficiently representative coverage, for a corpus of that size, of styles and domains.[4]

Aspect marker	POS tag	Frequency	
		Training corpus	Test corpus
Actual -*le*	ACTL	1,019	119
COS *le*	COS	164	11
Double-role *LE*	DBL	23	4
Experiential -*guo*	EXP	75	9
Durative -*zhe*	DUR	196	42
Progressive *zai*	PROG	77	11
Inceptive -*qilai*	INC	18	2
Continuative -*xiaqu*	CONT	8	0
Delimitative verb reduplication	VDUP	34	4
Completive RVC (*wan*, *guo* and *hao*)	RVCC	33	12
Directional RVC	RVCD	740	92
Result-state RVC	RVCS	780	84
TOTAL		3,167	390

Table 1.1 Frequency of aspect markers in the Weekly corpus

The FLOB/Frown corpora. While this book is principally concerned with aspect in Chinese, the model developed in this book is tested by contrasting Chinese and English. As English corpora are readily available for research purposes, we did not have to build an English corpus in order to do this. After reviewing available corpora, we decided to use the Freiburg-LOB Corpus of British English (i.e. FLOB, cf. Hundt, Sand & Siemund 1998) as its sampling period is close to that of the Weekly corpus built by us. A further attraction of FLOB is that it has a matching American English corpus, the Freiburg-Brown corpus (i.e. Frown, cf. Hundt, Sand & Skandera 1999).

Code	Text category	No. of samples	Proportion
A	Press reportage	44	8.8%
B	Press editorials	27	5.4%
C	Press reviews	17	3.4%
D	Religion	17	3.4%
E	Skills, trades and hobbies	38	7.6%
F	Popular lore	44	8.8%
G	Biographies and essays	77	15.4%
H	Miscellaneous (reports, official documents)	30	6.0%
J	Science (academic prose)	80	16.0%
K	General fiction	29	5.8%
L	Mystery and detective fiction	24	4.8%
M	Science fiction	6	1.2%
N	Western and adventure fiction	29	5.8%
P	Romantic fiction	29	5.8%
R	Humour	9	1.8%
Total		500	100%

Table 1.2 Text categories of FLOB

---

[4] At the time of writing, a much larger balanced corpus, the Lancaster Corpus of Mandarin Chinese, is under construction at Lancaster University (see below). However, as the corpus is released in the later stages of this research, we decided to take the Weekly corpus as the major source of empirical evidence while shifting our focus to trying to minimise any adverse effects arising from the limited size of the Weekly corpus.

FLOB is a balanced corpus of present-day British English compiled at Freiburg University in 1991-1992. The sampling frame of the corpus is exactly the same as that used in the compilation of LOB (the Lancaster-Oslo-Bergen corpus, see Johansson, Leech & Goodluck 1978) with the notable exception that LOB was sampled from texts produced in 1961 whereas FLOB was sampled from texts current in 1991-1992. The corpus contains 500 text segments of approximately 2,000 words sampled from 15 text categories (see Table 1.2), totalling roughly one million words.

The two Freiburg corpora share the exactly same parameters except that FLOB sampled British English while Frown sampled American English. They will be used in combination with the LCMC corpus, in chapter 6, to contrast aspect marking in Chinese, British English and American English.

The Lancaster Corpus of Mandarin Chinese (LCMC). While the main concern of this book is to model aspect in Chinese, we will also contrast aspect marking in English and Chinese, in chapter 6, on the basis of LCMC and FLOB/Frown. LCMC is a one-million-word balanced corpus of written Mandarin Chinese.[5] The corpus was designed as a Chinese match for FLOB (see McEnery, Xiao & Mo 2003). In addition to monolingual studies of the Chinese language, LCMC, in combination with FLOB/Frown, is also a sound basis for contrastive studies of Chinese and English. As McEnery & Xiao (forthcoming) observe, two well-matched monolingual corpora serve as a more reliable resource than a parallel corpus for contrastive studies, whether one wishes to compare the two languages as a whole or compare them by text type (see section 6.1).

In LCMC, the FLOB sampling frame is followed strictly except for two minor variations. The first variation relates to the text categories covered – we replaced *western and adventure fiction* (category N) with *martial arts fiction*. There are three reasons for this decision. Firstly, there is simply no western fiction in China; secondly, martial arts fiction is broadly a type of adventure fiction and it is a very popular and important fiction type in China and hence should be represented; thirdly, the language used in martial arts fiction is a distinctive language type and hence once more one would wish to sample it. Most stories of this type, even though they were published recently, are under the influence of vernacular Chinese, i.e. modern Chinese styled to appear like classical Chinese. While the inclusion of this text type has made the tasks of POS (part-of-speech) tagging and post-editing more difficult, it may also make it possible to compare representations of vernacular Chinese and modern Chinese.

