Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Using an integrated feature set to generalize and justify the Chinese-to-English transferring rule of the 'ZHE' aspect*

Yun-hua QU†[1], Tian-jiong TAO[2], Serge SHAROFF[3], Narisong JIN[4],
Ruo-yuan GAO[5], Nan ZHANG[6], Yu-ting YANG[7], Cheng-zhi XU[8]

([1]*School of International Studies, Zhejiang University, Hangzhou 310058, China*)
([2]*Institute of Fluid Physics, China Academy of Engineering Physics, Mianyang 621900, China*)
([3]*Center for Translation Studies, University of Leeds, Leeds LS2 9JT, UK*)
([4]*Center for Applied Linguistics, Hangzhou Normal University, Hangzhou 310012, China*)
([5]*School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*)
([6]*Sensor Network and Application Research Center, Graduate School, Chinese Academy of Sciences, Beijing 100049, China*)
([7]*School of Economics, Zhejiang University, Hangzhou 310027, China*)
([8]*Zhejiang University City College, Hangzhou 310015, China*)
†E-mail: qu163hua@163.com

Received Apr. 18, 2010; Revision accepted Aug. 4, 2010; Crosschecked June 30, 2010

Abstract: In machine translation (MT) practice, there is an urgent need for constructing a set of Chinese-to-English aspect transferring rules to define the transferring conditions. The integrated feature set was used to generalize and justify the Chinese-to-English transferring rule of the 'ZHE' aspect (ZHE Rule). A ZHE classification model was built in this study. The impacts of each set of temporal, lexical aspectual, and syntactic features, and their integrated impacts, on the accuracy of the ZHE Rule were tested. Over 600 misclassified corpus sentences were manually examined. A 10-fold cross-validation was used with a decision tree algorithm. The main results are: (1) The ZHE Rule was generalized and justified to have a higher accuracy under the two metrics: the precision rate and the areas under the receiver operating characteristic curve (AUC). (2) The temporal, lexical aspectual, and syntactic feature sets have an integrated contribution to the accuracy of the ZHE Rule. The syntactic and temporal features have an impact on ZHE aspect derivations, while the lexical aspectual features are not predictive of ZHE aspect derivation. (3) While associated with active verbs, the ZHE aspect can denote a perfective situation. This study suggests that the temporal and syntactic features are the predictive ZHE aspect classification features and that the ZHE Rule with an overall precision rate of 80.1% is accurate enough to be further explored in MT practice. The machine learning method, decision tree, can be applied to the automatic aspect transferring in MT research and aspectual interpretations in linguistic research.

Key words: ZHE aspect transferring rule (ZHE Rule), Machine learning, Decision tree, Aspect classification, Integrated feature set
doi:10.1631/jzus.C1000104 **Document code:** A **CLC number:** TP391.1

1 Introduction

1.1 Related work

Both tenses and aspects are concerned with temporal information of the sentences, but they are entirely different from each other in their semantic connotations. Tenses are temporal indicators that tell speakers/readers whether the events happened in the past, are happening at present, or will happen in the future. There are three tenses catering for the three circumstances mentioned: past, present, and future. In contrast, aspects tell speakers/readers whether an action is in the process of happening and, has been completed, or is a continuous state; its hypogynies are imperfective, progressive, perfective, perfect, and

* Project supported by the National Social Science Foundation of China (No. 08BYY001) and the Worldwide Universities Network 2009 Research Mobility Programme

perfect progressive. Both tense and aspect play a crucial rule in MT. As tense and aspect are embodied in every sentence, their mistranslation will undermine the integrity of the meaning of the whole sentence.

Different languages have different hypogynies and ways to express aspect. Aspect in Chinese (imperfective, progressive, perfective, perfect, etc.) is expressed by constructions with predicate verbs and aspect markers like ZHE, ZAI, LE, and GUO.

In current Chinese-to-English translation practice of aspect, one Chinese aspect can be transferred to several English ones. For instance, the imperfective composed of 'verb+ZHE' can be transferred to English progressive, perfect, perfect progressive, the simple tense with imperfective, etc. The examples are listed as follows:

(1) Ta (She) shuo (spoke) shi (as) shou (hand) shijin (furiously) pai (pounded) ZHE chuang (bed). (She pounded furiously on the bed, as she spoke.) (English simple tense with imperfective implications)

(2) Ta (He) cishi (this moment) zheng (just now) pan (look forward to) ZHE guonian (the New Year celebration). (At this moment he was looking forward to the New Year.) (English Progressive)

(3) Zhanqian (Before the war) you (there are) san'ge (three) nvren (girls) qiang (vie) ZHE jia (marry) ta (him). (Before the war three girls had vied to marry him.) (English perfect)

(4) Xiangzi benlai (originally) pan (look forward to) ZHE guonian (celebrate the new year). Xianzai (now) keshi (but) yidian ye bu qijin (felt nothing). (Xiangzi had been looking forward to the New Year, but now he felt nothing.) (English perfect progressive)

The undefined transferring conditions for this one-to-many matching relationship have been the underlying cause of aspect errors in Chinese-English machine translation (MT) practice. There is an urgent need for a set of Chinese-to-English aspect transferring rules to diminish the errors in MT.

Until now much research has been done on some theoretical issues concerning Chinese aspect, such as the establishment of Chinese aspectual systems (Dai, 1997; Chen, 2003; Sun, 2007), the description of the relationships between Chinese aspect and situation aspect (Xiao and Mcenery, 2004), the collocations of temporal adverbials with Chinese aspect markers (Xiao, 2002), the relationships between Chinese aspects and modality (Pen, 2007), the grammatical and tense meanings that Chinese aspect markers can express (Jin, 2003; Lin, 2004; Robson, 2005), the non-obligation status of the Mandarin aspect markers, LE and ZHE (Wu, 2005), the dependency of LE, ZHE, and GUO on the extra-sentential context (Ljungqvist, 2007), the semantic ambiguity of LE to be interpreted as perfective or imperfective (Chen, 2009), and the connection between the 'change of state' and the 'contrary to expectation' interpretations of Mandarin sentence-final-LE (Soh, 2009).

Most Chinese aspect researchers contend that the aspect marker ZHE expresses the duration of an action and a continuous state (Chen, 1990; Smith, 1991; 1997; Shi, 1992; Shi and Hu, 1998; Lu, 1999; Qian, 2000; Xiao and Mcenery, 2004). Shi and Hu (1998) contended that ZHE denotes the durative situation of an action, when associated with the active verbs, and denotes a stative duration, when associated with the stative verbs.

Comparative studies have focused mainly on English-to-Chinese aspect transferring (Xiao and Mcenery, 2004; Shang, 2007), while few were conducted for Chinese-to-English aspect transferring (Zhao and Shen, 1984; Chen, 1990; Pan, 2004). However, not all of them were corpus-based studies. A small number of examples selected by these studies are not enough to reveal the panorama and true nature of the Chinese-English aspect transferring phenomenon.

All these studies have nothing to do with the practical issues for automatic Chinese-to-English aspect transference.

