The expression of stance in Mandarin Chinese:
A corpus-based study of stance adverbs

Haiyang Ai

304 Sparks Building, University Park, PA 16802, United States
Department of Applied Linguistics, The Pennsylvania State University
hua126@psu.edu

_____

Abstract

Stance-taking is considered as one of the fundamental properties of human communication (Jaffe, 2009). It is pervasive, intersubjective, and collaborative. While a good deal of research has investigated the expression of stance in English, much less has been done in Chinese. In this study, we draw upon the five-million-word Academia Sinica Balanced Corpus of Modern Chinese to investigate a comprehensive set of 34 stance adverbs expressing certainty, likelihood, attitude, and style (Biber, 2006b) across various modes of communication, diverse genres, and media channels. In addition, we also zero in on a pair of synonymous stance adverbs i.e. dique vs. queshi, to illustrate the subtle and nuanced functional and distributional differences across different genres and registers. Implications for Chinese dictionary compiling, pedagogy and sentiment analysis are discussed.

Keywords

Stance adverbs, Mandarin Chinese, corpus-based, register variation, near synonym

_____

1   Introduction

Taking a stance towards the content of our utterances or the propositions of our interlocutors is considered as one of the fundamental properties of human communication (Jaffe, 2009). In fact, Jaffe (2009) argues that there's no such thing as a completely neutral position, even neutrality itself indicates one's stance. Biber (2006b) suggests that personal stance reflects how certain we feel towards the truthfulness of a preposition, how we obtain access to the information, and what our perspective we are taking. Du Bois (2007) maintains that stance-taking helps us to assign values to objects of interests, position ourselves, calibrate alignment with our interlocutors, and evoke the presupposed system of sociocultural value and ideologies. As such, the understanding of stance is closely related to the understanding of social aspects of human conduct. Research has shown that the expression of stance carries such pragmatic functions as hedging (Hyland, 1996), mitigating (Fraser, 1980; Homes, 1984), showing evident (Chafe, 1986), indicating appraisal (Martin & White, 2005), and evaluating propositions (Hunston & Thompson, 2000).

Stance can be expressed by a range of lexico-grammatical features including grammatical devices, value-laden word choice, and paralinguistic devices (cf. Biber, et al. 1999, pp. 966-969). The use of adverbials is a common way to mark personal stance (Biber & Finegan, 1988). Biber *et al.* (1999) described and characterized four groups of stance adverbs: certainty (e.g. absolutely, definitely), likelihood (e.g. possibly, probably), attitude (e.g. surprisingly, amazingly), and style (e.g. frankly, honestly). Based on such

classification, Biber and colleagues (e.g. Biber, 2006a, 2006b; Biber and Conrad, 2000; Helt, 1997) have conducted a series of studies examining the expression of stance across different genres and registers. For instance, Biber and Conrad (2000) compared adverbial markings of stance in speech and writing. Biber (2006a, 2006b) examined the expression of stance in American university spoken and written registers. Helt (1997) analyzed stance adverbial variation in spoken American English.

While a large number of studies have been conducted on the expression of stance adverbs in English, relatively few efforts have been made in examining stance-taking in Chinese, a typologically distant language from English. In this study, we draw upon the five-million-word Academia Sinica Balanced Corpus of Modern Chinese to systematically investigate a comprehensive set of stance adverbs across different communicative modes, diverse genres, and various media channels. In addition to the macro-level analysis, we also zero in on a pair of synonymous adverbs at a more micro-level in order to illustrate the subtle and nuanced differences across different genres and registers. In section 2, I outline the literature on stance-taking in Mandarin Chinese, and the benefits of the corpus-based approach. I then describe the corpus data and research questions. In section 4 I report and discuss the different variations of stance adverbs followed by a presentation of the findings of a pair of synonymous stance adverbs. I conclude the article with a discussion of the implications for Chinese dictionary compilation, pedagogy, and sentiment analysis.

2   Stance-taking in Mandarin Chinese

In Mandarin Chinese, adverbs are used to express time, attitude, manner, frequency, or duration. They are often placed after the subject or after the topic if no subject is present, or at the beginning of sentence (Li & Thompson, 1981). In (1), the epistemic adverb *yiding* "defintely" expresses the sense of certainty of willingness to cooperate, and was placed immediately after the subject *women* "we":

(1)   *mei wenti, women **yiding**   quanli   peihe*
         NEG problem we   definitely full-power cooperate
         "No problem. We will definitely fully cooperate (with you)."

