1. Background

Time is a basic concept that exists independently of human language. Temporal information is encoded in human languages by two related, yet distinct, linguistic categories: tense and aspect. While tense and aspect are both temporal notions, they are different in nature. While tense and aspect are both temporal notions, they are different in nature. Tense is deictic in that it indicates the temporal location of a situation, i.e., its occurrence in relation to a specific reference time. Aspect, in contrast, 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. While, as noted, time is a cognitive concept based on common human experience, expressions of temporal information may vary across languages. For example, while Romance languages like French and Spanish focus on when an action happens rather than its temporal distance and length, the plotting of action, so important in tense languages, is not important in an aspect language like Chinese, which is concerned with whether the action is completed or not, whether the action is in progress or not.

Aspect is related to the temporal properties of linguistically described situations in the world (i.e. situation aspect) and how these situations are presented (i.e. viewpoint aspect). As the temporal notion denoted by aspect is essential to human languages, aspect has long been the subject of study by both semanticists and grammarians. The cross-linguistic contrast of aspect, however, has only become possible with the development of modern aspect theory in the last decade, notably the two-component aspect theory proposed by Smith (1997). Contrastive studies of features such as aspect in different languages have proved useful for typological research as well as for pedagogical purposes. Previous studies, however, have been largely been limited to closely related languages such as English and Portuguese or conducted without recourse to corpus data. As English and Chinese are genetically distinct languages that differ considerably, the corpus-based contrastive study of aspect and tense for this language pair represents a worthwhile research challenge.

2. Objectives

The project had the following aims and objectives: 1) to construct a one-million-word corpus of written Chinese produced by native speakers (an L1 corpus); 2) to examine, on the basis of L1 corpora, how aspectual and temporal meanings are conveyed in English and Chinese; 3) to identify the translation patterns of English aspect into Chinese; 4) to use corpora to examine the interaction of situation aspect and viewpoint aspect in the two typologically different languages; 5) to verify, and if necessary, to extend the aspect model presented Xiao and McEnery (2002). All of these aims and objectives have been met by the project. We will outline below how each of these aims and objectives is achieved. See sections 4 and 6 of this report for details of the results and outputs of our research.

The Lancaster Corpus of Mandarin Chinese was created to fulfil aim 1. The corpus, which is distributed by the ELRA, has been announced in several mailshots and has been made publicly available for use in non-profit making research. As of 08/04/2004, over 100 institutions hold a licence. A Chinese mirror site of the corpus has also been established and hosted by the Chinese Academy of Sciences, where instructions on downloading and using the corpus are given in Chinese. The corpus was publicised at the 4th Workshop on Asian Language Resources and a demonstration of the use of the corpus will be given at LREC 2004. The study of the temporal and aspect system, as well as the interaction between situation aspect and viewpoint aspect, in Mandarin Chinese is to be published in Aspect in Chinese (Benjamins), where the expression of English aspectual meaning is detailed in one chapter. The translation patterns of English aspect into Chinese were explored in our paper ‘Aspects of translation in Chinese and English: A corpus based approach to the translation of aspect’. A cross-linguistic contrast of aspect marking in Chinese, British and American English was published in ‘Aspect marking
in English and Chinese: using the Lancaster Corpus of Mandarin Chinese for contrastive language study’ in
*Literary and Linguistic Computing* 18/4. The modified two-level aspect model of situation aspect is to be

3. Methods

While linguistic analysis typically benefits from being based on attested language use, previous studies of aspect
have largely been conducted without recourse to such data. They have, rather, been based on a handful of
confected examples which, if not intuitively unacceptible, are atypical of attested language use. Furthermore,
those proposals have not, to date, been tested on corpus data. Yet corpora have a role to play both in developing
and testing such theories.