The second variation was caused by problems we encountered while trying to keep to the FLOB sampling period. Considering the availability of texts of some categories (notably F, D, E, and R), we decided to modify the FLOB sampling period slightly by also including some samples for ±2 years of 1991 when there were not enough samples readily available for 1991. We assume that varying the sampling

---

[5] LCMC is created as part of our research project "Contrasting tense and aspect in English and Chinese" funded by the UK Economic and Social Research Council (Grant Ref. RES-000-220135). The corpus is distributed free of charge for use in non-profit-making research. The manual accompanying the corpus, as well as the details for ordering, can be accessed online at the corpus website http://www.ling.lancs.ac.uk/corplang/lcmc or its Chinese mirror site in China at http://www.cass.net.cn/chinese/s18_yys/dangdai/LCMC/LCMC.htm.

frame in this way will not influence the language represented in the corpus significantly.

The LCMC corpus has been constructed using written Mandarin Chinese texts published in Mainland China to ensure some degree of textual homogeneity. It should be noted that the corpus is composed of written textual data only, with items such as graphics and tables in the original texts replaced by gap elements in the corpus texts. Long citations from translated texts or texts produced outside the sampling period were also replaced by gap elements so that the effect of translationese could be excluded (McEnery & Xiao forthcoming) and L1 quality guaranteed. LCMC became available in the later stages of this research, and hence is used only in chapter 6 of this book.

The English-Chinese Parallel Corpus (ECPC). In addition to contrasting aspect marking in the two languages, chapter 6 also explores how aspectual meanings in English are translated into Chinese. For this purpose, we built an English-Chinese parallel corpus. A parallel corpus can be defined as a corpus that contains source texts and their translations. Corpora of this type are particularly useful for translation studies (see McEnery & Xiao forthcoming). Parallel corpora can be uni-directional (e.g. from English into Chinese or from Chinese into English alone) or bi-directional (e.g. containing both English source texts with their Chinese translations as well as Chinese source texts with their English translations). As in this book we are interested in Chinese expressions of translated aspectual meanings from English, we are using a unidirectional parallel corpus for our research where English is the source language and Chinese is the target language. The corpus is composed of bilingual texts taken from *English World*, a web-based journal published in China.[6] The sampling period is between October 2000 and February 2001, during which 121,493 English words and their translation in the form of 135,493 Chinese words were gathered.

We have now established what we will be studying, outlined why we wish to study it and described the methodology and data we will use in our research. Yet one important question remains to be answered – what theoretical model guides our approach to aspect?

1.4. The theoretical framework and an overview of this book

The basic theoretical framework used in this book is the two-component aspect model proposed by Smith (1991, 1997). According to this theory, aspect is compositional in nature. The aspectual meaning of a sentence is the synthetic result of 'situation aspect' and 'viewpoint aspect' (i.e. grammatical aspect). The former refers to the intrinsic aspectual properties of idealised situations while the latter refers to the speaker's choice of a perspective from which a situation is presented. The two are independent yet interacting components of aspect. It will be argued in this book, however, that Smith's theory while useful, contains some flaws as it stands (see section 2.2). It needs to be modified significantly in order to model aspect in Chinese. In developing our model, it must be noted that on the one hand, aspect consists of situation aspect at the semantic level and viewpoint aspect at the grammatical level (see section 2.5); on the other hand, situation aspect is modelled as 'verb classes' at the lexical level (see section 3.3) and as 'situation types' at the sentential level (see

---

[6] The web-based journal can be accessed at http://www.bentium.net/cgi-bin/getlsts?listname=enwd. The corpus is a component of the Babel English-Chinese Parallel Corpus, which is available online at http://www.ling.lancs.ac.uk/corplang/babel/babel.htm.

