CHAPTER TWO

THEORETICAL FRAMEWORK

In the literature on Lexicase, it is always emphasised that grammar is generative and constrained. The term generative, according to Noam Chomsky, means that grammar is "a formal and explicit description of the internalised system which accounts for a speaker's ability to recognise well-formed sentences in his language" (Starosta, 1995), and is therefore, applicable to all human languages. A 'constrained' grammar means that, the grammar will not be 'too powerful'. A theory that is too powerful claims that all things are possible and will fail to exclude precisely what is ungrammatical in a given sentence. Conversely, one that is constrained will be able to make a testable claim or be disproved if it makes a wrong claim. For a grammar to be effective, the two elements, that is, 'generative' and 'constrained' are essential.

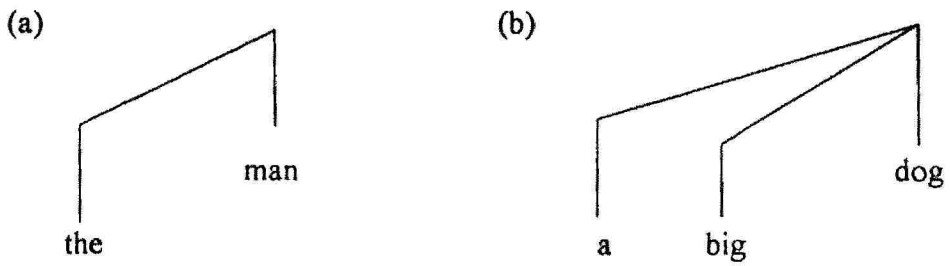
The present chapter will begin with a detailed description of the framework of Lexicase in Section 2.1, with an emphasis on the properties of the theory. Section 2.2 shows how the theory is applied in the present study of tense and aspect in Chinese, and provides a description of the subcategorisation of tense/aspect markers. Concluding remarks will be presented in Section 2.3.

2.1 Theoretical Framework

2.1.1 Lexicase Grammar

Lexicase, including its latest development in morphology, Seamless Morphology, is a theory of syntax developed over the last three decades by Professor Stanley Starosta of the University of Hawai'i. The theory is a hybrid of the European-style of dependency grammar and the American tradition of case in linguistics. According to Starosta (1994), the *Lexi-* part of the name of the theory shows that this is a kind of "word grammar" and it stands for "a strategy that captures all linguistic generalisations as statements about the internal structure and external distribution of words", while *–case* indicates "the syntactic-semantic relations that obtain between nouns and the words on which they depend, and the morphological and syntactic mechanisms that overtly mark the presence of these relations". The former is a characteristic of dependency grammar and the latter is the significant property of case grammar.

Lexicase is simple and straightforward. In the grammar, a sentence or phrase is a list of words connected by lexically specified links. Being a dependency grammar, it characterises the sisterhood-like relationship between a pair of words in terms of a regent and a dependent. The regent is the one in the pair of words that governs the other; and the dependent is the word that depends on its regent. The notational conventions of regent-dependent relationship can be seen graphically in a stemma, a tree-like graphic representation of dependency relations in Lexicase as in Figure 2.1 below:



Dependency Links in Lexicase

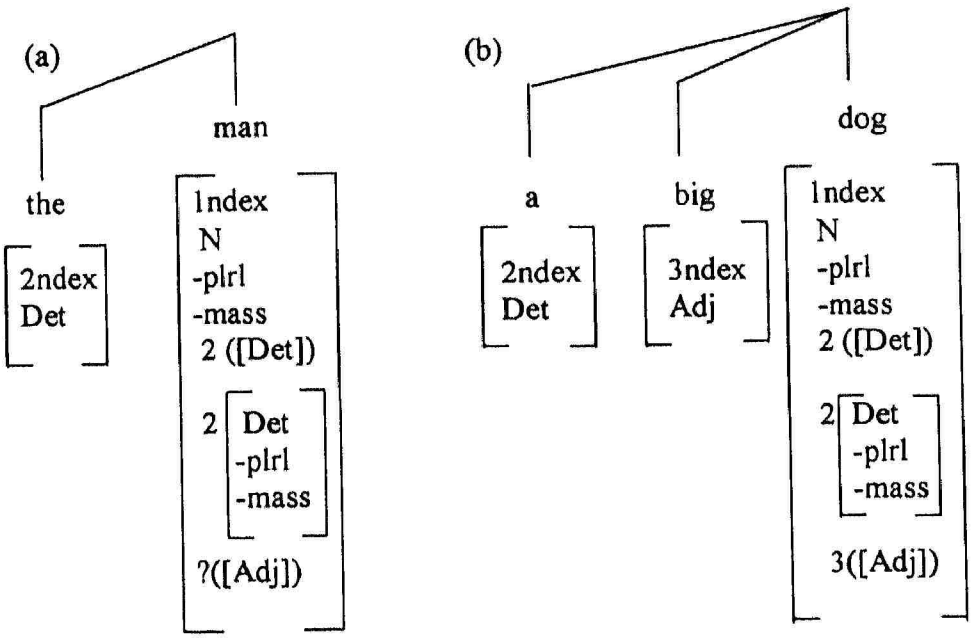
Figure 2.1

In (a) above, the two words are placed graphically at different heights, with a vertical 'mast' each above the words. *Man*, the regent of the NP is placed at a higher position in the stemma, while its dependent *the* is at a lower position. The slanting line that links the tops of the two mast shows that *man* is the head that controls its dependent *the*. In (b), the regent, *dog*, has two dependents, *a* and *big*. Both the dependents are related to *dog* individually but there is no local relationship that exists between them. That is, neither of them is dependent on the other. Since *dog* does not depend on the other two words in the stemma, it is the head of the NP *a big dog*.

On top of the tree-liked notational representation, the syntactic relationship between the regent and its dependent(s) is shown in the contextual and non-contextual features marked in the lexical matrices under the individual entry in a stemma. A contextual feature is a feature that is enclosed in [ ] or ( ), indicating the dependency of a dependent on its regent. The square bracket [ ] means that the feature in the bracket is a complement and it is grammatically required by the regent, compared to the feature in the round bracket ( ) which indicates that the element is an adjunct that is grammatically optional, unless the element is a skeletal feature that is required for the

syntactic analysis of a sentence.[1] The skeletal features are basic features that establish the primary external dependency links in the structure by various rules in the grammar, such as the Linking Rules and Chaining Rules in the grammar.[2]

Non-contextual features are 'inherent' properties of a lexical item. It could be a part of speech such as Det (determiner), N (noun) and Adj (adjective), or any morphosyntactic semantic feature, such as number, gender and agreement in grammar.



plrl : plural[3]

Figure 2.2

[1] In the matrix of the regent in a stemma in Lexicase, there is one or more 'skeletal contextual feature(s)' that is categorised by the dependents, such as the 2([Det]) in Figure 2.2.

[2]We will not go into the theoretical part of how the Rules are applied in this dissertation. Details on such rules are available in Starosta (1988, 1995, On-line Reference Manual) and many other Lexicase writings.

[3] With few exceptions, it is the convention in Lexicase to abbreviate a word by taking the first four consonants of the word. 'Plural' therefore becomes plrl.

Figures 2.2 (a) and (b) above show how the contextual and non-textual features help to establish the grammatical relationship between a regent and its dependent(s) in Lexicase stemmas. Note that in the stemma, each word is given an index based on its linear position, as references for analysis. Except for 1ndex (index 1) that is always used to mark the head of the stemma, the remaining index may be assigned arbitrarily as long as it is clearly indicated.