With that said, we do not mean that the corpus-based approach and the intuition-based approach are
antagonistic, rather the two are complementary. With the intuition-based approach, researchers can invent purer
eamples instantly for analysis, because intuitions are readily available and invented examples are free from the
language-external influences existing in naturally occurring language. However, intuition should be applied with
cautio. Firstly, it is possible to be influenced by one’s dialect or sociolect. As such, what appears unacceptible to
one speaker may be perfectly felicitous to another. Secondly, when a researcher invents an example to support or
disprove an argument, they are consciously monitoring his/her language production. Therefore, even if their
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 perceive and a corpus can yield reliable quantitative data. Hence
we decided to undertake a corpus-based approach to modelling situation aspect.

The use of corpus data as an input to the semantic analysis of aspect, a methodology used on our project,
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.

The English corpora used on this project are FLOB and FROWN, which represent modern British and
American English respectively. As there was no publicly available Chinese corpus which was suitable for the
contrastive study of English and Chinese, we created the Lancaster Corpus of Mandarin Chinese, a one-million-
word Chinese match for the FLOB/FROWN corpora. These three corpora are the major source of evidence in our
research.

4. Results

The results of our research cover six areas: 1) Corpus building, 2) translation patterns of English aspect/tense in
Chinese, 3) modelling situation aspect, 4) the Chinese aspect system, 5) interaction between situation aspect and
viewpoint aspect, and 6) contrasting aspect marking in English and Chinese. Below is a summary of our research
findings.

1) Corpus building

The Lancaster Corpus of Mandarin Chinese (LCMC) is designed as a Chinese match for the FLOB and
FROWN corpora for modern British and American English. The corpus is suitable for use in both monolingual
research into modern Mandarin Chinese and, when used in conjunction with its English match FLOB/FROWN,
for the cross-linguistic contrast of Chinese and British/American English. The corpus sampled 15 written text
categories including news, literary texts, academic prose and official documents etc published in China in the
earlier 1990s, using the same sampling frame as FLOB/FROWN. The corpus, together with the web-based
exploration tool WebConc, has been released for use in non-profit-making linguistic and educational R&D. It is
currently licenced to 55 academic institutions and over 40 private and non-academic users.

2) Chinese translation patterns of English aspect/tense

English has four viewpoint aspects: the *progressive*, the *perfect*, the *perfect progressive* and the *simple.*
Our data show that around 88% of tensed verbs take the simple aspect. The perfect accounts for roughly 8% while
the progressive takes up 4%. The perfect progressive is extremely rare (less than 0.5%). Although both English and Chinese have a progressive aspect, the progressive has different scopes of meanings in the two languages. About 58% of situations referred to by the English progressive take the progressive or the durative aspect, marked either overtly or covertly in Chinese translations. The interaction between situation aspect and viewpoint aspect also influences the translator’s choice of viewpoint aspect. Most telic situations (around 65%) and situations incompatible with progressiveness (92.5% of individual-level states and 75.9% of achievements) are more likely to undergo viewpoint aspect shift and be presented perfectly. In contrast, atelic situations are normally translated with the progressive or the durative aspect (80% of stage-level states and 87.2% of activities).

There are 4 types of perfect in English. The perfect of result, accounting for 59% of all instances of perfect, goes only with telic situations. This type of perfect is most frequently (about 70% of the time) translated with the actual or the completive viewpoint in Chinese. The English perfect of experience and the experiential aspect in Chinese share the same meaning. Translations of this type of perfect are therefore rather straightforward. The perfect of persistent situation is typical of English and imperfective by nature. Translations of this type of perfect mainly take the LVM (lack-viewpoint-marker) form, though there are variations because of the interaction between situation aspect and viewpoint aspect. Situations referred to by the English perfect of recent past are normally translated with the actual aspect in Chinese, because all situations in the recent past must be actualized and the focus is on their temporal nearness. The pluperfect (past perfect) refers to the “past in the past”. As a situation with a past time reference must have been actualized, completed or mentally experienced, the pluperfect is most likely (90% of the time) to be presented with the actual, the experiential or the completive aspect in Chinese translations. When situations referred to by the English future perfect are translated into Chinese, the perfect meanings are lost. Such situations are normally presented perfectly with a future time reference.