Much work has been done on determining Chinese tenses in automatic Chinese-to-English translation. Olsen *et al*. (2001), using headline, grammatical aspect, adverbials, and the lexical aspectual features of telicity, suggested a tense and discourse structure in the English translation of a Chinese newspaper corpus. Ye *et al.* (2006), using surface and latent features, tested the precision rate of English tense generation with the classification tree. Ye and Zhang (2005), employing a moderate number of linguistic features, did the automatic tense tagging for Chinese verbs by training a tense classifier upon a small amount of manually annotated data. Li *et al.* (2004), taking into account the effects of linguistic features and lexical aspects, typed the temporal relations using four dif-

ferent classification approaches and examined the impact of each feature. Wong *et al*. (2005) developed a framework for modeling and representing temporal discourse structure, the temporal relations (i.e., 'before', 'after', and 'same-as') among the sentences. However, their work was concerned with only three tenses: past, present, and future; the Chinese aspects (imperfective, progressive, perfective, and perfect) were not probed.

Ye *et al.* (2007) conducted a pilot study on Chinese aspect marker generation in English-to-Chinese MT. They built an aspect marker classifier based on a maximum entropy model with promising classification accuracy, and also identified the highest utility of the different clusters of features.

Little work has been done on the Chinese-to-English aspect transferring rules by MT researchers. Cheng *et al*. (2004) proposed a set of Chinese-English tense and aspect transferring rules. However, it was not a bilingual corpus-based study and lacked semantic analysis. Moreover, the transferring rules were not formalized or tested for their accuracy. Thus, there are obviously some problems with the generalization, reliability, and feasibility of their rules.

Although experienced translators can now correctly translate the Chinese aspects to English, their cues for translating aspects have not been generalized as formalized rules or applied to the real practice of MT.

As a whole, linguists, MT researchers, and human translators have reported only the achievements of studies in automatic Chinese-to-English tense transferring and automatic English-to-Chinese aspect. They have left the problem of automatic Chinese-to-English aspect transferring rules untouched, except for Qu (2008), who proposed a set of ZHE aspect transferring rules (ZHE Rule) from Chinese to English based on a comparative study of Chinese and English aspects.

1.2 Qu (2008)'s ZHE Rule from Chinese to English

Qu (2008)'s ZHE Rule (Fig. 1) embodies three parts: the Chinese source ZHE imperfective aspect, the transferring conditions, and its target English aspects. Compared to a horizontal tree, it has two sub-rules with two roots.

The first root has the Chinese source ZHE aspect formed by 'active verb+ZHE' and has four branches with four target English aspects on its leaves, while the second root has the Chinese source ZHE aspect formed by 'stative verb+ZHE' and has two branches with two target English aspects on its leaves.

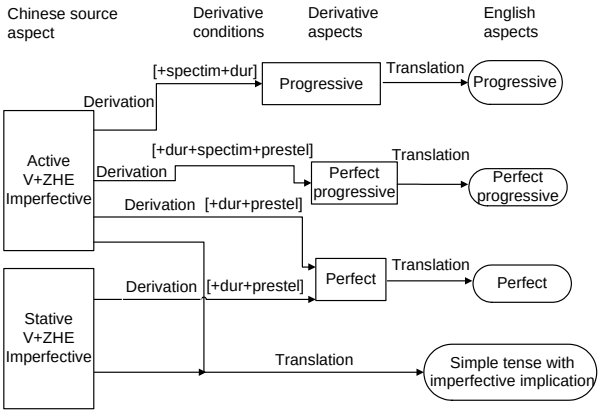

Fig. 1 Qu (2008)'s ZHE Rule from Chinese to English
The semantic features of ZHE imperfective and the derivative aspects are elaborated. dur: durative; spectim: specific time; prestel: present telic

Each root has two transferring paths. One is the direct transferring path without transferring conditions, like the generation of the target English simple tense with the imperfective implication. The other path describes a two-phase-transferring process: the first is the derivative process of Chinese aspect, embodying the original aspects, the derivative conditions, and the derivative aspects; the second is the translation process, embodying the derivative aspects and target English aspects. The derivative conditions are the transferring conditions in the form of semantic features of temporal adverbials. A derivative principle was proposed in Qu (2008) and Qu and Feng (2008) to interpret the underlying causes for the derivation of ZHE aspect by revealing the semantic relationships between adverbials, verbs, and the aspect markers involved in the derivative process. The derivative principle says that, in a derivative aspect pattern 'adverbial+verb+aspect marker', the lexical meanings of the adverbial can override the aspectual meanings of the 'verb+aspect marker' and consequently, an aspect is derived. The examples of Qu (2008)'s ZHE Rule are listed in Table 1.

Qu (2008)'s ZHE Rule stipulates the Chinese source ZHE imperfective, the transferring conditions, and its target English aspect types. However, the rule and the principle need statistical support to justify their rationality and accuracy.

Table 1 Examples of Qu (2008)'s ZHE Rule*

Chinese source aspect	Examples of Chinese aspect sentences	Derivative conditions	Derivative aspect sentences & their English correspondents	Derivative aspects & semantic features
Active V+ZHE imperfective [+dyn+dur]	(5) Ta (He) pan (look forward to) ZHE guonian (the New Year). He looked forward to the New Year.	Null	He looked forward to the New Year.	Simple tense with imperfective implication [+dyn+dur]
		cishi, zheng (this moment) [+spectim]	(6) Ta (He) *cishi* (this moment), *zheng* (just now) pan (look forward to) ZHE guonian (the New Year). At this moment he was looking forward to the New Year.	Progressive [+dyn+dur+spectim]
		tiantian (everyday) [+dur+spectim +prestel]	(7) Ta (He) *tiantian* (everyday) pan (look forward to) ZHE guonian (the New Year). Every day he has been looking forward to the New Year.	Perfect progressive [+dyn+dur+spectim +prestel]
	San'ge (Three) nvren (girls) qiang (vie) ZHE jia (marry) ta (him). Three girls vied to marry him.	zhanqian (before the war) [+dur+prestel]	(8) San'ge (Three) nvren (girls) *zhanqian* (before the war) qiang (vie) ZHE jia (marry) ta (him). Before the war three girls had vied to marry him.	Perfect [+dyn+dur+prestel]
Resultative stative V+ZHE imperfective [+stat+dur]	(9) Zhongguo (China) de ('s) jihua shengyu (family planning) baohan (include) ZHE liangge fangmian (two aspects). China's family planning includes two aspects.	Null	China's family planning includes two aspects.	Simple tense with imperfective implication [+stat+dur]
		shizhong (always, indicating a constant state) [+dur+prestel]	(10) Zhongguo (China) de ('s) jihua shengyu (family planning) *shizhong* (always) baohan (include) ZHE liangge fangmian (two aspects). China's family planning has always included two aspects.	Perfect [+stat+dur+prestel]

* Adapted from Table 2 in Qu *et al.* (2008). stat: stative; dyn: dynamic; dur: durative; spectim: specific time; prestel: present telic

2 Task description

The goal of the study is to generalize a linguistic rule, a ZHE Rule for Chinese-to-English MT, by testing its accuracy to examine whether it is correct and applicable to MT. The ZHE Rule, based on Qu (2008)'s ZHE Rule, is generalized by identifying classification features and tested by a machine learning method, the decision tree.