The dependency link between words is formalized by copying the linear index of the dependent into a valency feature of the regent in Lexicase. In Figure (a) above, the regent *man*, a singular count N (Noun) which is –plrl (non-plural) and –mass (non-mass), has a feature ?([Det]) in its matrix. It means that it needs a determiner as an obligatory dependent. The dependent *the* is labeled as 2ndex (index 2) in the stemma. Hence, the index of *the*, which is '2', is copied into the valency feature of its regent, forming the feature 2([Det]). 2([Det]) is the skeletal feature of the regent *man*, The [ ] symbol shows that the dependent is obligatory.[4]

The contextual feature 2[Det, -plrl, -mass] means, whenever it occurs, a Det must co-occur with a complement that is a non-mass non-plural noun. Again, the requirement is satisfied by 2ndex.[5] The ?([Adj]) is a feature that is unindexed. It

[4] The ( ) indicates, on the other hand, that the noun may or may not have a dependent Det (Determiner) that is labeled as 2ndex (index 2) in the Figure. The symbol seems to contradict with what we have just said about [ ] in this paragraph. In fact, it is not contradictory. The ( ) is meant to cover cases in human languages such as English in which a mass noun may not take a Determiner. Since Det is a skeletal feature in the tree, by definition, the bracket remains although it is not relevant in this particular Figure. The same reason holds for the ( ) in ?([Adj]) in both Figures (a) and (b).

[5] By theory, the index 2 in 2([Det, -plrl, -mass]) here is deduced by linking that feature to the skeletal feature(s) by the Linking Rules in the grammar. In simple terms, it means that since ?([det]) has an index 2ndex, the same index number is copied or linked to ?[Det, -prlr, -mass], forming 2[Det, -prlr, -mass]. It shows that the latter also refers to the same lexical entry *the*. As the indexes used in a stemma are quite transparent and easy to interpret, we will not go into the theoretical part of it.

indicates that the N allows an Adj (adjective) as dependent, but it is not present in this example. Unlike (b), the ?([Adj]) is satisfied with the presence of the adjective *big* in the matrix for *dog* as 3([Adj]).

The affiliation between the regent and the dependent is constant throughout all sentence structures in the Lexicase. In a longer phrase or clause or a sentence, the same principle is applied to indicate how a word is attached to another word in dependency. Figure 2.3 is an example of the stemma notation of a simple sentence. This stemma is taken from Starosta (1994:2168) but modified according to the latest development in the grammar in the late 1990s.

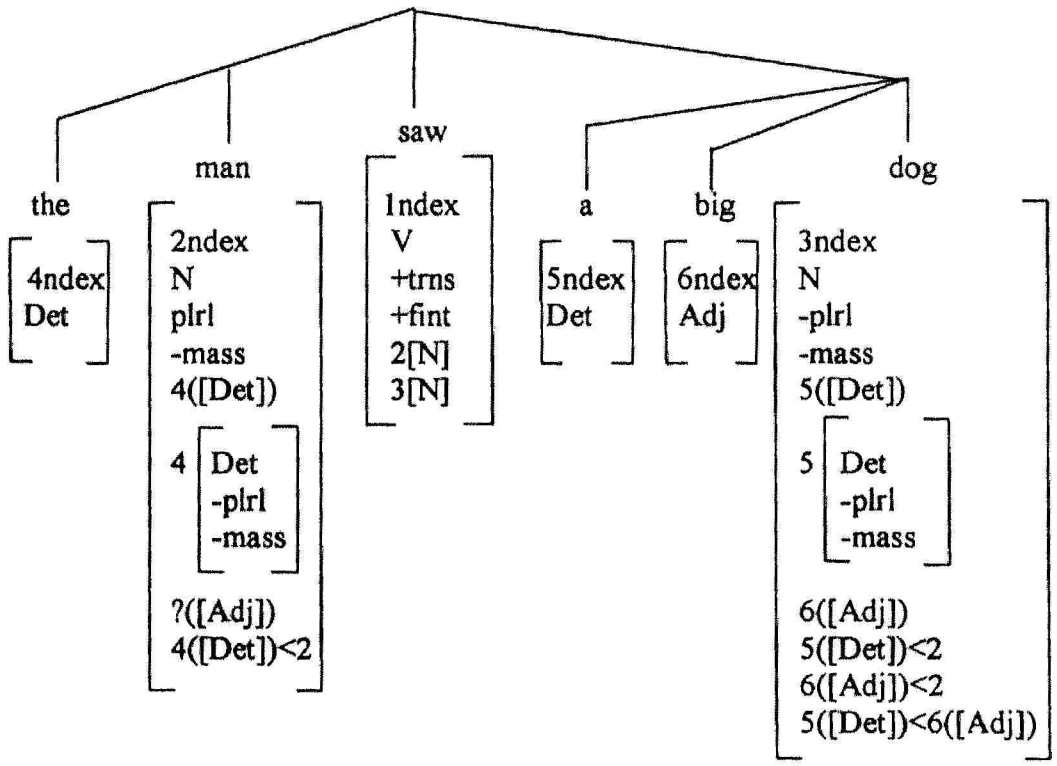


Figure 2.3

In this stemma, 1ndex (index 1) marks the head of the sentence, as it is the one word that does not depend on any other word in the stemma. Normally, the head is the main verb or root verb of a sentence. In this sentence, the head *saw* has two obligatory dependents, *man* and *dog* that are labeled as 2ndex and 3ndex respectively. The determiner *the* (4ndex) is the dependent of *man*, and *a* and *big* (5ndex and 6ndex respectively) are dependents of *dog*. We have explained in Figure 2.1 that the dependents *the*, *a* and *big* only depend on their regent but are unrelated to the verb *saw*. *The* may appear linked to the head in the Figure. However, it should be noted that the relationship is blocked by the presence of a node above the 'mast' of *man*. The sister-hood constraint does not allow any dependency to go beyond the regent in Lexicase.

The binary branches and the features in the matrices in the stemma show how the words are syntactically related to the head. The matrix of *saw* says the verb is transitive and finite as indicated in the non-contextual feature of +trns and +fint. The verb also requires two nouns as obligatory dependents, one from 2ndex and the other from 3ndex, which are the subject and object of the transitive verb respectively. The assignment of features for the subject and object will be shown when we discuss the assignment of Case in the later part of this chapter.

Constraints on word order are formalised by using the symbol '<' for linear precedence in a lexical matrix. 4([Det])<2 states that the Det that has taken the 4ndex must precede its regent NP *man* that is indexed 2, and 5([Det])<6([Adj]) means that the Det should precede the Adj when both of them serve as dependents to the same regent. Hence, word orders such as **man the* and **big a dog* are considered

ungrammatical because the index for linear order is not satisfied as shown in (a) and (b) below.

(a) * man the
2ndex 4ndex
?([Det])<2

(b) * big a dog
6ndex 5ndex 3ndex
?([Det])<6([Adj])

In (a), *man* is looking for a Det that precedes it, but there is no such lexical entry in the structure. The mark ? cannot be filled, resulting in an ill-formed construction. Similarly, (b) is ungrammatical because the Det *a* (5ndex) occurs after 6ndex, so, *big* cannot index its feature for linear order. In Lexicase, a phrase or sentence is considered ill-formed or ungrammatical if there exists features that cannot find an appropriate index.

After seeing the *lexi-* part of the grammar and the 'surface' linear order of all the words in a sentence, one may expect to see a fully illustrated stemma that contain features that are related to the *–case* part of the theory. However, prior to that, it is necessary to understand what 'case' is in Lexicase.

Case, in traditional grammar, is the inflectional category that is marked on nouns to indicate the type of semantic-syntactic relation they bear to other words in the sentence. Typical examples of cases are the nominative case for the grammatical subject and the accusative case for the direct object in accusative-type languages such as English. Case grammar, on the other hand, is the general term for various developments in grammatical theory which concerns the semantic functions in the

organisation of grammar, originating in mid-to-late 1960s. In a narrow sense, it refers to Fillmorean case grammar, due to Fillmore's significant use of case relations in his grammar. The case in Lexicase is an offshoot of Fillmore's case grammar, but differs from the former in several ways, as will be explained below.