The perfect progressive is an interaction between the perfect and the progressive. Chinese translations of the perfect progressive may shift towards the progressive or the perfect meaning, depending on the situation type involved and the translator’s choice of viewpoint. But in most cases both perfect and progressive meanings can be retained, with the perfect being lexicalized by temporal adverbs such as "yi zhi" “all the time” while the progressive is either signaled by the progressive aspect marker "zai" or it is implied by context. The pluperfect progressive is similar to the perfect progressive with the exception that it signals progressiveness with a relative past time reference. Situations referred to by the English pluperfect can be translated into Chinese with the progressive or the durative aspect unless the translator chooses to present them perfectly or there is a shift in situation type which prohibits them taking the progressive or the durative aspect.

The simple aspect in English presents a situation without aspectual modification. The simple past refers to a single situation completed or terminated in the past or a past habitual situation. More than 80% of the situations referred to by the English simple past are presented with a perfective viewpoint in Chinese translations. Past habitual situations referred to by the English simple past are not marked in Chinese translations. The simple present in English normally refers to states, namely, statements made “for all time” and habitual situations, though in highly specific circumstances, it can be used to mark a future or past situation or a situation in progress. When situations referred to by the English simple present are translated into Chinese, they are most likely (over 80% of the time) to take the LVM form. Specifically, timeless situations are more likely to take the LVM form than other situations, and [-telic] situations are more likely to take the LVM form than [+telic] situations. The simple future refers to situations with a future time reference. When these situations are translated into Chinese, over three quarters of them take either adverbs indicating futurity (e.g., jiang “will, be going to” and/or modals (e.g., hui “be going to”), while around one quarter take the LVM form.

3) Modelling of situation aspect

On this project we extended Smith’s (1997) two-component aspect theory by developing a two-level model of situation aspect in which situation aspect is modelled as verb classes at the lexical level and as situation types at the sentential level. In our model, we developed a 5-way classification system for situation aspect, which
includes [±dynamic], [±durative], [±bounded], [±telic] and [±result]. At the lexical level, verbs are classified into six verb classes: individual-level state, stage-level state, activity, semelfactive, accomplishment and achievement using our classification system. Situation types are the composite result of the rule-based interaction between verb classes and complements, arguments, peripheral adjuncts and viewpoint aspect at the nucleus, core and clause levels. We identified 12 rules for this interaction on the basis of English and Chinese corpus data. At the sentence level, there are six basic situation types which share the same name and features as lexical verbs. Except for accomplishments, all of the others have various derived situation types which vary from their basic types with respect to their durativity or boundedness value.

With a framework consisting of a lexicon, a layered clause structure and a set of rules mapping verb classes onto situation types, the model is developed and tested using English and Chinese corpora. The corpus-based two-level model provides a more refined aspectual classification and gives a more systematic account of the compositional nature of situation aspect than previous models. Mourelatos’ (1981: 199) criticism of Vendler and Kenny also applies to all of the models reviewed in our research, though it should be noted that Mourelatos himself does not provide an explanation of the ‘determinants’ of situation aspect. Vendler (1967) is confined to the lexical level whereas Verkuyl (1993) works only at the core level. While Smith and Shirai (1991) utilize a classification system similar to ours, they do not differentiate between the lexical and sentential levels of situation aspect. In fact, with their classification systems (note the difference between Shirai’s definition of result and ours), it is quite impossible to treat verb classes and situation types separately. However, the distinction between their one-level approach and our two-level approach is more than cosmetic as a two-level approach can model situation aspect in a more structured way and provide a clearer account of the composition of situation aspect. The dichotomous treatment of states in our model also gives a unified explanation of the felicitous co-occurrence of some states with the progressive, which is absent in previous proposals. Last but not least, our model represents an innovative attempt to marry a corpus based approach and a theory-driven approach to aspect.