The whole study involves preparations and testing experiments such as selecting and preprocessing corpus, prescribing the classification features, processing the multi-aspect alternative translation sentences, and building a classification testing model.

The proposed ZHE Rule is expected to stipulate the source ZHE aspect, the transferring conditions, and the target English aspects, forming one-to-one matching transferring versions.

In addition to the aspect translation version prescribed by the ZHE Rule, there may exist other correct alternatives of English aspect translation versions. However, the ZHE Rule is expected to be applied to MT. As long as its aspect transferring version is correct, it can be applied to MT regardless of whether the other versions are correct or not.

2.1 Experimental method

The machine learning method, by building statistical models, has a faster speed in training and testing. Now it has been widely and successfully applied in economics, medical science, biology, chemistry, etc. It has been used in aspect generations from English to Chinese (Ye *et al.*, 2007) and in tense translation from Chinese to English (Ye and Zhang, 2005; Ye *et al.*, 2006), but has not been applied to aspect-classification-testing for generalizing linguis-

tic rules, such as aspect transferring rules from Chinese to English.

The proposed ZHE Rule should stipulate the Chinese source aspect 'verb+ZHE', the transferring conditions, and the target English aspect types. Thereby, to test the aspect classification is to test whether the transferring conditions and other features in the Chinese source sentences are the correct vectors for classifying the correct Chinese aspect types. If the aspect types are correctly classified in the Chinese source sentences, their target English aspects can be defined accordingly. Therefore, the justifying task in this study is a classification task of the Chinese source aspect types with transferring conditions, and other relevant vectors. Thus, the decision tree algorithm, a machine learning method for classification, can be employed.

The Weka toolbox is a collection of machine learning algorithms for data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. The J48 algorithm embedded in the Weka toolbox is an improved decision tree algorithm for classification. It is an implementation of C4.5, and can conduct the classification task based on the discrete data by generating a decision tree. Based on the vector on each note, J48 performs the decision results by following a path from the root to the leaves corresponding to the classes of the expected tasks. As the transferring conditions and other relevant vectors are the discrete data, J48, a decision tree learner, was selected as an instrument to conduct the aspect classification task.

2.2 Corpus for generalizing and testing the rules

As Chinese fictions are the most abundant carriers of grammatical aspects (Wang, 2004), this type of text was selected as the main pool for constructing the Chinese-to-English parallel aspect corpus. The bilingual parallel corpus employed in this study was built upon five modern Chinese fictions written by famous Chinese writers, and 18 white papers issued by the Information Office of the State Council of the People's Republic of China in the years from 1991 to 2002, together with their target English translated versions.

2633 pairs of Chinese source and their target English sentences, embedded with 'verb+ZHE' phrase structures, extracted from the above bilingual parallel corpus, were manually annotated with three types of grammatical aspect, two types of tense, and one type of non-grammatical aspect in their target English sentences, based on the aspectual systems proposed by Qu (2008). The three aspects are progressive aspect, perfect aspect, and perfect progressive aspect. One type of non-grammatical aspect involves all the non-predicative verb constructions such as gerund verbs, present and past participles, infinitive verbs, prepositional phrases, nouns and noun phrases, adjectives, and adverbials. All these non-predicative verbs constructions are annotated as 'OTHERS' in this Chinese-to-English ZHE aspect parallel corpus.

The two tenses are the tense with imperfective implications and the tense with perfective implications. Unlike Chinese with ZHE imperfective and LE perfective, grammatically, English has no single categories to express imperfectivity and perfectivity separately. However, semantically, English has simple tenses in the forms of past, present, and future with imperfective and perfective implications (Olsen, 1997). Thus, the simple tense, with imperfective and perfective implications, was simplified as imperfective and perfective in the ZHE Rule and, accordingly, was annotated briefly as imperfective and perfective in this Chinese-to-English ZHE aspect parallel corpus.

In light of Vendler (1957)'s verb classifications, three lexical aspects, STATE, ACTIVITY, and ACCOMPLISHMENT, were also manually annotated in this Chinese-to-English ZHE aspect parallel corpus in 2008.

As ZHE, expressing the durative meaning cannot be used with Vendler's ACHIEVEMENT verbs, the lexical aspect ACHIEVEMENT was not found or annotated in our corpus.

All the manual grammatical and lexical aspect annotating work was done by Qu (2008) and checked by her three times in 2009. As she has studied aspect semantics since 2002 and gained much knowledge in grammatical and lexical aspects, her annotating of grammatical and lexical aspects is more likely to have higher accuracy.

As the annotation of some syntactic features is dependent on the annotation of the part of speech, the Chinese-to-English ZHE aspect parallel corpus was first segmented and annotated with part of speech, using the Chinese Lexical Analysis System

(ICTCLAS). The system was designed by the Institute of Computing Technology in Chinese Academy of Sciences and had a precision rate of 99.50% in a closed test. Later, the segmented and annotated corpus was checked and revised manually by an MA graduate candidate of English, a native Chinese speaker, at Zhejiang University. Xiandai Hanyu Cidian (5th Ed.) (Institute of Linguistics in Chinese Academy of Social Sciences, 2005), with part of speech annotation, and known as the highest quality in modern Chinese dictionaries, was used as the part of speech manual checking standard for the part of speech annotation.

3 Experimental studies

As ZHE aspect classification is quite different from other classification tasks dealt with by a decision tree, in the complexity of features for data mining and alternatives of Chinese-to-English translation, this study highlights the improvement of these basic conditions of the classification rather than the core of the decision tree algorithm.

Our study primarily involves three tasks: (1) prescribing the features, (2) identifying the sentences with multiple alternatives of Chinese-to-English aspect translation, and (3) building the classification model with J48 in Weka.

3.1 Prescribing the feature sets

As the prediction of the target English aspects in the ZHE Rule is based on the types of verb, the transferring conditions, and other features of Chinese source aspects, the feature sets in this classification test are built on the source language Chinese.

The proposed ZHE Rule is theoretically formulated in Smith's two-component aspect theory (Smith, 1991; 1997), in which the aspectual meaning of a sentence is a composite of the information from the components of viewpoint and situation type. Accordingly, the lexical aspectual features, as the situation type in the two-component theory, have their contributions to the aspectual meanings. Hence, they are the aspectual cues employed as feature sets in our experiment.

The lexical aspectual features refer to the inherent temporal properties encoded in verb stems and other lexical items (Olsen, 1997). Vendler (1957)'s temporal criterion is similar to the clusters of values for lexical aspectual features, (a)telicity, (non)dynamicity and (non)durativity. In terms of these lexical aspectual features, Vendler (1957) typed the verbs into four classes: STATE, ACTIVITY, ACCOMPLISHMENT, and ACHIEVEMENT.