The word 'case' in Lexicase is similar to the use of 'case' in Fillmorean convention, the 'case form' in traditional grammar, the use of 'case' in Chomsky's Government and Binding (GB) theory or the 'final terms' in Perlmutter and Poster's Relational Grammar. In Lexicase, it is defined as the "syntactic-semantic relations between a noun and its regent" (Starosta, 1988: 114-115). It covers three categories: role relations, case forms and case markers. Role relations include case relations (CRs) and the macrorole actor. Case form (CF) is the configuration that marks case relations. In the grammar, every noun and preposition is marked with a case form, and among them, Nom (nominative) is universal and Acc (accusative) is present in all accusative languages such as English and Chinese. Case markers include the morphological shape ('case inflection') and the lexical subclass, word order and prepositions/ postpositions (Starosta, An On-line Reference Manual [henceforth 'On-line'] 28.06.2001). The macrorole actor is similar to Halliday's 'logical subject' (Starosta, 1997). It is the entity in which the action of the verb is attributed. In Lexicase, macrorole co-occurs with the AGT (agent) of a transitive clause or the PAT (patient) of an intransitive one (Starosta, 1988).

Lexicase differs from Fillmorean grammar in the number of case relations. Fillmore (1968) suggests 5 case relations in his grammar, which are replaced by a slightly different list in Fillmore (1971). Case notions in Fillmore (1971) are Agent

(A), Counter-agent (C), Object (O), Result (R), Instrument (I), Source (S), Goal (G) and Experiencer (E). Lexicase (1988), however, limits case relations to only five: Patient (PAT), Agent (AGT), Locus (LOC), Correspondence (COR) and Means (MNS).[6] The PAT and AGT are complements, that is, as grammatically required dependents of their regent lexical items, while LOC, COR and MNS may appear either as complement or as adjuncts. The latter are dependents that are grammatically allowed but not necessarily required. Case relations in Lexicase correspond to Chomsky's thematic relation or θ-role in GB Theory. Examples below illustrate aspects in which Lexicase differs from Fillmorean-Chomskyan analysis in case assignment. Based on the differences, we may deduce major properties of Lexicase which will be verified in Section 2.1.2.

Sentences (1)-(3) below are pairs of controversial examples often cited in linguistics:

(1) a

	Mary	hits	John.
Fillmore :	A		O
Starosta :	AGT		PAT

b.

	John	is hit by	Mary.[7]
Fillmore :	O		A
Starosta :	PAT		MNS

[6] For definitions of individual case relations in Lexicase, refer to 'Glossary in Lexicase' in this dissertation.

[7] Note that Fillmorean-Chomskyan analysis marks cases on N only but Lexicase on N as well as prepositions. The difference is a matter of tradition and it does not affect the outcome of the analysis. It is therefore not included in the discussion.

(2) a. The post office is to the right of the bank.

Fillmore :	O	L
Starosta :	PAT	LOC

b. The bank is to the left of the post office.

Fillmore :	L	O
Starosta :	PAT	LOC

(3) a. John hit his stick against the fence.

Fillmore :	A	O	L
Starosta :	AGT	PAT	LOC

b. John hit the fence with his stick.

Fillmore :	A	L	O
Starosta :	AGT	PAT	MNS

2.1.2 Properties of Lexicase

2.1.2.1 Perceptual (versus Situational)

The most striking difference between (a) and (b) in the pairs of sentences above is the criteria used in case assignment in the two grammars. Note that in all cases, Fillmore assigns the same case to the same N in both (a) and (b). For example, in the active sentence (1), the subject NP, *Mary*, is identified as A (Agent) in (a). The same case is found in its corresponding passive sentence (b), although *Mary* is no longer the subject of the sentence. Similarly, the cases for the *post office* and *bank* in (2a), and *John*, *stick* and *fence* in (3a) remain unchanged in both the corresponding (b) sentences. In other words, case relation in Fillmorean tradition is situation-oriented. If two sentences characterise the same situation, the case relation for a particular Noun will remain constant in all paraphrases of that sentence.

As we know, it is common in human languages that a situation be expressed in more than one way or narrated from different perspectives, such as the active and passive construction in (1)-(3). In Fillmorean and Chomskyan tradition, when such cases happen, one of the sentences will be chosen as the underlying or deep structure whiles the other(s) as the surface structures, assuming that there is one and only one basic construction for all paraphrases or parallel structures. Therefore, the case relations for the same NP should be the same in all relevant sentences. This principle is said to apply to all languages.

Starosta (1988:117) formulates the Fillmorean tradition as:

> If two sentences S_1 and S_2 describe the same situation, and if x_1 is a participant in that situation, then all NP's in S_1 and S_2 which refer to x_1 has the same case relation.

Problems arise with the Fillmorean's principle. For example, the sentences in (2) are symmetrical constructions that describe the same situation that can be expressed in either way. Which is the underlying and which is the surface construction? If the choice is arbitrary, case assignment would become subjective, inconsistent and language independent, and finally fails to capture grammatical generalisations about human languages.

As an alternative, Lexicase proposes case assignment in terms of sentence-specific perspectives of paraphrases. Starosta believes that 'sentences do not represent external situation, but rather speakers perception' (On-line 5.2.2001). As such, Lexicase is perceptually or grammatically-oriented [emphasis mine]. For example, the NPs in sentences (a) and (b) in (1)-(3) above are assigned cases according to their distribution in the respective sentences by grammatical criteria per se. From the

Lexicase point of view, these sentences are not related to or correspond to each other, as they are different perspectives or different ways of observing the same situation. For example, in (3), *the stick* is treated as PAT in (a) because it is the recipient of the action *hit*. However, *with the stick* is assigned MNS in (b) because it is now the Instrument used for the action of hitting. Case in Lexicase is the marking of dependent nouns for the type of relationship they bear to the head in a sentence.

As stated in Starosta (On-line), "If two different sentences refer to the same situation but portray it from different viewpoints they may contain a quite different array of case relations."

The perceptual approach of Lexicase is not ungrounded. Clark (1997:1) has pointed out that, in human languages, the choice of perspectives is fundamental to how speakers categorise the world they talk about.

> When speakers plan an utterance, they choose a specific PERSPECTIVE on what they wish to speak about. This perspective, marked by word-choice, allows them to present to their addressees a specific conceptualization of an object, property, relation or event. Word choices allow speakers to conceptualise the same entities and events in different ways. And they therefore allow speakers to highlight properties or events in different ways.

In her paper on conceptual perspective and lexicon choice in acquisition, Clark concludes that as children grow and acquire their language, they will show a contrast in difference in meanings by shifting from one conceptual perspective to another through the lexicon options they employ. A similar finding is shown in P. Li (1990) that explores the acquisition of aspect in children speaking Mandarin. It was found