4) The Chinese aspect system

Aspect in Chinese consists of two components: situation aspect and viewpoint aspect. Situation aspect is universal cross-linguistically. As such, the two-level model of situation aspect outlined above also applies to Chinese. In addition to situation aspect, which is inherent in linguistic expressions of situations in human languages, our research also identifies, on the basis of corpus data, four perfective and four imperfective viewpoints in Chinese. Perfective aspects in Chinese include the actual aspect marked by -le, the experiential aspect marked by -guo, the delimitative aspect marked by verb reduplication and the complete aspect marked by RVCs (resultative verb complements). While these four viewpoints all present situations perfectly, they have different focuses. Specifically, -le focuses on the actuality of a situation, -guo on its experientiality, verb reduplication on its delimitativeness and RVCs on its completivity. These viewpoints also interact with situation aspect in different ways. Imperfective aspects in Chinese include the durative aspect marked by -che, the progressive aspect marked by -zai, the inceptive aspect marked by -qilai and the successive aspect marked by -xiaqu. Although these viewpoints all present a situation as imperfective, each of them has its own focus. Specifically, -che focuses on the durativeness of a situation; -zai focuses on its progressiveness; -qilai focuses on its inceptiveness; and -xiaqu focuses on its successiveness. These viewpoints also behave differently in respect to their interactions with situation aspect.

While the aspect markers listed above typically signal individual viewpoint aspects, we found that in Chinese discourse, covert aspect marking, namely, the utterances which take the LVM form, is a frequent and important strategy to express aspectual meanings. The LVM form typically occurs in three situations. Static situations normally take the LVM form because statives do not have to be marked aspectually. For dynamic situations, there are two types of LVM sentences. They are either irrealis imperfectives or they are perfectives without overt aspect marking. We assume that sentences without overt aspect marking have the ‘zero aspect’, because they convey aspectual meanings but take the zero form. When taken in isolation, these LVM sentences
may be ambiguous between perfective and imperfective readings; but in discourse, their aspectual meanings are made explicit by context. As such, the zero aspect can be either perfective or imperfective, depending on context. It is not an independent viewpoint aspect.

Our research is the first corpus-based study of aspect in Chinese. While some of the viewpoint aspects have already been identified in previous studies of aspect in Chinese, our research has corrected many intuition-based misconceptions and associated misleading conclusions readily found in the literature. Some viewpoints, e.g. the complete aspect marked by resultative verb complements, have for the first time been considered as independent viewpoints based on their behaviours in attested language use.

5) Interaction between situation aspect and viewpoint aspect

As different languages may have different viewpoint aspects, and even the same viewpoint may vary in the range of its use in different languages (e.g. see the discussion below on the English progressive and the Chinese progressive marker *zai*), the interaction between situation aspect and viewpoint aspect shows different behaviours cross-linguistically.

In Chinese, the actual viewpoint interacts with all situation types, though there is a strong tendency for -le to co-occur with spatially or temporally bounded situations. With unbounded states, -le demonstrates the feature of ingressive dynamicity and coerces these situations into derived activities at the clause level. As a perfective marker, -le only indicates the actualization and focuses on the totality of a situation but does not provide any final endpoint. The experiential aspect interacts with situations of any type, irrespective of their values of dynamicity, telicity or boundedness, though its distribution pattern is unbalanced, reflecting the unbalanced distribution of situation types. The delimitative aspect can only interact with dynamic situations encoding no result. As resultative verb complements can be affixed to almost all verbs classes other than achievement verbs, and the resulting compound verbs are all turned into derived achievements encoding a result, the complete aspect only interacts with the situation type of achievement. The durative aspect is sensitive to the [+durative] and [±result] values of a situation. Basically, it is only compatible with [+durative] and [±result] situations. The progressive aspect is sensitive to the aspectual feature of dynamicity, hence it is only compatible with non-stative situations. The inceptive aspect is sensitive to the aspectual feature of durativity. That is, -qilai can only be affixed to durative situations felicitously. Like the inceptive aspect, -xiaqu is sensitive to the feature of durativity. This means that -xiaqu can only be affixed to durative verbs, though semelfactives can also take -xiaqu when they denote multiple events.

In English, viewpoint aspects include three simplex viewpoints (the progressive, perfect, and simple aspects) and one complex viewpoint (the perfect progressive). As English tense and viewpoint aspect are often combined morphologically, the interaction between viewpoint aspect and situation aspect is not subject to constraints as rigid as those present in Chinese. While the progressive and perfect progressive aspects also tend to interact with [+dynamic] situations, the simple and perfect aspects interact with all situation types.