In light of the properties of the lexical aspect, Vendler (1957)'s three classes of lexical aspect, STATE, ACTIVITY, and ACCOMPLISHMENT, were employed as the lexical aspectual features for aspect classification. This is because the ZHE aspect marker is used mainly to signal the durative nature of a situation (Xiao and Mcenery, 2004), and cannot be used with ACHIEVEMENT situations. Therefore, in our corpus, no such case of ZHE usage was found or annotated.

Some researchers have proposed that the proper domain of lexical aspect classification is the sentence or proposition rather than the verb (Smith, 1991; 1997; Olsen, 1997). For instance:

(11) Zhang Shan zai shufang li zheng xieZHE zi. (Zhang Shan *was writing* in the study now.) (ACTIVITY)

(12) Zhang Shan yizhi zai xieZHE *yi ben shu*. (Zhang Shan *has been writing a book* all along.) (ACCOMPLISHMENT)

(13) *Qiang shang* xieZHE wu ge zi. (There *are* five words *on the wall*.) (STATE)

Thereby, under the guidance of the above generally recognized compositional theory of lexical aspect, the lexical aspects (verbs, their arguments, etc.) were manually and holistically annotated at a sentence level and later prescribed as the new features for aspect classification in this study (Table 2).

The transferring conditions, in the form of semantic features in Qu (2008)'s ZHE Rule, specified as temporal adverbs, etc. in our corpus, are hypothesized to have an impact on the derivation in Qu (2008)'s ZHE Rule. These specific temporal adverbs, etc. are the linguistic constraints for the transferring. Hence, they were employed as the temporal feature sets for the aspect classification test (Table 2).

Aspects are signaled by means of grammatical morphemes, but differ in their semantic meanings. The derivation of ZHE aspect is a derivation of semantic meanings of aspects, shifting from the semantic meanings of the ZHE aspect to other aspects. As is well known, the syntactic features have impacts

Table 2 Temporal, lexical aspectual, and syntactic features

Feature	Category	Example
Temporal	Nouns to express time interval	Zuijin jinian (in the recent years), … niandai yilai (since the year of …), etc. appear before the 'verb+ZHE'
	Temporal adverbs	Yijing (already), gangcai (just now), etc. appear before the 'verb+ZHE'
	Frequency adverbs & nouns	Shishi (*ever and again*), laoshi (*often*), etc. appear before the 'verb+ZHE'
	Temporal location adverbs & nouns	Zhengzai (*in the process of*), muqian (*at present*), etc. appear before 'verb+ZHE'
	Temporal location & time interval constructions	zicong … shishi (*since … ever and again*), yizhi yilai (*as always*), appear before 'verb+ZHE'
Lexical aspectual	Stative	(14) Zhuo shang fangZHE yiben shu. (*On the desk lies* a book.)
		(15) Qiang shang guaZHE yifu hua. (*On the wall hangs* a picture.)
	Activity	(16) Lin Peishah de yanjing xianzai shanZHE yiyang de guangmang. (Lin Peishah's eyes *were shining* with an unusual excitement.)
		(17) Tamen tiaoZHE wu. (They *are dancing*.)
	Accomplishment	(18) Ta zheng yu pengyou tanZHE yibi ruanjian shengyi. (He *was discussing a business of software* with some friends.)
		(19) Gongrenmen yizhi zai zaoZHE yisuo fangzi. (The workers *have been building* a *house* all along.)
Syntactic	'verb+ZHE' functioning as an adverbial modifier	(20) Ta shuoZHE, ye jiu zou le. (So *saying*, she went away again.)
		(21) Ding yisheng cuoZHE shou, kanZHE ta shuo: "duanqi le!" (Dr. Ding, *rubbing his hands together*, said to him, "It's all over.")
	'verb+ZHE' happening simultaneously with another action	(22) Ta yimian shuo, yimian yankanZHE ta de zhangfu. (*Said* she, *looking in* at her husband sitting.)
		(23) Nvgongmen yibian rangZHE, yibian jiu fenfen sanqu. (The crowd *broke up and drifted away.*)
	'verb+ZHE' followed by some sense verbs	(24) Gao xiansheng juede shen hou you ren genZHE. (Mr. Gao *had sensed* that he was being tailed.)
		(25) Ta zhidao jiali laopo dengZHE ta. (H*e knew* his wife was awaiting him.)

on the semantic meanings. Hence, the syntactic features may have an impact on the derivation of ZHE aspects and were supplemented as a new feature set for aspect classification in this study.

To obtain a higher precision rate of aspect transferring, an integrated feature set of temporal, lexical aspectual, and syntactic features were adopted in this classification testing (Table 2).

3.2 Identifying the multi-aspect alternative translation sentences

In Ye *et al*. (2007)'s experiment (Table 3), the precision rate of ZHE aspect marker generation from English to Chinese was only 50%, much lower than those of LE and GUO aspect marker generations (Table 3).

Our initial precision rate of ZHE aspect classification was only 57.6% (Table 4), slightly higher than that of Ye *et al*. (2007).

Ye *et al*. (2007) did not explain why ZHE had lower precision rates. After observing over 2000 pairs of ZHE aspect sentences and their English correspondents, we found that the primary cause was that ZHE aspect has multiple alternatives of aspects for the target English translation version, given the same syntactic structures and lexical structures in which ZHE is embedded. For instance, if ZHE is in the structure of 'Ta changZHE ge zoulu', it can have two correct versions of target English sentences, but with two different aspect annotations in our corpus as follows:

Table 3 Ye *et al*. (2007)'s classification accuracy of three Chinese aspects

Aspect	Precision	Recall	*F* score
LE	0.7827	0.8244	0.8051
ZHE	0.5000	0.5108	0.4986
GUO	0.6500	0.5000	0.5343
NULL	0.8038	0.7870	0.7862

Precision: the number of correct results divided by the number of all returned results. Recall: the number of correct results divided by the number of results that should have been returned. *F* score: a measure of a test's accuracy where both the precision rate and the recall rate are considered for computing its value

Table 4 Initial accuracy of ZHE aspect classification

Class	TP rate	FP rate	Precision	Recall	*F*-measure	AUC
IMPFCTIV	0.787	0.542	0.585	0.787	0.671	0.638
PROG	0.190	0.023	0.367	0.190	0.250	0.630
PFCT	0.200	0.004	0.524	0.200	0.289	0.567
PFCT-PROG	0.000	0.001	0.000	0.000	0.000	0.507
PFCTIV	0.266	0.003	0.708	0.266	0.386	0.642
OTHERS	0.422	0.196	0.578	0.422	0.488	0.634
Overall precision			0.576			

IMPFCTIV: the simple tense with imperfective implication; PROG: progressive; PFCT: perfect; PFCT-PROG: perfect progressive; PFCTIV: the simple tense with perfective implication; OTHERS: all the non-predicate constructions, gerund, present and past participle, infinitive, noun and noun phrase, adverb, adjective, prepositional phrase, etc. TP rate: true positive rate for the receiver operating characteristic curve (ROC). Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of 0.5 represents a worthless test. FP rate: false positive rate for the ROC curve. Precision: the number of correct results divided by the number of all returned results. Recall: the number of correct results divided by the number of results that should have been returned. *F*-measure: a measure of a test's accuracy where both the precision rate and the recall rate are considered for computing its value. AUC: area under the ROC curve

(26) She walked *while singing*. (OTHERS)

(27) She *sang* and walked. (IMPERFECTIVE)

These two alternative annotations in the corpus are marked as OTHERS, denoting non-predicate construction, and IMPERFECTIVE, denoting the simple tense with imperfective implication, respectively. Normally, the computer makes predictions based on the statistical probability of the target aspect annotation. In our corpus, the probability of target IMPERFECTIVE is larger than that of the target OTHERS, corresponding to the same Chinese source aspectual cues. Hence, all the target OTHERS annotations, corresponding to the same Chinese source aspectual cues as IMPERFECTIVE in our corpus, were predicted as the wrong classifications. As human beings have multi-alternative target aspect annotations corresponding to the same source aspect cues in the corpus, it is not possible for computers to give a correct predication.