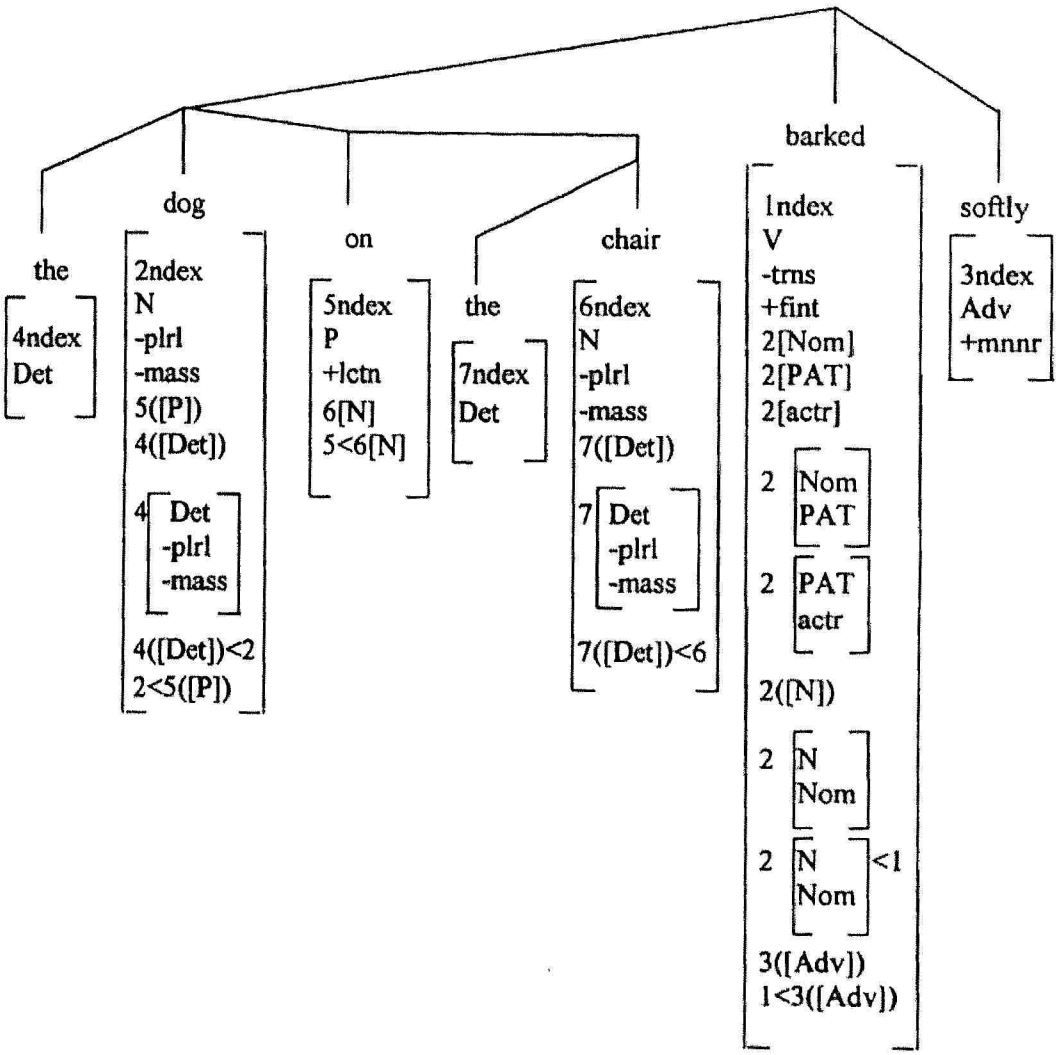
that the children would choose different aspect markers to produce utterances that vary according to their perspectives.

In sum, Lexicase believes that different perspectives produce different utterances and different syntactic constructions. We should refer to the speaker's perception of external reality, not the reality itself. Fillmorean-Chomskyan analysis, on the other hand, seeks to represent the situational perspective of a sentence by ensuring matching case assignments in logically equivalent sentence pairs.

2.1.2.2 Monostratal Representation

As Lexicase assigns case relations conceptually according to sentences per se, it does not really need more than one level of representation. It is monostratal, that is, a sentence's organisation comes from properties of its component words and there is just one level of sentence structure in grammar. It has no distinction such as deep structure/surface structure as in early Transformational Grammar, or the D-structure/ S-structure as in Government and Binding Theory. Neither does it have the f-structure/ c-structure as in Lexical Functional Grammar, nor there are no traces, PROs, empty categories nor node labels as in the Chomskyan analysis (Bender, 2002:255). Lexicase is believed to be the first generative framework to propose the total abolishment of transformations. Its representations, though extremely simple, are capable of capturing the same kind of grammatical facts as other theories (Starosta, 1988 & 1994).

The following is another stemma of Lexicase complete with all the cases specified.[8]



Stemma Notation of Lexicase

Figure 2.4

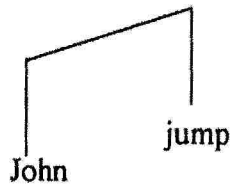
Since basic features in this stemma are similar to those in Figure 2.3, we will discuss the three new bits of information in Lexicase grammar that have been added

[8] Profesor Stanley Starosta personally provides this sentence analysis on 10 October 2001 in Hawai'i.

onto this tree, namely, the case relations, the assignment of PAT for intransitive verb in the tree, and the exocentric configuration of the prepositional phrase *on the chair*. The assignment of PAT in Lexicase will be dealt with separately in a later section. Before discussing case relations, it may be necessary to verify the syntactic function of the exocentric configuration in the grammar.

There are only two types of constructions in a stemma in Lexicase: the endocentric construction and the exocentric construction (Starosta, 1988). The former is what we have seen so far, that is, a construction with one lexical head and one or more optional or obligatory dependent(s), and the head is graphically drawn at a higher position. Figure 2.5(a) is an example for comparison. In this simple sentence, *jumps* is the head and *John* is the dependent.

(a) endocentric construction



(b) exocentric construction

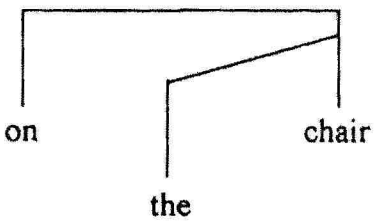


Figure 2.5

An exocentric construction, however, has more than one obligatory heads: a lexical head and one or more co-heads such as the prepositional phrase. The co-heads are linked together by a horizontal bar, such as the preposition *on* and the noun *chair* in the prepositional phrase *on the chair* in Figure 2.5(b) above.

By means of the exocentric construction, *chair* in Figure 2.4 is now able to link to the head *dog* through *on* which is a dependent on the latter and, by means of the constraint on linear precedence that 2ndex must precede 5ndex, and 6ndex must follow 5ndex, the sequential order of the sentence is satisfied.

Case relations only mark the lexical matrix of the head of the sentence in Lexicase, which is the verb *barked* in Figure 2.4. The contextual features in the matrix entail that the V may take a N, which is labeled a 2ndex, as dependent. As it occurs, the N *dog* will obligatorily be the PAT of the sentence as well as the grammatical subject, the Nom (Nominative) of the verb *barked.* A PAT is, by definition, the actor of the action.[9] Feature 2[actr] is therefore obligatory. The features 2[Nom, PAT], 2[PAT, actr] and 2[N, Nom] show how each of these four features is related to each other. In simple mathematical expression, it is a case of A is linked to B, B to C, C to D and therefore, A to D.

The non-contextual feature +mnnr (Manner) in 3ndex is a semantic feature of the adverb *softly*. Semantic non-contextual features are normally less emphasised in a stemma in Lexicase. It is believed that the syntactic relationship between a verb and an adverb does not really need to be represented semantically. For instance, the syntactic feature of 1<3([Adv]) is already sufficient to show that an Adv should not precede the verb that it modifies.

In other words, grammaticality can be represented syntactically. Semantic features may be necessary but not obligatory. As Starosta argues, "We have not lost

[9] The 'actr' (actor) in Lexicase is quite similar to Halliday's 'logical subject', but is grammatically constrained to matching either AGT (if the verbs is transitive) or PAT (if the verb is intransitive) (Starosta, 1997:8).

any semantic information at all: it is still there but distributed among several constituents" (On-line, 24.01.2000 & 28.06.2001).

Limiting grammatical information to a single level of representation is a first step in constraining the power of a grammatical framework. The discussion on stemmas above proves that Lexicase is capable of capturing all the syntactic relations between the nouns and other nouns although Lexicase has only one level of representation.

2.1.2.3 Patient Centrality

Referring to sentences (1) - (3), and the sentence in Figure 2.4, it is noticeable that the case relation of PAT (patient) is assigned to the grammatical subject of an intransitive sentence and the object of a transitive sentence in Lexicase – an untraditional convention that is very different from other contemporary grammars.