Only a few studies have examined the expression of stance in Mandarin Chinese. Adopting the conversation-analytical framework, Wu (2004) examined the expression of stance from the use of final particle *a* and *ou* in the unfolding development of talk-in-interaction. Wang, Tsai, and Yang (2009) examined two stance adverbs *qishi* ('actually') and *shishishang* ('in fact') in spoken discourse. They found that these two adverbs have face-saving and intersubjectivity functions, and are used more frequently in situations where politeness is expected, such as TV or radio interviews. Hsieh (2009) investigated the use of stance adverbs in press reportage. She reported that journalists often make strategic choices of epistemic stance markers to achieve special power. Other studies have taken a contrastive perspective. Zhen (2008) followed Biber's (2006b) classification of stance adverbs and examined epistemic (i.e. certainty and likelihood) stance adverbs in a Chinese-English parallel corpus. She found that although there were English translations for Chinese stance adverbs, the positions and linguistic forms were not always consistent. Long and Xu (2010) compared stance adverbs in Chinese EFL learners' English and Chinese argumentative essays of a shared topic. They reported that learners' use of stance in English has strong correlation with the stance use in Chinese. Taken together, these studies have contributed to our understanding of the expression of stance in Mandarin Chinese by examining different linguistic devices from different theoretical and methodological perspectives. However, a

systematic analysis of a comprehensive set of stance adverbs in Mandarin Chinese based on large-scale corpus data has yet to be done, which is the focus of the present study.

3   The corpus-based approach

The use of authentic and attested corpus data is likely to introduce more rigor into theory testing than using introspective and isolated data (Channell, 2000). Norrick (2009, p. 865) posits that the use of electronic corpora gives pragmatic research "a broader, more secure basis". A central feature of corpus-based studies is the focus on frequency and distribution (Gries, 2009). Within the field of corpus linguistics, it is well recognized that differences in frequency imply differences in function and use (cf. Firth, 1957; Gries, 2010). In other words, the distributional patterns of a linguistic item reveal its semantic and functional properties. The relationship between meaning and distribution has been captured nicely by Harris (1970):

> [I]f we consider words or morphemes *A* and *B* to be more different in meaning than *A* and *C*, then we will often find that the distributions of *A* and *B* are more different than the distributions of *A* and *C*. In other words, difference of meaning correlates with difference of distribution. (Harris, 1970, p. 785; cited in Gries, 2010, p. 122)

The role of frequency has also been recognized in usage-based linguistics, cognitive linguistics, and construction grammar (cf. Leech, 2011). A general underlying assumption is that the more frequent a linguistic expression is used, the more likely it is to be entrenched in cognition. On the other hand, it is also generally agreed that the identification of such frequent use of linguistic items and associated register or genre variation is difficult to come by based on native speaker's intuition. However, this is relatively easy to achieve by using the corpus-based approach. While some studies (e.g. Long & Xu, 2010; Zhen, 2008) used corpus data in examining the expression of stance in Chinese, they seemed to focus more on the comparison of stance-taking from a contrastive or second language acquisition perspective. In this study, we will draw upon large-scale corpus data to systematically investigate a comprehensive set of stance adverbs across different genres and registers.

4   The Sinica Corpus

The corpus data used in this study were retrieved from the five-million-word Academia Sinica Balanced Corpus of Modern Chinese (The Sinica Corpus thereafter) [1]. The Sinica Corpus was designed to be a representative corpus of modern Mandarin Chinese, containing texts from different communicative mode (written, spoken, written-to-be-read, written-to-be-spoken, spoken-to-be-written), genre (narration, argumentation, exposition, and description), media (newspaper, academic journals, conversation or interview, etc.). Table 1 shows its composition by media. All texts in the corpus are word segmented and part-of-speech tagged. This allows us to identify all the stance adverbs by grammatical category.

Sources	Words (per 10,000)	Percentage
Newspaper	162.57	31.28%
General magazine	151.80	29.18%
Journal article	3.62	0.70%
Textbook	21.25	4.08%
Reference book	0.70	0.13%
Thesis	7.06	1.36%
General book	43.96	8.45%
Audio/Visual	118.80	22.83%
Conversation/Interview	8.50	1.63%
Lecture	1.32	0.25%
Miscellaneous	0.53	0.10%