Thus, we revised the annotations of over 600 multi-aspect-alternative sentences for target English, and changed all the misclassified multi-aspect alternative annotations into the correctly predicated aspect annotations in the target English classifications in our corpus. For instance, the above source sentence, 'Ta changZHE ge zoulu', can have two aspect alternative annotations in the corpus. The computer predicates the annotation of IMPERFECTIVE as a correct classification. We just changed the other choices of annotation, OTHERS, into the correct annotation, IMPERFECTIVE. The experimental results after revising the corpus are presented later in Tables 5 and 7.

As MT is aimed to simulate a human's thought, if the human assesses these two alternatives of aspect translation to be correct, why can we not let the computer judge them correct?

3.3 Building a decision tree classification model in Weka

As the testing of precision rates and AUC (area under the receiver operating characteristic curve) is a predication and an estimation of how accurately a predictive model will perform in practice, a 10-fold cross-validation was employed in this experiment. The 10-fold cross-validation technique, with 90% of training data and 10% of validation data, is much more precise than the one-round validation. In 10-fold cross-validation, our experimental data were partitioned into 10 sub-data. Nine sub-data were employed as the training data and one as the testing data. After 10 rounds of testing one after another sub-data, a single estimation of the average result of 10 times testing with 10 folds was finally reached.

A supervised machine learning classification technique was used, as the learner must predicate the correct class of target translation through learning data vectors annotated with target aspect types. In addition, the data vector annotated with the features was discrete data. Consequently, the filter should be a supervised discrete filter.

As three feature sets are involved and the ZHE aspect marker can be used with two kinds of verbs, active verbs and stative verbs, three types of experimental results using 10-fold cross-validation are presented: the precision and AUC of ZHE aspect

general classification, the impact of each feature set and their integrated feature sets on the classification, and the precision and AUC of the two sub-classifications of ZHE aspect markers used with active and stative verbs.

3.3.1 Precision and AUC of ZHE aspect general classification

Before using J48, a decision tree algorithm in Weka, to calculate the overall precision and recall rates, a preliminary task was first conducted to transform the bilingual parallel Chinese-to-English ZHE aspect corpus into a '.arf' dataset using Python 26 software.

Table 5 presents the precision and AUC of the general classification of ZHE aspect calculated using the J48 algorithm in Weka.

3.3.2 Impact of each feature set and the integrated feature sets

To assess the impact of each individual feature set and their integrated impact on the accuracy of the general ZHE aspect classification, the integrated feature sets were partitioned into three subsets, the temporal, the lexical aspectual, and the syntactic feature set. First, each of the three feature sets was fed into Weka one by one for testing its individual effect by J48. Then, three classification rounds of two-cross combinations of each individual feature set were conducted to discover their integrated effect on the accuracy using J48 in Weka, respectively. The results are presented in Table 6.

3.3.3 Precision rate and AUC of two sub-classifications

To test the two sub-classifications of the ZHE aspect, the whole corpus was partitioned into two sub-corpora, one with 'active verb+ZHE', annotated by lexical aspectual features, ACTIVITY and ACCOMPLISHMENT, the other with 'stative verb+ZHE', annotated by a lexical aspectual feature, STAT.

Table 5 Accuracy of the general ZHE aspect classification

Class	TP rate	FP rate	Precision	Recall	*F*-measure	AUC
IMPFCTIV	0.899	0.290	0.815	0.899	0.855	0.808
PROG	0.492	0.009	0.739	0.492	0.591	0.768
PFCT	0.271	0.002	0.684	0.271	0.388	0.643
PFCT-PROG	0.520	0.005	0.667	0.520	0.584	0.834
PFCTIV	0.462	0.001	0.857	0.462	0.600	0.657
OTHERS	0.732	0.090	0.786	0.732	0.758	0.827
Overall precision			0.801			

IMPFCTIV: the simple tense with imperfective implication; PROG: progressive; PFCT: perfect; PFCT-PROG: perfect progressive; PFCTIV: the simple tense with perfective implication; OTHERS: all the non-predicate constructions, gerund, present and past participle, infinitive, noun and noun phrase, adverb, adjective, prepositional phrase, etc. TP rate: true positive rate for the receiver operating characteristic curve (ROC). Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of 0.5 represents a worthless test. FP rate: false positive rate for the ROC curve. Precision: the number of correct results divided by the number of all returned results. Recall: the number of correct results divided by the number of results that should have been returned. *F*-measure: a measure of a test's accuracy where both the precision rate and the recall rate are considered for computing its value. AUC: area under the ROC curve

Table 6 Impact of each feature set and their cross-integrated feature sets

Class	Precision						
	Temp	Lex	Synt	Temp & Lex	Temp & Synt	Lex & Synt	Temp, Lex & Synt
IMPFCTIV	0.601	0.744	0.618	0.766	0.631	0.788	0.815
PROG	0.250	0.000	0.753	0.000	0.725	0.753	0.739
PFCT	0.917	0.000	0.500	0.765	0.765	0.000	0.684
PFCT-PROG	0.550	0.000	0.000	0.659	0.549	0.000	0.667
PFCTIV	0.000	0.818	0.000	0.857	0.000	0.818	0.857
OTHERS	0.250	0.768	0.613	0.777	0.657	0.782	0.786
Overall precison	0.600	0.749	0.621	0.767	0.633	0.784	0.801

Temp: temporal feature set; Lex: lexical aspectual feature set; Synt: syntactic feature set. IMPFCTIV: the simple tense with imperfective implication; PROG: progressive; PFCT: perfect; PFCT-PROG: perfect progressive; PFCTIV: the simple tense with perfective implication; OTHERS: all the non-predicate constructions, gerund, present and past participle, infinitive, noun and noun phrase, adverb, adjective, prepositional phrase, etc. Precision: the number of correct results divided by the number of all returned results

The two corpora embodied 2052 and 581 pairs of sentences, respectively. The two sub-classifications, the classification of the 'active verb+ZHE' and the 'stative verb+ZHE' aspect, were tested on two datasets formed from the two sub-corpora one by one, respectively, by J48 in Weka. The test results are presented in Table 7.