The Patient Centrality Hypothesis in Lexicase assigns the case relation of PAT to the verb. The hypothesis postulates that every verb (the intransitive verb in particular) has a PAT complement, and that every transitive verb has an AGT (Agent) complement.[10] Hence, PAT is obligatory in the case frame of every predication. It corresponds to Halliday's Medium and Actor as in the case assigned in sentences (4) and (5) below (Starosta, 1995).[11]

---

[10] This assumption in Lexicase is meant to capture the large number of universal and language-specific generalizations in both accusative languages (such as English and Chinese) and ergative languages (such as Tagalog).

[11] As cited in Starosta (1995 & On-line), Halliday (1985) describes that "Every process has associated with it one participant that is the key figure in the process...Let us call this element the Medium...Thus the Medium is the nodal participant throughout: not the doer, or the causer, but the one that is critically involved, according to the particular nature of the process", and "Hence in a material process the

(4)		The dog	barked.
	Lexicase:	PAT	
	Halliday	Medium	
		Actor	

(5)		The dog	bit the mailman.
	Lexicase:	AGT	PAT
	Halliday	Actor	Medium

It is argued that, given the characterisation of PAT, the grammar will be able to distinguish the complement versus adjunct grammatically, and moreover, capture various generalisations about the semantic scope of a sentence, especially the semantic distinction in intransitive sentences which is somewhat more difficult to detect in grammar (Starosta, 1988, On-line; Blake, 1994).

The Patient Centrality principle leads to another difference between Lexicase and Fillmorean analysis in comparative constructions with the subjects of the same predicate. Lexicase regards all sentences in (6) below as intransitive constructions with a patient subject and the main verb is derived from the same root *warm*. The difference is that the root belongs to different semantic subclasses (marked by subscripts in *warm*) (Starosta, 1988). Fillmore, on the other hand, assumes that the roots have the same meaning and the subjects differ in the type of case assigned to each of them (stated in parenthesis after each sentence) (Anderson, 1994).

Medium is equivalent to Actor in an intransitive clause and Goal in a transitive clause." (pp.146-147). The PAT of Lexicase is therefore similar to the Medium in Halliday's grammar. The exception is that in sentences such as *It's raining*, there is no Medium in Halliday's theory but a PAT in Lexicase.

(6) a. I am $warm_1$. (Experiencer)

b. This jacket is $warm_2$. (Instrument)

c. Summer is $warm_3$. (Time)

d. The room is $warm_4$. (Locative)

The reason for the sub-categorisation of the semantic meaning of *warm* is based on the definition of words in Lexicase, another property of the grammar that will be discussed below.

2.1.2.4 Triune Sign and Lexical Entries

As we have seen so far, all entries in the Lexicase discussion deal only with words or lexical items. Lexicase is a grammar of words in the sense that a word is an atom with its own valence; a phrase is a molecule with the atoms linking to each other in accordance with their individual valences, and a sentence is the linking of molecules with well-established valences. What is 'word'?

Traditionally, word is defined as the smallest unit that makes up a sentence. In Bloomfieldian terminology, it is the smallest minimal free form that can form an utterance on its own. According to Saussure, a word is a sign with arbitrary combination of sound and meaning and it is a unit that has to be memorised individually for each lexical item. The sign is represented in Figure 2.6 below.

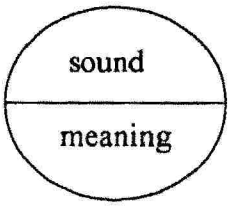


Figure 2.6

Starosta (1988, On-line) points out that the sign has omitted an important bit of information, that is, distribution which sets the condition for a word to co-occur with other lexical items, especially in an utterance. He argues that knowing the sound and meaning of a word does not necessarily entail knowing where the word could appear in a phrase or clause or sentence. For example, both *house* and *apartment* mean 'the name of a place or thing' in English, however, while the sentence *How are you going to feed and house all those bureaucrats?* is fully grammatical, **How are you going to feed and apartment all those bureaucrats?* is not (Starosta, 1988: 44-45). The reason is that the substitution of *apartment* in the sentence is distributionally unacceptable. It is therefore mandatory for a speaker to memorise the distribution of the word to produce a well-formed utterance.

Consequently, Lexicase goes one step further to modify the Saussuarean model into a triune sign that contains all the three components: sound, meaning and distribution, a significant property of Lexicase grammar. Figure 2.7 below is the symbol of a Triune Sign.

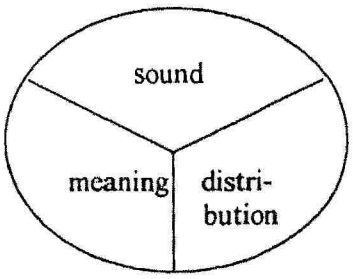


Figure 2.7

In accordance with the definition of the sign is a principle in Lexicase that a form with two different meanings or two different distributions is two different words. [emphasis mine] The four *warm* in sentence (6) mentioned earlier should, following this principle, be treated as different lexical items. The $warm_1$ in *I am warm* (6a) is in the semantic class as *happy*; $warm_2$ in *This jacket is warm* (6b) is in the semantic class as *comfortable*; $warm_3$ in *Summer is warm* (6c) is in the semantic class as *long*, and finally the $warm_4$ in *The room is warm* (6d) is in the semantic class of *spacious* (Starosta, 1988:24).

Another example is that English verbs such as *eat* which can function both transitively and intransitively as in sentences (7a) and (7b) below are treated as separate entries. In (7a), *eat* is transitive. Hence, *horses* which is the subject NP of transitive *eat* is assigned a AGT case. The entry in (b) is intransitive with only one argument; consequently, Patient Centrality obligatorily assigns PAT the subject NP of the intransitive sentence, namely *horses*.

(7) a. Horses eat grass (Somers, 1987:90)
AGT PAT

b. Horses eat.
PAT

Sentences in (8) below are controversial sentences in linguistics. Let's see how Lexicase analysis differs from the Fillmorean-Chomskyan analysis consequent to the triune sign principle that the same verb in (a) and (b) are treated as two different lexical items, mainly because the sentences differ in construction.

(8) (i) Fillmorean-Chomskyan analysis (Starosta, 1993:241)

a. Hiram sprayed the floor with paint.
Nominative Accusative Comitative
Agent Object Instrument

b. Hiram sprayed the paint on the floor.
Nominative Accusative Locative
Agent Instrument Object

(ii) Lexicase analysis

a. Hiram sprayed_1 the floor with paint.
Nom +trns Acc Com
AGT +lfct PAT MNS

b. Hiram sprayed_2 the paint on the floor.
Nom +trns Acc Lcv
AGT +lctn PAT LOC

Due to Patient Centrality and the Triune sign principles in the grammar, Starosta's Lexicase has been criticised. Blake (1994:88) comments that the grammar has deviated from the Fillmorean tradition.

> As can easily be seen, there is no attempt to keep case relations constant under paraphrases.

Meanwhile, Somers (1987:90) also agrees with Van Valin's (1980:302) criticism of Lexicase:

> This analysis has [the result that] the lexicon is virtually flooded with homophonous verb forms; one of the original positive features of Fillmore's model was that it allowed a unified treatment of verbs with both transitive and intransitive forms, and this is lost in the lexicase account.

Lexicase is criticised because it appears that it would produce unlimited numbers of lexical items and therefore the grammar is too powerful.

2.1.2.5 Seamless Morphology

The accusations above need not be true. As mentioned at the beginning of this chapter, Starosta is fully aware of the fact that a good grammar must be constrained and Lexicase is claimed to be a constrained grammar. In the early development of the grammar, and at word level, Lexicase uses various sets of Rules to reduce redundancy features in the grammar. Moreover, it tries to deduce generalisations by categorising data or features of the same properties together. The Rules include RRs (Redundancy Rules), SRs (Subcategorisation Rules), DRs (Derivation Rules), MRs (Morphological Rules) and IRRs (Inflectional Redundancy Rules) (Starosta, 1988). All these Rules are well defined in many papers on Lexicase.