6) Contrasting aspect marking in English and Chinese

Our contrastive study of aspect marking in English and Chinese shows that while Chinese and English are typologically different, aspect markers in the two languages show a strikingly similar distribution pattern, especially across the two broad categories of narrative and expository texts. It is also interesting to note that while British English and American English have developed variations in spelling, word choice, and grammar, their use of aspect is strikingly similar – the curves for the distribution of aspect markers for FLOB and Frown are almost identical to each other.
On the other hand, the study also reveals some important differences in the distribution of aspect markers in Chinese vs. English and British vs. American English across fifteen text categories. In expository texts, perfective aspect markers in Chinese generally occur more frequently than those in English, whereas in narrative texts perfective markers in English are generally more frequent than those in Chinese. The relatively higher frequency of perfective markers in narrative texts and their lower frequency in expository texts in English can be accounted for by the fact that English aspect markers express both temporal and aspectual meanings. Narrative texts are normally related to what happened in the past whereas expository texts are typically non-past. Hence the relatively higher frequency of perfective markers in narrative as opposed to expository texts is understandable. In marked contrast, imperfective aspect markers show a totally different distribution pattern from perfective markers. In expository texts, imperfective markers in English typically occur more frequently than those in Chinese whereas in narrative texts, imperfective markers in Chinese are generally more frequent than those in English. This phenomenon can be explained as follows. First, the Chinese progressive marked by *zai* can only signal progressiveness literally. In contrast, ‘the progressive in English has a number of other specific uses that do not seem to fit under the general definition of progressiveness’ (Comrie 1976:37). While the different uses of the progressive in English and Chinese account for the slightly higher frequency of the English imperfective markers in expository texts, it cannot explain the relatively lower frequency of these markers in narrative texts. Nevertheless, we can find an answer in the Chinese imperfective marker *-zhe*. This marker has three basic functions: to signal the durative nature of a situation, to serve with a verb as an adverbial modifier to provide background information and to occur in locative inversion to indicate existential status. Of the three functions, only the first is used in expository texts. In contrast, all three functions of *-she* apply to narrative texts. Furthermore, in addition to inducing a background effect, *-she* can also be used in an apparently ‘foregrounded’ situation to move narration forward. As such, it is hardly surprising that Chinese imperfective markers occur more frequently in narrative texts than English imperfective markers. British English and American English also differ in that the latter variety does not show such a marked fluctuation in aspect marking in narrative texts, notably in biography and the five types of fiction.

5. Activities

The activities of the project include the presentation of papers at three conferences and networking with Oxford University Computing Services in testing the XML-aware version of SARA (i.e. XAIRA).

At the international conference on German Corpus Linguistics: synchronic, diachronic and contrastive, we presented our paper ‘Aspects of translation in Chinese and English: A corpus based approach to the translation of aspect’. At the 4th Workshop on Asian Language Resources, we presented our paper ‘Developing Asian language corpora: standards and practice’. The paper proposes standards for developing Asian language corpora so as to facilitate international data exchange. We also presented two corpora of Asian languages developed at
Lancaster University, the LCMC and EMILLE corpora and demonstrated how to explore these corpora using Xara and other Unicode-compliant XML-aware corpus tools. At LREC 2004, we will our paper ‘The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study’. The poster/interactive demonstration session shows users how to explore LCMC for monolingual research into Chinese and for contrastive research.

In addition to the three conferences, we have also networked with the OUCS in beta-testing the new Xara suite on Chinese and 14 South Asian languages. This activity has benefited both parties as it improved the quality of our corpora (i.e. XML and CES-compliant structures) and identified and solved a wide range of problems of the software encountered with Asian languages.