Table 7 Accuracy of two sub-classifications of ZHE aspect

Class	Precision		AUC	
	AVZ	SVZ	AVZ	SVZ
IMPFCTIV	0.814	0.753	0.819	0.786
PROG	0.716	1.000	0.816	0.821
PFCT	1.000	0.400	0.614	0.436
PFCT-PROG	0.565	0.500	0.790	0.723
PFCTIV	0.783	–	0.690	–
OTHERS	0.770	0.799	0.818	0.766
Overall precison	0.796	0.761		

Precision: the number of correct results divided by the number of all returned results; AUC: area under the receiver operating characteristic curve; AVZ: 'active V+ZHE'; SVZ: 'stative V+ZHE'. IMPFCTIV: the simple tense with imperfective implication; PROG: progressive; PFCT: perfect; PFCT-PROG: perfect progressive; PFCTIV: the simple tense with perfective implication; OTHERS: all the non-predicate constructions, gerund, present and past participle, infinitive, noun and noun phrase, adverb, adjective, prepositional phrase, etc.

4 Discussion

As can be seen from the above results, our study has yielded valuable insights in three aspects: identifying the aspect classification features, probing into the causes of low accuracy by examining the corpus, and building the aspect classification model. We analyze the results in the following four ways.

4.1 Impact of multi-aspect alternative translations

In this experiment, over 600 sentences in the corpus with multi-aspect alternatives were identified and revised for their annotations, which raised the accuracy by about 22% (Tables 4 and 5).

The multi-aspect alternative of translation is well known in Chinese-to-English translation practice, as all the experienced and high proficient translators make a habit of avoiding the repetitions of the same structures in translation. Variety in structures is one of the ultimate goals they try to attain in translation. There are hundreds of multi-alternatively translated natural language sentences that highly undermine the accuracy improvement of data mining. However, in MT practice, this is just put aside temporarily by the MT community.

Thus, our results show that this problem needs to be addressed again in the area of data mining for Chinese-to-English MT. Most probably, there exist other similar problems with other grammatical structures in Chinese-to-English MT. How to cope with this problem has far-reaching implications for MT. Our approach to tackling this problem is only a pilot study for further discussion.

4.2 Identification of aspect classification features

Table 6 summarizes the contributions of each classification feature set and their integrated contributions to the overall precision rate. Surprisingly, the lexical aspectual features give the highest contribution with a precision rate of 74.9%, while the lowest one is the temporal feature set, 60.0%.

Although the lexical aspectual features have a significant impact on the classification performance, they can predicate only the classifications of the simple tenses with imperfective, perfective, and the non-predicate construction types, having null aspect classification functions. This is because, theoretically, aspects are primarily derived owing to the external influence of adverbials. Lexical aspectual features are the internal features of verbs, having no access to the derivation of aspects (Qu, 2008).

Theoretically, the temporal features have a greater impact on the derivation of ZHE aspect while overriding the semantic meanings of 'verb+ZHE' structure in Qu (2008) and Qu and Feng (2008) and should have a much higher accuracy. This low accuracy of 60.0% may have been caused by the misannotation of the features on the dataset, as the temporal features are difficult to define by the Phython program for feature annotation employed in this study. To some extent, the temporal features do have made some predications to the ZHE aspect classification and are identified as aspect classification features in this study.

Although the syntactic feature sets achieved the overall precision rate of only 62.1% for ZHE aspect classification, they had the predictions to the PROGRESSIVE and PERFECT with precision rates of 75.3% and 50.0%, respectively. This suggests that the syntactic feature also plays an indivisible rule in de-

termining the derivation of ZHE aspect by providing compensational meanings to Qu's derivation principle (Qu 2008; Qu and Feng, 2008). Accordingly, the syntactic features were identified as new ZHE aspect classification features. The derivational aspectual function of these syntactic features is worth further discussion for MT and aspect semantics researchers.

The integrated feature set with temporal, lexical aspectual, and syntactic features achieved the highest accuracy, 80.1% (Table 5). This shows that temporal, lexical aspectual, and syntactic features can make holistic contributions to the precision rate of the ZHE Rule. Thereby, they can be the ZHE Rule classification features, including the classifications of the transferences from Chinese-English aspect and from Chinese imperfective to English simple tenses with imperfective and perfective implications (Table 6, Fig. 2).

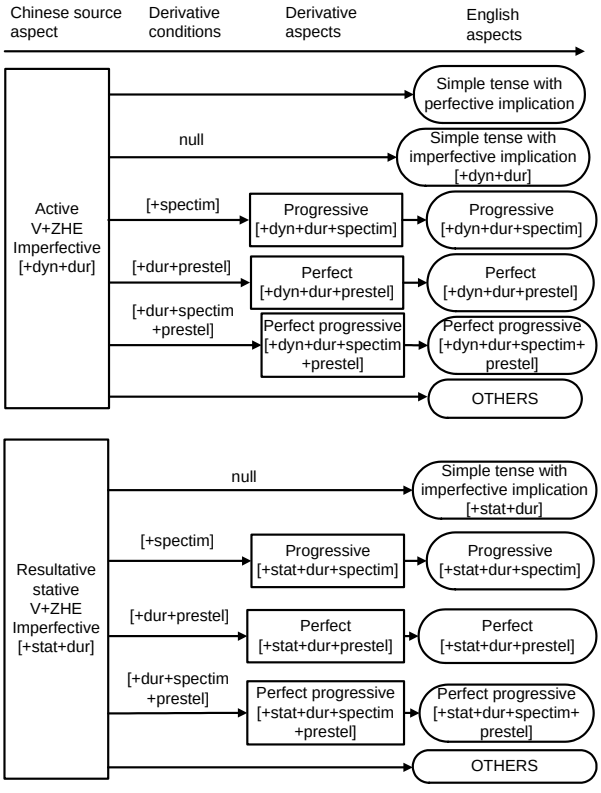

Fig. 2 New ZHE Rule from Chinese to English embodying two sub-classification rules

dyn: dynamic; dur: durative; spectim: specific time; stat: stative; prestel: present telic

4.3 Generalization of the ZHE Rule

As can been seen from Tables 5 and 7, the overall precision rate of the general classification of ZHE aspect transferring was 80.1%, tested employing J48 and justified by 10-fold cross-validation.

This suggests that the predictive result of the decision tree classification model was satisfactory and that the classification accuracy of ZHE aspect was high enough. Judging from AUC, the AUCs of three aspects, IMPERFECTIVE, PROGRESSIVE, and PERFECT-PROGRESSIVE, are high enough, being 0.808, 0.768, and 0.834, respectively. However, the AUCs of two aspects, PERFECT and PERFECTIVE, were a bit low, being 0.643 and 0.657, respectively. The overall precision rates of two sub-classifications of ZHE aspect (Table 7) are 79.6 % and 76.1%, a bit lower than those in the general classification. These results are consistent with most of the linguists' findings that ZHE is an imperfective marker while forming a collocation with 'active verb' and 'resultative stative verb'. They also indirectly show that the two sub-classifications of ZHE aspect are correct: 'active verb+ZHE' or 'stative verb+ZHE' added with some temporal adverbials will derive into progressive, perfect, and perfect progressive aspects, and accordingly be transferred into target English aspects.