Since the late 1990s, Starosta has developed a new strategy in word analysis, namely, the Word Formation Strategies (WFS). The Strategies is not meant to replace

DRs in the earlier version of the grammar, but to develop and serve as an alternative approach to deal with the same information. The Strategies initially gathers sets of related data that regularly correspond in form and content in a language, and then formally establishes the connection or relationship between the pairs of data by analogy. Note that the analogical analysis is performed at the word level – a consistent stand of Lexicase that analysis begins at the level of lexical items.

The analysis of words in WFS is "seamless", and the strategy to analyse words is termed Seamless Morphology (SM). In the traditional approach on the study of the grammatical structure of words in Linguistics, a word will first be sliced into morphs, and the slices will be grouped into morphemes for analysis. For example, the word *girls* will be divided into *girl* and *–s* and analysed later as 'plural'. In WFS, the term 'morphology' is preserved because it is also a study of words. By "Seamless", it means that, in this grammar, a word is analysed as a whole. There are (a) no morphemes, (b) no way of breaking words apart, (c) no need to recombine and (d) allomorphy is irrelevant (Starosta, 2002). Lexicase does not believe that there is internal structure in words as well as in compounds (Starosta and Ng, 1997).

A common example of analysis in Seamless Morphology is the rule for the passive form in English. For example,

(9) a. The cat eats the rat.
AGT PAT

b. The rat is eaten by the cat.
PAT MNS

Note that in sentence (9), the verb in (a) is transitive while in (b), it is intransitive. Consequently, the same nouns in (a) and (b) are assigned different case relations. However, there is a correspondence between the lexical properties of active and passive verbs such as the match between the Agent of an active verb with the Means actant of the passive verb, and the *∂n*] termination of the passive verb compared to the root in the active sentence. As such, a regular pattern is observed, and it can be analogised into rules by means of the WFS.

Rule (i),

$$\begin{bmatrix} \text{V} \\ \text{+trns} \\ \text{m[AGT]} \end{bmatrix} \quad : \quad \begin{bmatrix} \text{V} \\ \text{-trns} \\ \text{+pasv} \\ \text{m[MNS]} \end{bmatrix}$$

$$] \quad : \quad \partial n]$$

To paraphrase, the rule states that: For a transitive V, there may be a corresponding intransitive passive V ending in *∂n*. The AGT of the transitive V corresponds to the MNS actant of the corresponding intransitive. The rule proves that the characteristics of English active and passive sentences can be captured without the need for a transformation.

Another example is the WFS for the *ed*] ending group of verbs in English in the past tense. The analogical pattern is,

Rule (ii),

$$\begin{bmatrix} V \\ -past \end{bmatrix} \quad : \quad \begin{bmatrix} V \\ +past \end{bmatrix}$$

$$[X] \quad : \quad [X\textit{ed}]$$

where X is the variable and *-ed* is the constant in Seamless Morphology.

Finally, we may formulate the general pattern of Word Formation Strategies in Seamless Morphology as follows:

Rule (iii),

$$\begin{bmatrix} F_\alpha \\ F_\beta \end{bmatrix} \quad : \quad \begin{bmatrix} F_\alpha \\ F_\gamma \end{bmatrix}$$

$$[aXb] \quad : \quad [cXd]$$

This general rule implies that, corresponding to a word with the features F_α and F_β, there may be another word sharing the features F_α. The two words will share some part of their phonological representations (X, the 'variable'), but differ in one or more parts (a, b, c, d; the 'constants').[12]

It should be noted that WFS is a fairly new development in the grammar. It accounts for inflection and derivational variation and it is meant to generalise common characteristics of some particular group of words into analogical formulae; hence, the Rules derived will only be applied to words in respective groups. Different groups of words have their own characteristics and are thus subject to different WFS rules. It means that the WFS rule in Rule (ii) above is only applied to the set of verbs that have

[12] Definition provided by Stanley Starosta by verbal communication, 8 May 2002 in Hawai'i.

an *-ed* ending in the past tense. For verbs that do not have an –ed ending, such as *gave*, *stood* and so on, different rules need to be established.

The scenario is that Lexicase may have, by the Patient Centrality and Triune sign principles derived many 'homophonous verb forms' (as criticised by Van Valin), and big sets of WFS rules for, say, verbs. However, if the grammar can correlate and then formulate the verbs according to their individual characteristics,

> instead of a vast number of small verb classes with Gerrymander-like semantic ranges and syntactic distributions, there will be a fairly small number of verb classes, each with a neatly circumscribed distribution, meaning, and structure-meaning mapping pattern, and some analogical derivation rules relating word forms which appear in more than one class. The fact that two semantically similar forms have different distributions is then a simple consequence of a typical property of lexical derivation: non-productivity. (Starosta, On-line 24.07.2000)

As such, Lexicase still abides by its principle of keeping the grammar generative, but at the same time constrained

2.2 Application of Lexicase in This Study

In this Section, we will explain how the grammar is applied in this study of Tense and Aspect in Chinese. Lexicase has been used in a number of syntactic studies of Chinese in the past. Major papers include Starosta (1985) that examines case marking in Chinese, Her (1985-1986) that studies the word order of the language, and Ng (1997) that presents a detailed study of the syntactic structures of Mandarin in Singapore. This study will adopt the tradition that has been established in these works.

Various assumptions that have been generally accepted in Chinese linguistics will be included as a starting point for this work. For instance, in terms of aspect markers, the focus of the discussion will be confined to *le*$_1$, *le*$_2$, *guo*$_1$/*guole*, *guo*$_2$, *zhe* and *zài* which have been accepted as aspect markers in many studies. The present work emphasises the syntactic function of these markers as Tense and Aspect markers, not their semantic functions. Hence, we will not question the number of uses nor the semantic implications of *le*, *guo* and *zhe*, which have been heatedly disputed by linguists as discussed in Chapter One.

In most cases, the two *le* in Chinese are distributed differently. We shall follow the tradition by identifying the verbal suffix as *le*$_1$ and the clausal or sentential final particle as *le*$_2$, although it is still difficult to determine at this time the status of sentence final *le* when it is preceded by an intransitive verb. The status of the experiential *guo* has never been questioned in Chinese linguistics but it has been argued that there exists another *guo* which has the same syntactic distribution as the experiential one but differing in meaning. In this study, we also accept that there are two *guo* and we shall follow the convention by naming the non-experiential one as *guo*$_1$ and the experiential one as *guo*$_2$. We notice that *guo*$_1$ may co-occur with *le*$_1$. Moreover, the combination of *guo*$_1$*le*$_1$ (henceforth *guole*) is syntactically and semantically identical with that of *guo*$_1$ as shown in (10a) and (10b). Hence, the two markers are treated as synonyms in this study according to the Triune Sign principle in Lexicase, and they are different from *guo*$_2$ (10c).

(10) a. Wǒ chīguo$_1$ fàn jiù zǒu
I eat-mrk rice then leave
'I will leave after eating.'

b. Wǒ chīguole fàn jiù zǒu
I eat-mrk rice then leave
'I will leave after eating.'

c. Wǒ chīguo$_2$ líulián
I eat-mrk durian
'I have eaten durian before (as experience).'

2.2.1 The Triune Sign

The Triune Sign plays an important role in Lexicase analysis of Chinese. Unlike English and many other languages that have overt marking of parts of speech, tenses, agreements, case inflections and other grammatical properties on verbs, Chinese verbs do not change their form to provide such information. Whether a word is functioning as a verb or noun or preposition depends on where it occurs in a sentence. In other words, distribution plays a crucial role in deciding the meaning of the word. Here are examples of the distribution of two stative verbs:

(11) a. Píngguǒ hóng le
apple red mrk
'The apple has ripen.' (lit. 'The apple has become red.')

b. Nàge yǎnyuán hěn hóng
that-CL actor very popular
'That actor is very popular.' (lit. 'That actor is very red.')

c. Tā hóngle liǎn
she red-mrk face
'She is blushing.' (lit. 'She is red in the face.')