6. Outputs
The major outputs of our project include:

A. Journal papers

B. Books and book chapters

C. Refereed conference papers

D. Machine-readable resources
Release notes for the Lancaster Corpus of Mandarin Chinese:
1. June 2003: Corpus mounted for free downloading on the website of Corpus-based Language Studies (http://www.ling.lancs.ac.uk/corplang/lcmc) hosted by the Linguistics Department, Lancaster University and announced at the UCREL website;
2. August 2003: Chinese mirror site for the corpus established and hosted by the Chinese Academy of Sciences, Beijing (http://www.cass.net.cn/chinese/s18_yys/dangdai/LCMC/LCMC.htm);
3. December 2003: Corpus release announced at CORPORA-list, ELSNET-list and CLUK-list;
4. May 2004. Corpus taken over by ELRA/OTA.

E. Software
We have also developed a web-based XML-aware, Unicode-compliant concordancer for the LCMC corpus (WebConc) to help users to explore this corpus (see the corpus website).

7. Impacts
The potential longer-term impacts of the research undertaken on this project include contributions to the following areas of studies:

1. Methodology: previous research on aspect has largely been conducted without recourse to attested language data. Our work takes a corpus-based approach to the study of aspect.
2. Aspect theory: the corpus-based aspect model proposed on this project should advance research in aspect theory significantly. It promises to provide a more refined classification of situation aspect and a better account of the interaction between situation aspect and viewpoint aspect.
3. Contrastive linguistics: our research has made it clear how similar (or different) the mechanisms are that English and Chinese uses to express aspectual and temporal meanings.

4. Linguistic typology: the study of aspect and tense in English and Chinese will shed new light on the typological studies of Asian languages and European languages.

5. The LCMC corpus has been used widely in language study and language engineering. As of 08/04/2004, we have over 100 licencees for the corpus.

8. Future Research Priorities

We found on this project that many grammatical categories in English and Chinese are potentially related to aspect. They typically encompass the following: predicates; arguments; determiners and classifiers; adverbials; time words, localisers and prepositions; negation; ba, bei, de and shi structures and series verb constructions; aspect markers and temporally/aspectually related particles. We also found that while the semantic functions of these categories in the two languages are quite similar, they differ considerably morphologically and/or syntactically. For example, while arguments in both English and Chinese play an important role in determining the telecity value of a situation, nouns in Chinese do not inflect themselves to reflect the singular/plural distinction as happens in English. While Chinese has aspect markers signaling different viewpoint aspects, English combines tense and aspect markers morphologically/syntactically. These findings encourage us to further investigate these grammatical categories with the aim of uncovering the potential similarities and differences between the two languages.

The compositional nature of aspect is well recognized. Indeed, the effects of verbs and their arguments on aspect have been intensively studied on this project. Nevertheless, the potential contributions of other grammatical categories such as negation to aspect remain largely unexplored. The priority for future research is identifying grammatical categories that may potentially contribute to aspectual meaning. The categories listed above, while not comprehensive, form the core of a grammar. As tense/aspect is a central distinction between English and Chinese, this list reflects important differences between the two languages. Hence, a further proposal, which forms an extension of this project, is to contrast English and Chinese by comparing aspect-related grammatical categories.

This future research will have an aspectual focus and a contrastive focus on the grammars of the two languages. While research on aspect has a contrastive focus, as demonstrated by works on aspect by Comrie (1976), Dahl (1985; 1999), Bybee et al (1994), Smith (1997) and Miller (1999), a contrastive study is also supposed to have an aspectual focus, because aspect is an important grammatical category. In Brown & Miller (1999), for example, aspect takes up a lion’s share of the work presented. While some languages may not have tense (e.g. Chinese), aspect appears to have been found in all human languages investigated so far (cf. Dahl 1985). The convergence between aspect research and cross-linguistic contrast is a natural tendency as while aspect is a common grammatical category (cf. Miller 1999:42), languages may express aspectual meanings in different ways (cf. Lehmann 1999:48). For example, English uses morphologically combined tense/aspect markers whereas Chinese uses aspect markers to express aspectual meaning. Yet subtle distinctions occur: while English and Chinese both have some grammatical aspects, e.g. the progressive, they differ in their use of the progressive. Differences such as these are useful in accounting for the phenomena observed in attested language use.

The future research outlined here has already been funded (ESRC RES-000-23-0553) and the results of the project will be published by November 2007.