Table 1. *The Sinica Corpus as Organized by Media*

The stance adverbs examined in this study were selected from previous studies of stance adverbs, notably from Biber (2006b) and Zhen (2008), as well as consultation with the reference thesaurus dictionary (Mei, Zhu, Gao, & Yin, 1983). We have also taken account of limitations of the tool used in this study, i.e. the online version of the Sinica Corpus via its web-based interface. While the web-based interface is a convenient tool for exploring the distributional information of the corpus data, it does not allow for more complicated phrasal or Regular Expression search. Hence it is not possible to search for *zai...chengdu shang* "to … degree or extent". Second, we selected these stance adverbs with an intention to make comparison to Biber's (2006b) study, whose classification of stance adverbs was also followed in this study. Third, the list of stance adverbs could be overwhelming long if multi-word expression and constructions were to be included in the study. While it is certainly desirable to include as much stance adverbs as possible, considering that each stance adverb requires dozens of searches to determine its frequency in different communicative modes, diverse genres, and media channels, a line has to be drawn at some point in order to make the project manageable. In total, 34 stance adverbs have been included in this study (see Appendix B). Frequency counts were all normalized to per million words, and mean values of normalized frequency were used for comparison.

5   Result and discussion

5.1   Register variation of stance adverbs

The overall distribution of stance adverbs in Mandarin Chinese shows considerable variation across four categories: certainty, likelihood, attitude, and style adverbs occur 188.5, 158.8, 117.7, and 77.2 times per million words respectively. Epistemic stance adverbs i.e. indicating certainty and likelihood are relatively more frequent, while style adverbs are less so in our corpus data. In addition, the epistemic certainty adverbs occur more frequently than likelihood adverbs. These findings are consistent with Biber's (2006a) study of English stance adverbs in university registers.

Considerable variation has been found in the distribution of stance adverbs across different genres. Specifically, stance adverbs are extensively used in narration, but seldom in description (see table 2). This might not be surprising necessarily, as description of an

object or an event tend to be objective and impartial, while narration of an event tends to afford more opportunity for speakers to convey their opinions, judgments, or evaluation.

	Certainty	Likelihood	Attitude	Style
Argumentation	19.3	14.0	11.4	7.7
Narration	76.1	79.0	63.1	35.7
Exposition	34.3	25.7	15.0	18.4
Description	1.2	1.9	1.1	0.2

Table 2. *Distribution of Stance Adverbs Across Different Genres*

Within the genre of narration, a number of stance adverbs stand out as having higher frequency than other members of the same category. For instance, *yiding* and *bixü* are pervasive in expressing certainty, while *keneng* is frequently used in expressing likelihood, and *qishi* emerged as a popular choice for marking personal attitude. These three stance adverbs are illustrated in examples (2)-(4).

(2)   *zhou xiansheng* **yiding** *haobuchiyi de   shuo: "wo   hen   kuaile"*
       zhou Mr   surely no-doubt   GEN say: "1SG very happy".
       "Mr. Zhou will surely say: 'I am very happy'".

(3)   *jiu   zai women* **bixü** *likai lundun de   qian   yi tian,*
       NOM zai 1PL   must leave London GEN before one day
       "Just before the day we left London,"

(4)   *ta   **qishi** zaoyi   zhidao le   da'an*
       1SG in fact already know   PRF answer
       "In fact, she already knew the answer"

Within each genre, large variation was observed among different groups of stance adverbs. Attitude and style adverbs, for instance, occur less frequently than certainty and likelihood adverbs. This seemed to suggest that the expression of epistemic stance (i.e. certainty and likelihood) is more of a concern than the way they are expressed (i.e. style).

        Next we explore the distributional patterns of stance adverbs across different communicative modes: written, written-to-be-read, written-to-be-spoken, spoken, and spoken-to-be-written. Such differentiation—according to the Sinica Corpus manual—is theoretically motivated. Specifically, written refers to the typical sense of written documents as in corpus construction. Written-to-be-read, however, refers to documents that are essentially written for the purpose of giving a speech or lecture. It is argued that this type of text is different from common spontaneous speech, given that it is often carefully constructed and edited. Written-to-be-spoken refers to scripts for plays or dramas. They are written in such a manner as to mimicking real spoken interaction. Because they are repeatedly rehearsed in advance, they are said to be somewhat different from real spontaneous speech. Spoken-to-be-written refers to texts such as notes taken at a conference. Although they are spoken, they are often post-edited, and thus warrant a separate category in the corpus construction. Finally, spoken genre consists of transcripts of spontaneous speech.