(12) a. Tā sǐ le
He die mrk
'He dies.' / 'He died.'

b. Wáng Miǎn sǐle fùqin
Wang Mian die-mrk father
'Wang Mian lost his father.'
(lit. *'Wang Mian has died his father.')

The word *hóng* 'red' in (11a) and (11b) are stative verbs in intransitive sentences. In (11c), it occupies a non-stative verb position. By convention, in traditional Chinese linguistics, when a word occurs in different positions in different sentence constructions, its root meaning will be regarded as the basic meaning, while the other(s) will be regarded as extended meanings of the word. *Hóng* in (a) is thus the basic meaning while that in (b) and (c) are the extended meanings of 'red'. This practice, however, faces difficulty with sentences such as (12b). In terms of distribution, the telic intransitive verb *sǐ* 'die' in (12b) function as a transitive verb as it takes an object, but this is certainly not acceptable to many Chinese linguists.

The solution to the problem provided by Lexicase is to identify words according to their syntactic distributions and to analyse the respective sentences according to the grammatical criteria advocated in the theory. So, *sǐ* 'to die' in (12a) will be treated as an intransitive verb *die* that requires a Patient as subject while that in (12b) is regarded as a transitive verb 'to lose (someone)' that needs an Agent and a Patient as subject and object respectively. This syntactic-semantic relation of the two *sǐ* 'to die' is thus clearly defined.

In view of these problems, it is reasonable to assume that Lexicase will be a useful tool in the analysis of languages like Chinese that have no verbal inflection, and that have numerous homonyms. From here onwards, we will use Lexicase to verify the syntactic analysis of tense and aspect in Chinese.

Here is another application of the Triune Sign in the identification of words in this analysis. Compare these two sentences:

(13) a. Tā chī fàn
he eat rice
'He eats.' / 'He is eating.'
(lit. 'He eats rice.')

b. Tā chīle fàn
he eat-mrk rice
'He has taken his meal.'
(lit. 'He has eaten rice.')

Both sentences in (13) contain a root word *chī* 'eat'. The only difference is that the verb in sentence (b) is suffixed with *le*. Both the verbs are .transitive and distributionally identical, but the translations show that the sentences are different in meaning. Sentence (a) indicates the action of eating is still ongoing. From the point of view of tense, it is the present tense, and from the point of view of aspect, it is imperfective. In (b), the action is completed. From the point of view of tense, it is the past tense, and from the point of view of aspect, it is the perfective. Obviously, the verbs are responsible for the differences. According to the definition of the Triune Sign, if two verbs have different perspectives, regardless of whether or not they are homonyms, they should be treated as separate verbs. We shall, therefore, regard V and V-*le* (or any other tense/aspect marker) as different lexical words in this study. The

sentences below elaborate further the differences in meaning. *Chī* in (14) co-occurs with different markers in Chinese:

(14) a. Tā $chīle_1$ fàn — 'He has taken his meal.'
(lit. 'He has eaten rice.') [sentence (13b)]

b. Tā $chīle_1$ fàn le_2 — 'He has already taken his food.'
(lit. 'He has eaten rice (now).')

c. Tā zài chī fàn — 'He is eating rice (now).'

d. Tā chīzhe fàn — 'He is eating rice.'

e. Tā $chīguo_1$ fàn le_2 — 'He has taken his meal.'
(lit. 'He has eaten rice.')

f. Tā chīguole fàn le_2 — 'He has taken his meal.'
(lit. 'He has eaten rice.')

g. Tā $chīguo_2$ fàn — 'He had eaten rice before.'
('He' is someone who comes from a place where eating rice is uncommon)

Sentences (a) to (g) above show that the meaning of the basic sentence 'He eats' varies according to the type of aspect marker that is affixed to the verb.[13] We may again assume *$chīle_1$*, *chīguo / chīguole*, *$chīguo_2$*, and *chīzhe* to be different verbal lexical items according to the Triune Sign principle in Lexicase. *Le_2* and *zài* occur in different positions syntactically. The former is clausal or sentential final and

[13] Note that the verb is not the only factor that decides the aspectual property of a sentence. This statement refers only to a basic SVO sentence in Chinese.

the latter occurs before the verb *chī* 'eat'. As they differ in distribution, they are, therefore, two separate lexical items.

Following the Lexicase convention, we shall then treat *le*$_1$, *guo*$_1$ / *guole*, *guo*$_2$, and *zhe* as inflectional markers. The V-marker construction will be treated as a single word in the discussion for the rest of this study. It might produce many tensal/aspectual verbs as has happened to other Lexicase studies. However, we will attempt to examine the related features between these verbs and formulate them using the rules in Seamless Morphology.

2.2.2 Subcategorisation

In Figures 2.2 to 2.4 above, the lexical items in Lexicase are marked by either a '+' (plus) or a '–' (minus) sign, showing that Lexicase characterises lexical items on a binary basis. The same principle applies to the subcategorisations of verbs, prepositions and other parts of speech in the grammar. We shall follow the same strategy in the analysis of tense and aspect in Chinese in this study.

In a model designed for the purpose (Table 2 below), we shall adopt the following procedure: (A) to subcategorise the tense/aspect markers on a binary basis, (B) to classify the syntactic subcategorisation of verbs in Chinese. The purpose is to identify the distinctive properties of individual markers and all relevant verbs. Based on the investigation, we will be able to discover syntactic constructions that are tense/aspect related, and (C), to tabulate the sentences according to their tensal and aspectual properties with respect to the markers in (A). Schematically, the procedure can be seen as follow:

Table 2

Syntactic Analysis Model: Tense and Aspect in Chinese

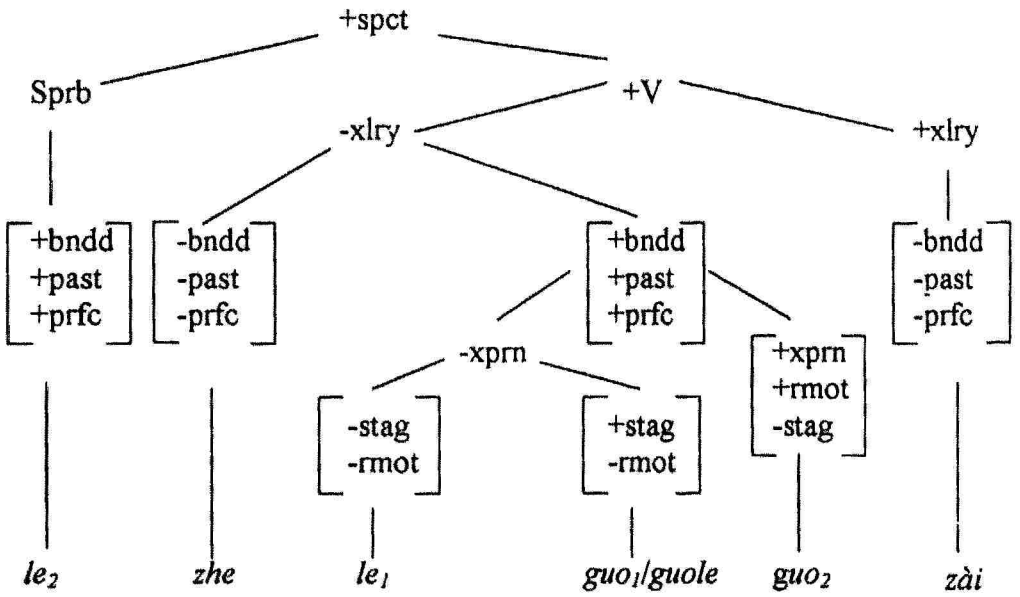
	(A) Subcategorised Tense/Aspect markers
(B) Syntactic subcategorisation of verbs	(C) Syntactic constructions with respect to Tense and Aspect

For example, we will have *le*$_1$ as a bounded past perfective marker in (A) and a verb *chī* 'eat' which is [+trns, -xtns, -lctn, -sttv] (transitive, non-extension, non-locative and non-stative) in (B). Based on the inherent properties of *chī* 'eat' discussed earlier, we place sentences (14a) - (14g) into the appropriate columns for *le*$_1$ in (C).