section 3.5), with the latter being the composite result of the interaction between verb classes and their complements, arguments and non-arguments such as peripheral adjuncts and viewpoint aspect at the 'nucleus', 'core' and 'clause' levels (see section 3.4).
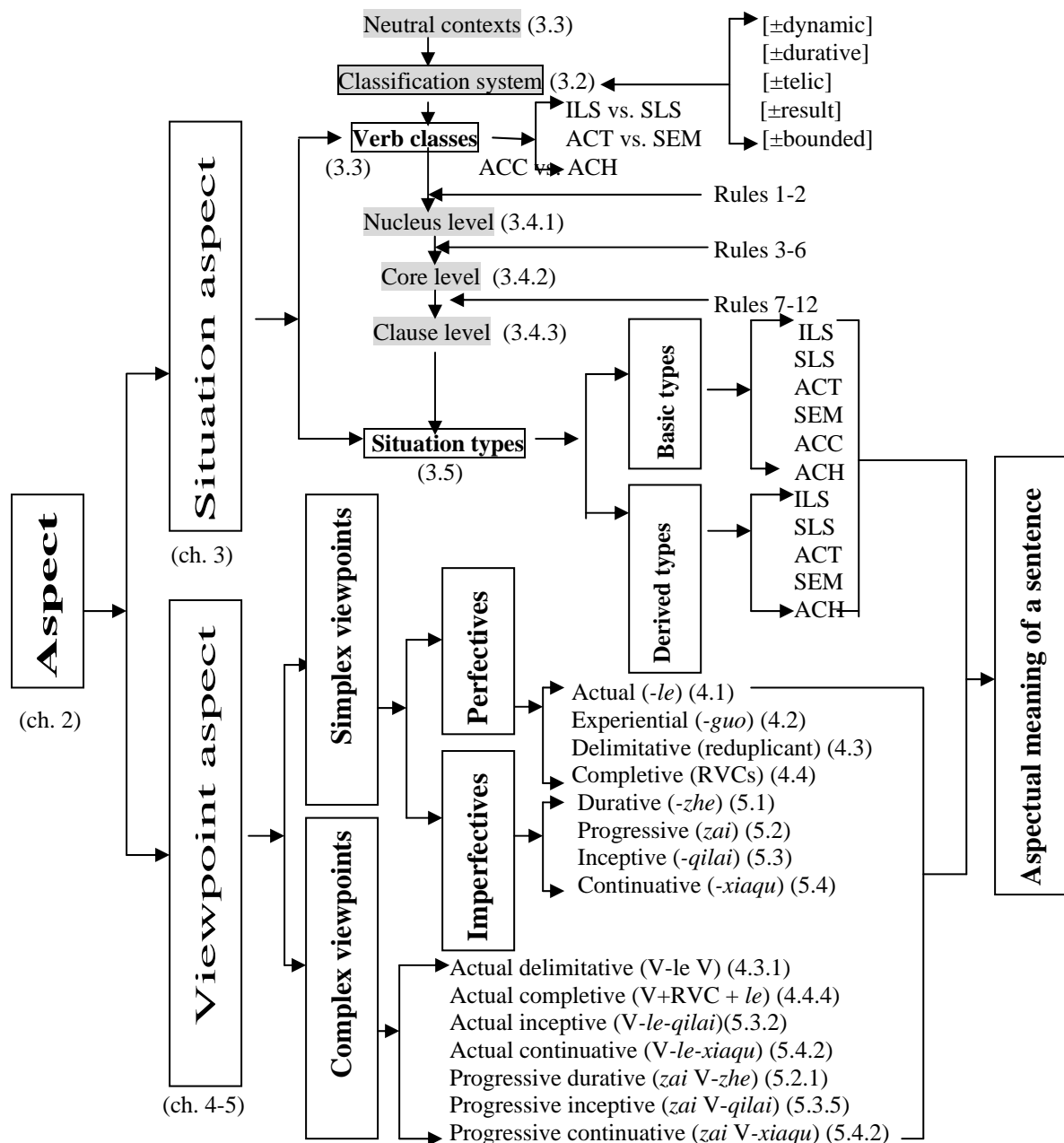
Figure 1.1 The two-component model of aspect in Mandarin Chinese

Figure 1.1 is an overview of our model of aspect in Chinese. The numbers in the brackets indicate the chapter or section numbers for the corresponding topics. In addition to chapters 2-5 outlined in the figure, we will use three comparable L1 language corpora and an English-Chinese parallel corpus, in chapter 6, to contrast aspect marking in Chinese and British English and American English, and to explore

how aspectual and temporal meanings in English are translated into Chinese. It is this complexity that is modelled in this book.

At this point, the model itself clearly has not been fully justified and presented. Consequently this figure should be viewed as a useful reference point for readers seeking to understand a specific element of the work presented in this book in the overall context of this book. However, with the overall model outlined, we can now proceed to outline the goals we have in presenting this research.

In terms of goals, our most important goal is to refine and expand Smith's model of aspect based upon the corpus-based research we have undertaken to provide an explanatorily adequate account of aspect. Yet we wish to base this explanatory account on a descriptively adequate account of aspect in Chinese. Our work presents a new, corpus-based, description of aspect in Chinese to replace the numerous, partial, published accounts. This new account of Chinese aspect will argue, on the basis of corpus evidence, that even in the case of aspect markers as widely studied as *-le*, important aspects of their meaning/use have been overlooked. Omissions such as this, we believe, are unacceptable. Before we proceed to present our account of aspect in Chinese, however, it is appropriate to refine our definition of aspect, which will be done in chapter 2.