	Certainty	Likelihood	Attitude	Style
Written	117.5	87.7	68.6	55.0
Written to be read	119.7	88.6	69.0	55.6
Written to be spoken	138.6	119.1	89.2	61.2
Spoken	19.0	29.3	20.0	6.2
Spoken to be written	137.9	117.5	88.2	61.4

Table 3. *Distribution of Stance Adverbs Across Different Communicative Modes*

The five different modes of communication exhibit different distributional patterns across the four groups of stance adverbs. The spoken register has relative infrequent occurrences of all four groups of stance adverbs (See Table 3). This finding contradicts with Biber's (2006a, 2006b) research where he reported that spoken interaction generally has more instances of stance adverbs than written texts in university language. While it might be tentative to attribute this to the relatively small component of spoken data (7.3%) in the Sinica Corpus, a closer examination suggests that both written-to-be-spoken (0.82%) and spoken-to-be-written (0.36%) have even less proportion than spoken data in the Sinica Corpus, yet they have more instances of stance adverbs than the spoken register (note that all frequencies were normalized to per million words).

Another point to note is that likelihood adverbs exceed certainty adverbs in spoken register. This also contrasts with Biber's (2006b) study where he found the reverse was true in all spoken registers in American university settings. Clearly, studies with larger samples of spoken data are needed to account for this difference.

In contrast to spoken register, written register exhibits a different distributional pattern whereby certainty adverbs occur more than likelihood adverbs. A potential explanation is that spontaneous spoken interaction tends to be more casual, containing various hedging devices, while written texts often have stronger binding effects, thus containing more certainty adverbs. Biber (2006a) pointed out that written language is generally considered as "more precise and accountable than speech" (p. 114).

	Certainty	Likelihood	Attitude	Style
Newspaper	7.7	5.5	4.4	3.7
General magazine	53.3	40.7	31.9	22.1
Academic journal	0.4	0.7	0.2	0.1
Textbook	6.1	4.7	2.5	2.5
Reference book	0.2	0.1	0	0
Thesis	2.3	5.0	0.9	3.1
General book	22.4	10.4	14.0	3.7
Audio/Visual	42.5	46.1	30.2	25.3
Conversation interview	3.0	6.0	5.1	1.4

Table 4. *Distribution of Stance Adverbs Across Different Media*

With respect to media channels, considerable variation was observed in the distribution of four categories of stance adverbs (see Table 4). In general, stance adverbs occurred frequently in magazines, general books, and audio/visual content, such as TV/Radio recordings and online forum discussions. This finding is consistent with Wang et al. (2009)

study in which they reported that certainty adverbs were more frequent in occasions where politeness is expected, such as TV or radio interviews. In other words, publications intended for a wider readership or audience such as popular magazines, books, or TV shows tended to employ more adverbs marking personal stance for the purpose of making it attractive. Conversely, we noted that such adverbs were almost nonexistent in sources like academic journals and reference books, exhibiting a "faceless stance" (Biber & Finegan, 1989, p. 108). The relative scarcity of stance adverbs in textbook is also line with Biber's (2006a) finding that English stance expressions in university text book were rare. This is because in these registers "factual reporting" (Biber, 2006a, p. 113) is more of the norm. In newspaper register, we found that certainty and likelihood adverbs were more frequent than attitude and style adverbs. This finding is consistent with Hsieh's (2009) study in which she reported that journalists tended to use epistemic i.e. certainty, likelihood, stance markers to achieve special power.

5.2 Near synonym: *dique* vs. *queshi*

The corpus-based approach is not only useful in revealing the general distribution of stance adverbs across diverse registers and genres, but also helpful in enabling researchers to examine near synonyms (Arppe & Jarvikivi, 2007; Liu, 2010; Tao, 2003; Xiao & McEnery, 2006). In this section, I will show how the corpus-based approach can help reveal subtle and nuanced differences between near synonyms of stance adverbs in Mandarin Chinese. In order to provide a succinct analysis within the confines of a publishable paper, I selected one pair of stance adverb, *dique* vs. *queshi*, meaning "indeed", for illustration.

In *Xiandai Hanyu Cidian* "The Contemporary Chinese Dictionary" (Lü, 1973), *queshi* is used in defining *dique,* suggesting that they are interchangeable and have the same denotational meanings. In what follows, however, I will show that the corpus-based approach offers us more insights into the differences between these two seemingly synonymous stance adverbs.

By examining the corpus data, we found important formal and syntactic differences between *dique* and *queshi.* First, we found that 14 out of 75 instances of *dique* were placed at sentence-initial position, followed by a comma, as illustrated in (5), while none of the 128 instances of *queshi* was used in this way.

(5)   ***dique****, wo ziji ye you he le banzui de jingyan*
      Actually, 1SG self too have drink PRF half-drunk GEN experience
      "Actually, I myself have the half-drunk experience too."