In subsequent chapters, we will examine the grammatical functions of the markers in sentences in (C) and find out whether or not they function as tense and/or aspect markers.

A detail study of (B) and (C) will be discussed in Chapter Three. In the remaining part of the chapter, we will only elaborate on how the tense/aspect markers are subcategorised and identified.

Figure 2.8 below shows that the subcategorisation of aspect markers begins with the specification that spct (Aspect) markers are found in two parts of speech in Chinese: V (Verb) and Sprb (Sperrbaum).[14]



Complete Chart of the Subcategorisation of Tense and Aspect Markers in Chinese

Figure 2.8

Those categorised under V are further divided into binary features of xlry (Auxiliary), bndd (bounded), past (Past), prfc (Perfective), xprn (Experiential), stag

[14] As Chinese is generally believed to have grammaticalised aspect and the presence of tense has yet to be proven, we shall use spct (Aspect) as the basis of subcategorasation rather than tense. Sperrbaum (Sprb), is a German term that means 'an obstruction, especially at border crossings, which can be placed in a vertical position'. It is the concept of 'obstruction' and 'at border crossings' in the term that fits nicely into the complicated function of *le2* in Chinese.

(stage), rmot (remote) and so on. The features are generalisations established in the findings of other scholars and observations deduced from our data.

Sperrbaum (Sprb) is the part of speech that is traditionally known as particles in linguistics. In Lexicase, it is defined syntactically as "a class of words that occur as the rightmost dependent of a root or a topic word".[15] The word is chosen because it is found to be more suitable for the capturing of the unique properties of *le*$_2$ in Chinese than the traditional term called 'particles'. In Chinese, the function of *le*$_2$ does not only express aspect and mood like some particles do, but also functions as a clausal or sentential boundary in syntax. It can also co-occur with certain particles, for example, in the sentence *Tā mǎi fángzi le ma* ?'Did he buy a house?' Particles generally do not have such properties.

Figure 2.8 above serves as an assumption that tense and aspect markers in Chinese are distinguishable by features, and the features will assist in the examination of the correlations between the verbs, the aspect markers and the syntactic structures in the language. In Lexicase, subcategorisation may be represented by a series of corresponding linear representations such as the Subcategorisation Rules (SRs) or the Redundancy Rules (RRs).

The SRs characterise choices that are available within a particular category. For example, the first Rule to the chart above would be:

SR-1 [V] → [±xlry]

that reads as 'Verb is subcategorised into Auxiliary and non-Auxiliary', which means that a V is either an auxiliary or a non-auxiliary. Similarly,

[15] This definition was provided by Professer Stanley Starosta on 15.5.2001 in Hawai'i.

SR-2	[-xlry]	→	[±bndd]
SR-3	[+bndd]	→	[±xprn]
SR-4	[-xprn]	→	[±stag]

'Bounded' is a common term in aspect. Li and Thompson (1981:185-186) claim that an event can be bounded: (a) by being a quantified event, (b) by being a definite or specific event, (c) by being inherently bounded because of the meaning of the verb, and (d) by being the first event in a sequence. A bounded event is always perfective and +past because the perfective is 'the continuing present relevance of a past situation' by definition (Comrie, 1976:56).

We distinguish the experiential [±xprn] from *guo*$_2$, *le*$_1$ and *guo*$_1$/*guole*. *Guo*$_2$ is used for an event that has taken place at least once and no longer holds true (Ma, 1977). It marks an indefinite past event. Hence, compared to *le*$_1$ which refers to an event that has just happened at the reference time, it is therefore more 'remote' (+rmot) than the latter (Ma, 1977; Huang, 1988).

Le$_1$ and *guo*$_1$/*guole* are further distinguished by ±stag (stage). The term 'stage' is used in the study to refer to only daily activities or an event that is expected to happen in a series of actions or processes. It can take the marker *guo*$_1$/*guole* (Zhang, 1986; Y. Liu, 1988). It is like something that has happened at a certain stage of the whole process. *Guo*$_1$/*guole* does not have such properties (Egerod, 1994). It is used for an event that has happened not long ago, it is therefore, not as 'remote' as *guo*$_2$.

Figure 2.8 can be simplified as it contains many redundant features that can be omitted. In Lexicase, there are Redundancy Rules that capture implicational or

predictable features. When the features are understood, they can be omitted. For instance, it is always true that a pronoun is a noun. The RR-13 in Starosta (1988) has the following formulation:

[+prnn] → [N]

The rule says that 'all pronouns are nouns'. Hence, the feature [N] can be left out of all pronoun matrices in the lexicon since it is predictable and redundant.

It can be seen in Figure 2.9 that a bounded lexical item is always perfective and past. Conversely, an unbounded one is always imperfective and non-past. We may therefore formulate two corresponding RRs that omit the feature [αpast] and [αprfc] in the tree, where α is '+'or '–'.

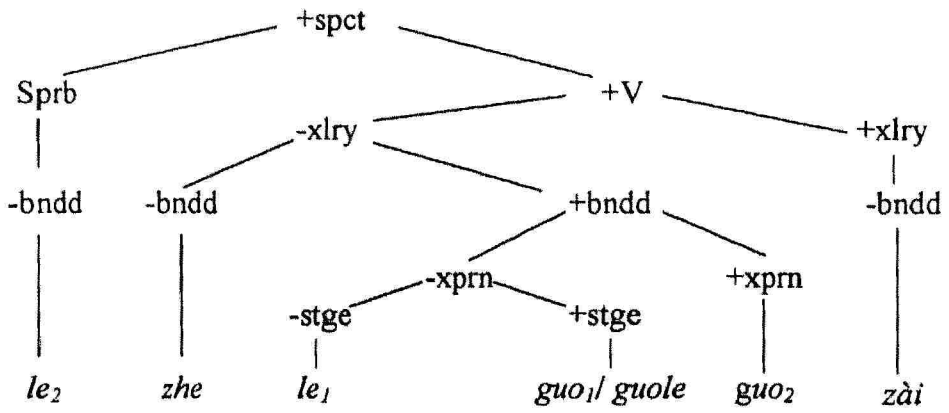
RR-1 [αbndd] → $\begin{bmatrix} \alpha\text{past} \\ \alpha\text{prfc} \end{bmatrix}$

Similarly, the co-existence of stag (stage) and rmot (remote) can be stated as follows:

RR-2 [αstag] → [αrmot]

RR-3 [+xprn] → $\begin{bmatrix} +\text{rmot} \\ -\text{stag} \end{bmatrix}$

Figure 2.9 below is the simplified version of Figure 2.8 after removing all the features that are redundant. The omitted features will not be specified in the remaining analysis unless it is necessary.



Simplified Chart of the Subcategorisation of Tense and Aspect Markers in Chinese

Figure 2.9

Figure 2.9 shows how neatly individual markers are distinguished from each other, but at the same time, related to each other through syntactic and semantic properties. As such, we have linked the markers into a paradigm of interrelated elements. The problem that remains here is: will these markers perform grammatical functions systematically? The answer will be provided in the later chapters.

2.3 Concluding Remarks

In this chapter, we have provided a detailed description of the grammar of Lexicase, with emphasis on its distinctive properties that differ from other contemporary theories. It is also explained how the theory will be applied in this study, as well as how the properties of the markers can be distinguished from each other. Chinese lacks explicit distinction between categories or function of words. It appears

that the Lexicase approach of distinguishing words by syntactic criteria and the regrouping of the lexical items by Seamless Morphology will help us to investigate the characteristics of tense and aspect in the language.