As the temporal features adopted from Qu (2008)'s ZHE Rule can predicate the ZHE aspect classifications and are identified as aspect classification features (Section 4.2), the derivative principle embedded in Qu (2008)'s ZHE Rule is justified to be correct, with statistical support. Accordingly, the derivative process involving the derivative conditions in the form of temporal features is displayed in the new ZHE Rule.

Therefore, based on this promising classification accuracy, classified target aspects, and the original temporal features of the source aspects in Qu (2008)'s ZHE Rule justified in this study, a new ZHE Rule embodying two sub-classification rules with the transferring conditions (Fig. 2) can be generalized.

Like Qu (2008)'s ZHE Rule, the revised ZHE Rule with two sub-rules involves two paths of the transferring process, the direct translation path and the indirect transferring path. However, the revised ZHE Rule involves more target English aspect categories, with six aspects in the first sub-rule and five aspects in the second sub-rule. As the decision tree is a black box, it can only holistically justify that the ZHE Rule is correct, and that the lexical and syntactic

features are the classification features for the ZHE Rule. It cannot expose which specific lexical aspectual and syntactic features can be the transferring conditions for which specific target aspects. Consequently, neither the lexical aspect nor syntactic features can be labeled on the derivative conditions of the rule.

Compared with Qu (2008)'s ZHE Rule, this ZHE Rule is formulated on the corpus-based statistical data manipulated by machine learning methods with 10-fold cross-validation. Hence, justified to be valid, the ZHE Rule containing the aspect derivation process proposed by Qu (2008) has been provided with statistical support. Based on this rule, provided with the transferring conditions in the form of specific temporal adverbials, the source ZHE aspect can be transferred to one definite target English aspect. It is expected that the rule will have some implications for MT research and practice. The derivative process embedded in the ZHE Rule also reveals semantic relationships between the Chinese temporal adverbials, the verbs, and the aspect marker ZHE. Now that the rule has been tested with high accuracy, Qu's hypothesis for aspect derivation is confirmed.

4.4 Perfective meanings of ZHE aspect associated with active verbs

Table 7 shows that the 'active verb+ZHE' construction is classified into target English simple tense with perfective implication, suggesting that this construction can be the perfective aspect with telic meanings. This is supported by 67 examples retrieved from our corpus, accounting for 2.5% of the total corpus.

This result is inconsistent with Qu (2008)'s ZHE Rule and the generally acknowledged view that ZHE is an imperfective aspect marker (Smith, 1991; 1997; Chen, 2003; Xiao and Mcenery, 2004).

Below are two examples selected from our corpus:

(28) Cai Zhen qiangZHE shuo, she le ta duimian de Su Lun yi yan. (*Broke in* Tsai Chen, glancing quickly across at Su Lun.)

(29) Wo jiu fucong duoshu. Tang Yunshan youshi qiangZHE shuo. (*Interjected* Tang Yun-shan hastily, "I'll accept the decision of the majority.")

According to the 5th edition of Xiandai Hanyu Cidian (p.1096), 'qiang' associated with ZHE is a verb and its second semantic meaning is "qiang xian, zhengxian. E.g., qiangZHE shuo le ji ju; dajia dou qiangZHE canjia yiwu laodong (vie for, compete for. E.g., vie for a few words; all of us vied with each other to take part in voluntary labour.)". Given this ZHE's semantic meaning, the verb associated with ZHE, 'qiang', means that a durative action 'qiang' starts, and ends with the result of 'xian' in the end, having the telicity in semantic meanings. Therefore, 'qiangZHE' denotes perfective situations.

In light of Olsen (1997)'s view, English simple past tense may denote perfective situations while the verbs are specified for telicity. The target English translations of 'qiang' are 'break in' and 'interject', with the durative periods and telicity in semantic meanings. Thereby, a perfective implication is generated for 'break in' and 'interject' in the above two examples.

Thus, judging from the telicity specified by the Chinese source verbs associated with ZHE, 'qiang' and its target English translations, 'break in' and 'interject', ZHE can be affiliated with verbs with telicity, denoting the perfective situations.

5 Conclusions

We have carried out this experiment by selecting the features, investigating the misclassified corpus sentences and calculating the accuracy of the aspect classification using a decision tree with 10-fold cross-validation.

Based on the above findings, five tentative conclusions can be drawn as follows:

1. The multi-alternative of aspect translation may be the primary cause for lower accuracy of Chinese-to-English grammatical structure transferring, judged by this aspect transferring study. This needs to be proved in further studies in MT.

2. The syntactic and temporal features having impact on ZHE aspect derivations can be employed as ZHE aspect classification features, while the lexical aspectual features have no predictions for ZHE aspect derivation. This is a significant contribution to research in aspectual semantics and the automatic aspect classifications, but needs further theoretical explanation.

3. The ZHE Rule, a linguistic rule, is generalized and evaluated with high accuracy. This linguistic rule

reveals the one-to-many matching relationships among the source ZHE aspect, the transferring conditions, and the target English aspects. Hence, it has the feasibility of application in MT practice. It also exposes the semantic relationships between Chinese temporal adverbials, the verbs, and their affiliated ZHE aspect marker. Consequently, the generalization of the rule may throw some light into Chinese aspect semantics research.

4. While associated with active verbs, ZHE aspect can denote a perfective situation.

5. Decision tree, a machine learning method, can be used for testing the aspect transferring rule from Chinese to English, provided that the corpus for testing is preprocessed and the features are prescribed by experienced experts.

In short, this pilot study, using a machine learning method to indirectly generalize the linguistic rule and prove its application to MT, is an innovative study in natural language processing, having practical implications for Chinese-to-English MT, Chinese aspectual semantics studies, and aspectual studies.

This study has some limitations in its corpus scale, the types of machine learning method, and the labeling of the lexical aspectual and syntactic features on the transferring conditions of the ZHE Rule. Only 2633 pairs of sentences with 'verb+ZHE' and one machine learning method, decision tree, were employed. As the decision tree is a black box, it can justify only that the ZHE Rule is correct, but cannot expose which specific lexical aspectual and syntactic features can be the transferring conditions for which specific target aspects. Therefore, further studies need to be conducted based on an enlarged corpus, comparison with other machine learning methods, and additional statistical methods.

Acknowledgements

Many thanks to Dr. Huan-huan CHEN, associate Prof. Bao-bao CHANG, Dr. Wen-qin SHANG, and Prof. Tony HARTLEY for their help with this research.

References

Chen, C.C., 2009. Ambiguity of LE in Chinese: the perfective as well as imperfective. *J. Chin. Ling.*, **37**(1):108-129.

Chen, G., 1990. On the Usage of "ZHE" and Its Comparison with English Progressive. *In*: Yang, Z.J. (Ed.), Collected Papers on English-Chinese Contrastive Study. Shanghai Foreign Language Education Press, Shanghai, China, p.379-391 (in Chinese).