This seemed to suggest that the adverb *dique* is capable of functioning at a more global level within the discourse than *queshi*, which tends to occur at a more local level, typically in closer proximity to the verbs it modifies. Further, *dique*'s capability of being stand-alone signals the tendency of being grammaticalized as a discourse marker, expressing the speaker's subjective evaluation, i.e. certainty towards the prepositions of event being communicated. Second, while *queshi* was found to have modal auxiliaries such as *xü,* or *bixü* "must", *dique* did not have such modal auxiliary collocates. Third, collocational behavior is another interesting aspect in the corpus-based study of near synonyms (Xiao & McEnery, 2006). In our corpus data, two most common verb collocates for *dique* and *queshi* are the existential *you* and the copula *shi*, the two of which accounted for more than one-third of the total occurrences of all the two collocates of the two stance adverbs. The copula *shi* collocates more frequently with *dique* than *queshi* while existential *you* occurs

more frequently with *queshi* than *dique* (see Table 5). This nuanced difference seems to suggest that the two seemingly synonymous stance adverbs have differential semantic preference, i.e., *dique* accentuates the accuracy of the situation while *queshi* tend to emphasize the existence of a verbal effect.

	you	*shi*
dique	15.1%	26.0%
queshi	18.6%	14.8%

Table 5. Distribution of Most Frequent Collocates for *dique* and *queshi*

In (6), *queshi* was used to emphasize the existence of the impact brought to the developers. In contrast, *dique* was used to describe the accuracy or truthfulness of using exhibition as a way to educate young people, as in (7). These subtle and nuanced functional differences are typically not readily accessible to native speaker intuition, nor are they available in Chinese dictionaries or grammar books.

(6)   *dui guonei   zhengdui DOS shichang kaifa   de   yezhe,*
        To domestic targeting DOS market   development GEN businesses
        ***queshi** you suo   yinxiang*
        queshi you some impact
        "There are indeed some impacts on businesses that specifically targeting domestic DOS market"

(7)   *zhanlan **dique** shi henhao   de   tuiguang jiaoyu   fangshi*
        exhibition dique COP very-good GEN promote education method
        "Exhibition is indeed a very good way to promote education."

In addition to the functional differences, there were also important register and genre differences between the two stance adverbs. With respect to communicative mode, while the two adverbs had similar frequency, *queshi* occurred slightly more frequently than *dique*. In general, the frequency of both adverbs in spoken mode were markedly lower than other communicative modes, a finding that is similar to the overall picture of stance adverbs discussed earlier.

	dique	*queshi*
Written	78.9	80.13
Written to be read	79.7	81.56
Written to be spoken	87.7	90.75
Spoken	8.4	9.40
Spoken to be written	87.5	91.16

Table 6. *Distribution of dique and queshi by Mode*

The distributional pattern of stance adverbs *dique* and *queshi* across different genres is given in Table 7. As can be seen, description category has the least amount of occurrences for both stance adverbs. In addition, *dique* and *queshi* occurred more frequently in

argumentation than description. This is not surprising necessarily because both adverbs carry subjective stance, a characteristic that is generally more common in argumentative genre. In other words, the expression of confidence, assurance, or certainty is of vital importance to argumentative genre, whose audience needs to be convinced and persuaded.

	dique	*queshi*
Argumentation	116.2	86.4
Narration	76.1	55.0
Exposition	98.3	172.4
Description	47.6	20.4

Table 7. *Distribution of dique and queshi by Genre*

An important distributional pattern to note between these two adverbs was that *dique* occurred more frequently than *queshi* in argumentation, narration, and description, but not exposition. The preference for *queshi* over *dique* in exposition seemed to be a distinctive stylistic feature. This interesting difference warrants future research focusing on exposition texts in particular. On the other hand, the certainty adverb *queshi* occurred frequently in exposition. An example is given in (8). As can be seen, in this informative sentence, *queshi* has been used in explaining the effects or significance of an act that has changed the appearance of the whole city. *dique*, on the other hand, occurs more frequent than *queshi* in the description genre. In (9), the author described a beautiful scene between the prince and princess. The stance adverb *dique* was used to further accentuate the fact that they had good time together.

(8)   *zhe-ge   faling* **queshi** *gaibian le*
        this-CLF act   indeed change PRT
        zhengge dushi de   rongmao
        whole   city GEN appearance
        "This   act   has   indeed   changed   the   appearance   of   the   whole   city."