Chen, Q.R., 2003. A Study on Chinese Aspectual System. PhD Thesis, Central China Normal University, Wuhan, China, p.142-152 (in Chinese).

Cheng, J.H., Dai, X.Y., Chen, J.J., Wang, Q.X., 2004. Processing of tense and aspect in Chinese-English machine translation. *Appl. Res. Comput.*, (3):79-80 (in Chinese).

Dai, Y.J., 1997. A Study on the Aspectual System in Modern Mandarin Chinese. Zhejiang Education Publishing House, Hangzhou, China (in Chinese).

Institute of Linguistics in Chinese Academy of Social Sciences, 2005. Xiandai Hanyu Cidian (5th Ed.). The Commercial Press Ltd., Beijing, China, p.1096 (in Chinese).

Jin, L.X., 2003. The grammatical meaning of "S+le" structure and its syntactic requirements. *Lang. Teach. Ling. Stud.*, **100**(2):38-48 (in Chinese).

Li, W.J., Wong, K.F., Cao, G.H., Yuan, C.F., 2004. Applying Machine Learning to Chinese Temporal Relation Resolution. 42nd Annual Meeting of the Association for Computational Linguistics, p.582-588.

Lin, Z., 2004. The Grammaticalization of "Le1" from a Marker of Perfective to a Marker of Tense. *In*: Jing, C. (Ed.), Hanyu Shiti Xitong Guoji Yantaohui Lunwenji. Baijia Publishing House, Shanghai, China, p.86-102 (in Chinese).

Ljungqvist, M., 2007. LE, GUO and ZHE in Mandarin Chinese: a relevance-theoretic account. *J. East Asian Ling.*, **16**(3):193-235. [doi:10.1007/s10831-007-9012-6]

Lu, J.M., 1999. A supplementary comment on "ZHE". *Chin. Lang.*, **272**(5):331-335 (in Chinese).

Olsen, M., 1997. A Semantic and Pragmatic Model of Lexical and Grammatical Aspect. Garland Publishing, Inc., New York and London, p.163-191.

Olsen, M., Traum, D., van Ess-Dykema, C., Weinberg, A., 2001. Implicit Cues for Explicit Generation: Using Telicity as a Cue for Tense Structure in a Chinese to English MT System. Proc. Machine Translation Summit VIII, p.34-41.

Pan, W.G., 2004. A Study on the Aspects in Mandarin Chinese from a Perspective of the English Translation of "Le". *In*: Jing, C. (Ed.), Hanyu Shiti Xitong Guoji Yantaohui Lunwenji. Baijia Publishing House, Shanghai, China, p.139-153 (in Chinese).

Pen, L.Z., 2007. A Study on the Modern Mandarin Chinese Modes. China Social Science Press, Beijing, China, p.209-309 (in Chinese).

Qian, N.Y., 2000. Aspect mark "zhe" not express "progressive" meaning. *Chin. Lang. Learn.*, (4):25-30 (in Chinese).

Qu, Y.H., 2008. A Chinese-English Comparative Study of Viewpoint Aspects: a Bilingual Corpus-Based Approach. Science Press, Beijing, China, p.74-84 (in Chinese).

Qu, Y.H., Feng, Z.W., 2008. A study on the derivations of viewpoint aspects in Mandarin Chinese. *J. Zhejiang Univ.*

(Human. Soc. Sci.), **38**(4):173-181 (in Chinese). [doi:10.3785/j.issn.1008-942X.04.020]

Qu, Y.H., Tao, T.J., Xu, C.Z., Feng, Z.W., 2008. The Formalization of 'Temporal Adverbials+ZHE Imperfective' Sentences. IEEE Int. Conf. on Natural Language Processing and Knowledge Engineering, p.375-382.

Robson, S.Y., 2005. The temporal relations and aspects expressed by the particle LE in Mandarin Chinese. *J. Chin. Ling.*, **33**(2):333-365.

Shang, X., 2007. The Comparative Studies between Chinese and English Aspect Systems. Shanghai People's Publishing House, Shanghai, China (in Chinese).

Shi, J.S., Hu, X.P., 1998. The "BA" structure with verbs followed by "ZHE". *Lang. Teach. Ling. Stud.*, (4):39-49 (in Chinese).

Shi, Y.Z., 1992. Aspects in modern Chinese. *Soc. Sci. China*, (6):183-201 (in Chinese).

Smith, C., 1991. The Parameter of Aspect. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Smith, C., 1997. The Parameter of Aspect (2nd Ed.). Kluwer Academic Publishers, Dordrecht, the Netherlands, p.273-277.

Soh, H.L., 2009. Speaker presupposition and Mandarin Chinese sentence-final-Le: a unified analysis of the "change of state" and the "contrary to expectation" reading. *Nat. Lang. Ling. Theory*, **27**(3):623-657. [doi:10.1007/s11049-009-9074-4]

Sun, Y.J., 2007. Studies of the Mandarin Chinese Aspect System. Heilongjiang People's Publishing House, Haerbin, China (in Chinese).

Vendler, Z., 1957. Verbs and times. *Philos. Rev.*, **66**:143-160.

Wang, K.F., 2004. The Construction and Application of the Bilingual and Parallel Corpus. Foreign Language Teaching and Research Press, Beijing, China, p.110 (in Chinese).

Wong, S.M., Li, W.J., Wong, K.F., 2005. A Framework for Modeling and Representing Temporal Discourse Structure. Proc. IEEE Int. Conf. on Natural Language Processing and Knowledge Engineering, p.213-218. [doi:10.1109/NLPKE.2005.1598737]

Wu, F.X., 2005. On the non-obligatory status of the Mandarin aspect markers "le" and "zhe". *Contemp. Ling.*, **7**(3):237-250 (in Chinese).

Xiao, X.Q., 2002. A comparative study on the functions of "ZHEN (ZAI)", "ZAI" and "ZHE". *Ling. Study*, (4):30-37 (in Chinese).

Xiao, Z., Mcenery, T., 2004. Aspect in Mandarin Chinese: a Corpus-Based Study. *In*: Abraham, W., Noonan, M. (Eds.), Studies in Language Companion Series 73. John Benjamins, Amsterdam/Philadelphia, the Netherlands, p.181-205.

Ye, Y., Zhang, Z., 2005. Tense Tagging for Verbs in Cross-Lingual Context: a Case Study. Proc. 2nd Int. Joint Conf. on Natural Language Processing, p.885-895.

Ye, Y., Fossum, V.L., Abney, S., 2006. Latent Features in Temporal Reference Translation. Proc. 5th SIGHAN Workshop on Chinese Language Processing, p.48-55.

Ye, Y., Schneider, K.M., Abney, S., 2007. Aspect Marker Generation in English-to-Chinese Machine Translation. Proc. Machine Translation Summit XI, p.521-527.

Zhao, S.K., Shen, J.X., 1984. Mandarin "LE" and its target English expressions. *Stud. Lang. Ling.*, **6**(1):114-126 (in Chinese).