 (9)   *wangzi yu gongzhu* **dique** *duguo le*
        prince and princess indeed spend PRT
        yiduan   tianmi de   rizi
        one-section happy GEN days
        "The prince and princess have indeed spent a wonderful time together."

Finally, distributional differences between *dique* and *queshi* in terms of media channels were also observed (see Table 8). In general, *dique* occurred more frequently than *queshi* in academic journals, textbooks, and general books, but not in thesis. In comparison among the different media, we observed that both adverbs occurred more frequently in general magazine, general book, and audio/visual materials than newspaper and conversation/interview. This is perhaps because certainty or assurance, as conveyed by the two stance adverbs, is subjective in nature and is a common trait of publication for general purposes. On the other hand, the sense of objectivity is usually a more desirable quality in registers such as academic writing.

	dique	queshi
Newspaper	16.6	14.8
General magazine	125.2	139.7
Academic journal	55.3	27.6
Textbook	47.1	23.5
Reference book	0	0
Thesis	0	99.2
General book	179.7	72.8
Audio/Visual	100.2	84.2
Conversation interview	25.5	23.5

Table 8. Distribution of *dique* and *queshi* by Media.

In sum, by examining substantive corpus data, we have identified important functional and distributional differences between the near synonymy *dique* and *queshi*. Functionally, *dique* are more frequently found to be at sentence-initial position with a more global scope, modifying the whole sentence; by contrast, *queshi* tend to occur at sentence-medial position and function more locally and in closer proximity with the main verb. In addition, we found that *dique* tended to collocate more often with copula *shi*, emphasizing the accuracy or truthfulness of the verbal action; *queshi*, on the other hand, tended to collocate more frequently with existential *you*, accentuating the existence of a certain verbal effect. In addition, our distributional analysis shows that *dique* occurred more frequently than *queshi* not only in spoken interaction and written text, but also in argumentation, narration, description, with the exception of exposition. These subtle and nuanced differences between the two seemingly synonymous adverbs are generally beyond native speaker intuition, and are not available in dictionaries or grammar books in Mandarin Chinese, but can be fruitfully exploited by the corpus-based approach.

6   Conclusion

In this study, we have conducted a corpus-based investigation of stance adverbs in Mandarin Chinese, by examining a comprehensive set of 34 adverbs expressing certainty, likelihood, attitude, and style across different communicative modes, diverse genres, and media channels, in the five-million-word Sinica Corpus. In addition, we have investigated a pair of synonymous stance adverbs and showed the nuanced functional and distributional differences that are otherwise inaccessible from introspection or existing reference books.

The compilation of dictionaries and reference books has increasingly benefited from the use of authentic and attested data from linguistic corpora (Walter, 2010). While much has been done in English (e.g. COBUILD; Longman Dictionary of Contemporary English), few efforts have been made in Mandarin Chinese (but see Xiao, Rayson, & McEnery, 2008 for an example of corpus-based frequency dictionary of Mandarin Chinese). This study has hopefully shown that the corpus-based approach is particularly resourceful in unraveling important linguistic information such as register variations that are not readily accessible by native speaker intuition.

Another area the research findings have useful implication is research on sentiment analysis [2]. In general, sentiment analysis seeks to determine the attitude of a speaker or a writer with respect to the overall contextual polarity of a document, i.e. positive, negative or neutral (cf. Pang & Lee, 2008; Liu, 2010). In addition, sentiment analysis also seeks to identify the subjectivity versus objectivity of a given text or sentence (Pang & Lee, 2004; Su & Markert, 2008). Because the stance adverbs examined in this study were capable of

The expression of stance in Mandarin Chinese: A corpus-based study of stance adverbs

showing personal stance at various levels in terms of certainty, they can be important words for sentiment analysis. Thus, sentiment analysis will benefit from taking into consideration the subjective evaluation, judgment, or opinion towards a proposition by factor in such subjective stance expressed by various epistemic stance adverbs. For instance, the epistemic certainty adverb *yiding* "surely" in (10) reveal a strong opinion towards the proposition that with such good background, the subject *ta* "he" will be even more successful (in whatever he is doing).

(10) *you name hao de   beijing,   ta **yiding** hui geng chenggong*
      exist that   good GEN background, 3SG surely will more successful
      "With such good background, he will surely be even more successful."

The research findings also have important pedagogical implications to the teaching of synonyms to L2 learners of Chinese. Xiao and McEnery (2006) noted that the explanation of meanings to learners by offering synonyms should be treated with caution, because even near synonyms can and usually do differ in collocational behavior. Using synonyms to teach vocabulary tends to emphasize the denotational meaning of words rather than their usage, and might be a potential trap for language learners (Tognini-Bonelli, 2001, p. 34). By designing concordance lines with clear examples showing the uncovered differences between near synonyms of stance adverbs, language instructors, and material and course developers can raise L2 Chinese learners' awareness of the subtleties and nuances as well as register variations of seemingly synonymous stance adverbs in Mandarin Chinese. In this way, they can use these stance adverbs more appropriately in different contexts and registers.

Biber's (2006b) classification of stance adverbs has usefully contributed to the understanding of personal stance, and has influenced a number of studies. However, as Helt (1997) noted, it is not always easy and straightforward to put adverbs into such discrete categories. Such difficulties has also evidenced in the present study of stance adverbs in Mandarin Chinese, a typologically distance language than English. Future research will benefit from extending such classification to include more categories (see Peng 2009 for nine different categories for Mandarin Chinese).

7   References
Arppe, A. and Jarvikivi, J., 2007, Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory,* vol. 3, no. 2, pp. 131-159.
Biber, D., 2006a, Stance in spoken and written university registers. *Journal of English for Academic Purposes,* vol. 5, pp. 97-116.
Biber, D., 2006b, *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
Biber, D. and Finegan, E., 1988. Adverbial stance types in English. *Discourse Processes* vol. 11, pp. 1-34.
Biber. D., and Finegan, E., 1989, Styles stance in English: Lexical and grammatical marking of evidentiality and effect, *Text,* vol. 9, pp. 93-124
Biber, D., and Conrad, S., 2000, Adverbial Marking of Stance in Speech and Writing. In S.

Hunston and G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse*. New York: Oxford University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E., 1999, *Longman Grammar of Spoken and Written English*. London: Longman.

Chafe, W. L., 1986, Evidentiality in English conversation and academic writing. In W. L. Chafe and J. Nichols (Eds.), *Evidentiality: The linguistic code of epistemology* (pp. 261-272). Norwood, NJ: Ablex.

Channell, J., 2001, Corpus-based analysis of evaluative lexis. In S. Hunston and G. Thompson (Eds). *Evaluation in text: Authorial stance and the construction of discourse* (pp. 38-55). Oxford: Oxford University Press.

Du Bois, J. W., 2007, The stance triangle. In R. Englebreston (Ed.), *Subjectivity, evaluation, interaction* (pp. 139-182). Amsterdam: John Benjamins.

Firth, J. R., 1957, *Papers in linguistics*. Oxford: Oxford University Press.

Fraser, B., 1980, Conversational mitigation. *Journal of Pragmatics*, vol. 4, pp. 341-350.

Gries, S. T., 2009, *Quantitative corpus linguistics with R: A practical introduction*. New York: Routledge.

Gries, S. T., 2010, Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal,* vol. 34, pp. 121-150.

Harris, Z. S., 1970, *Papers in structural and transformational linguistics*. Dordrecht: Reidel.

Helt, M. E., 1997, *Discourse marker and stance adverbial variation in spoken American English: A corpus-based analysis*. Unpublished doctoral dissertation, Northern Arizona University.

Holmes, J., 1984, Modifying illocutionary force. *Journal of Pragmatics*, vol. 8, pp. 345-365.

Hsieh, C.-L., 2009, Epistemic stance taking in Chinese media discourse. *Research in Theoretical Linguistics,* vol. 3, pp. 1-35.

Hunston, S., and Thompson, G. (Eds.), 2000, *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Hyland, K., 1996, Writing without conviction? Hedging in science research articles. *Applied Linguistics,* vol. 17, pp. 433-454.

Jaffe, A., 2009, Introduction: The sociolinguistics of stance. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives*. Oxford: Oxford University Press.

Leech, G., 2011, Why frequency can no longer be ignored in ELT? *Foreign Language Teaching and Research,* vol. 43, no. 1, pp. 3-20.

Li, C., and Thompson, S., 1981, *Mandarin Chinese: A functional reference grammar*. Berkely: University of California Press.

Liu, B., 2010, Sentiment Analysis and Subjectivity. In *N. Indurkhya and F. J. Damerau* (eds.). *Handbook of Natural Language Processing, Second Edition* (pp. 1-38). Boca Raton, FL: CRC Press, Taylor and Francis Group.

Liu, D., 2010, Is it a *chief, main, major, primary*, or *principle* concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics,* vol. 15, no. 1, pp. 56-87.

Long, M., and Xu, J., 2010, Comparing stance in Chinese EFL learners' English and Chinese argumentative essays of a shared topic. *Foreign Languages and Their Teaching,* vol. 3, pp. 21-24.

Lü, S. (Ed.), 1973, *Xiandai Hanyu Cidian [The Contemporary Chinese Dictionary* (2002 ed.).*]* Beijing: The Commercial Press.

Martin, J.R., and White, P.R.R., 2005, *The language of evaluation. Appraisal in English*. London: Palgrave.

Mei, J.-J., Zhu, Y.-M., Gao, Y.-Q., and Yin, H.-X., 1983, *Tongyici Cilin [Thesaurus of Synonyms]*. Shanghai: Shanghai Lexicographical Publishing House.

Norrick, N. R., 2009, Pragmatic markers: introduction. *Journal of Pragmatics,* vol. 41, no. 5, pp. 863-865.

Pang, B. and Lee, L. 2004, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Proceedings of the Association for Computational Linguistics (ACL)*. pp. 271–278.

Pang, B., and Lee, L., 2008, Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* vol. 2 no. 1-2, pp. 1–135

Peng, Y., 2009, *Studies on the Procedural Encoding Functions of Chinese Stance Adverbs*. Unpublished master's thesis, National Taiwan Normal University.

Su, F. and Markert, K., 2008, From Words to Senses: a Case Study in Subjectivity Recognition. *Proceedings of the 22nd International Conference on Computational Linguistics Volume 1* (pp. 825-832). Association for Computational Linguistics.

Tao, H., 2003, Toward an Emergent View of Lexical Semantics. *Language and Linguistics,* vol. 4, no. 4, pp. 837–856.

Tognini-Bonelli, E., 2001, *Corpus linguistics at work*. Amsterdam: John Benjamins.

Walter, E. (2010). Using corpora to write dictionaries. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. London: Rouledge.

Wang, Y.-F., Tsai, P.-H., and Yang, Y.-T., 2009, Objectivity, subjectivity and intersubjectivity: Evidence from qishi ('actually') and shishishang ('in fact') in spoken Chinese. *Journal of Pragmatics,* vol. 42, no. 3, pp. 705-727.

Wu, R.-J. R., 2004, *Stance in Talk: A conversation analysis of Mandarin final particles*. Amsterdam: John Benjamins.

Xiao, R., and McEnery, T., 2006, Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics,* vol. 27, no. 1, pp. 103-129.

Xiao, R., Rayson, P. and McEnery, A., 2008, *A Frequency Dictionary of Mandarin Chinese: Core vocabulary for learners*. London: Routledge.

Zhen, F., 2008, *Corpus-based Study of the Stance Adverbials in Chinese-English Political Texts*. Paper presented at the International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS 2008), Zhejiang University, Hangzhou.

Notes
1. The Academia Sinica Balanced Corpus of Modern Chinese is compiled by Institute of Information Science and CKIP group in Academia Sinica in Taiwan. It is available at http://www.sinica.edu.tw/SinicaCorpus/. Its online interface limits the number of hits to 2000 for users outside the Sinica Institute. This limitation, however, did not affect the present study, because the number of hits for any the stance adverbs did not exceed such limit. The function of searching by topic was not working at the time of the data collection. Thus, it was excluded from this study.
2. I thank an anonymous IJALP reviewer for drawing my attention to this issue.

Appendix A: List of grammatical terms
1SG   first person singular

1PL   first person plural
3SG   third person singular
CLF   classifier
GEN   genitive
NOM   nominalizer
PRT   particle

Appendix B: Character Version of Stance Adverbs Investigated in This Study

Certainty	一定、必然、完全、當然、絕對、無疑、必須、勢必、想必、根本、顯然、的確、確實
Likelihood	恐怕、也許、可能、或許、大概、大體、差不多、說不定、不見得
Attitude	居然、竟然、難道、畢竟、究竟、其實、反正
Style	通常、一般、主要、明明、誠然

Appendix C: Chinese character version of the examples cited.
1. 沒問題，我們**一定**全力配合。
2. 周先生**一定**毫不疑遲的說："我很快樂"。
3. 就在我們**必須**離開倫敦的前一天。
4. 他**其實**早已知道了答案。
5. **的確**，我自己也有喝了半醉的經驗。
6. 對國內針對 DOS 市場開發的業者，**確實**有所影響。
7. 展覽**的確**是很好的推廣教育方式。
8. 這個法令**確實**改變了整個都市的容貌。
9. 王子與公主**的確**度過了一段甜蜜的日子。
10. 有那麼好的背景，他**一定**會更成功。