# SPOKEN OR WRITTEN?: A CASE STUDY OF WRITTEN FEATURES IN ORAL TEXTS

Claire Hsun-huei Chang

張 郇 慧*

**摘 要**

傳統的分析認爲漢語的口語和書面語爲兩種截然不同的表現方式。本文以語料庫爲本，探討口說和書面語表達的差異，找出口語及書面語表現在辭彙、句法、語義及篇章層面上的語言特徵。並藉語料庫資料所顯示的語言特徵說明所謂的口語，書面語的分界並不是截然不同的，而是因使用者和文體的不同、口語和書面語的語言特徵互動，形成複雜的語言現象。

## 0. Introduction

The notion that there is a striking distinction between spoken and written Mandarin Chinese is widely accepted in the teaching of Mandarin Chinese as a second language. The nature and the relationship between the spoken and written languages has drawn great interest in linguistics, especially in the area of discourse analysis and language use (Tannen 1982, Chafe 1982, Lakoff 1981, for example). The subject has been discussed from theoretical, language acquisition, cultural evolution, and socio- and applied linguistics perspectives. The major interest of linguists and language teachers in these studies is the practical implications they would have on the teaching of a second language. However, most of these studies have been done in English. Our understanding of the relationship between written and spoken Mandarin is very limited. Studies in this area have been rather sporadic and concentrated on certain linguistic structures (Li & Thompson 1982, Biq 1995). The idea that there is a distinction between the spoken and written languages is the collective result of impressions held by individuals but these distinctions have never been explicitly

* 作者為本校英國語文學系副教授。

* This paper was first presented at a seminar in the Institute of Information Science, Academia Sinica and then at the First International Conference on New Technologies in Teaching and Learning Chinese in San Francisco, April 28-30, 1995. I am grateful to Yung-O Biq, Chu-ren Huang, and Lily I-Wen Su for their suggestions. I would like to thank the seminar and conference participants for their constructive comments. Thanks are also due to two anonymous reviewers for their very helpful comments on earlier version of this paper. Errors are, of course, mine. The presentation of the paper was made possible by the National Science Council travel fund. I would also like to thank CKIP group at the Institute of Information Science, Academia Sinica for their general support in providing news corpus. Special thanks go to Chan Hsiao-hui and CKIP programmers for their technical support.

explained. The teaching of the distinction between the spoken and written features of Mandarin is almost non-existent or rather sporadic at best. The subject is usually covered in the classroom with the remark "this is mainly used in written language not in spoken." The distinction in the dictionary (monolingual, bilingual, as well as learner's dictionaries) includes only the lexical distinction between spoken and written language[1] (see Pu-tong-hua Min-nan-yu Dialect Dictionary, for example). There is no mention of the distinction on a syntactic level. Nor is there any explanation of how and why a certain lexical item is considered a written or literal form. I believe that the distinction between spoken and written Mandarin is not as great as most people believe, though some features may occur in written texts more frequently than they do in spoken discourse. There is a great deal of interaction between written and spoken features and no text is completely free of either. Teachers of Mandarin Chinese should be aware of the fact that the language is complicated and should also be cognizant of its day-to-day usage. In this study, I attempt to delineate the so-called written and spoken features by analyzing a large corpus of data and noting the recurrent patterns in this body of data. I then summarize the findings and identify spoken and written features for pedagogical use. I hope that the study will provide some guidelines for Mandarin instructors in designing their curriculums. At the same time I hope this study provides the impetus for the creation of an electronic data base to be shared and used in teaching, textbook writing, and dictionary compilation purposes. In creating this data base we must bear in mind that it must be a balanced one. This balance is not to be understood as a balance of text styles or genres but rather a balance of written and spoken features.

The purpose of this paper is to provide an empirical base on lexical and syntactic differences between spoken and written language. After identifying written and spoken features, I will examine written features in oral texts. I will explain the theoretical and methodological bases of the study and their pedagogical implications. Section one presents some preliminary starting points for discussion. Section two focuses on lexical differences between spoken and written features as found in our corpus. Section three concentrates on a discussion of syntactic differences between the two modes of expression. Section four explicates the interaction between the spoken and written features in oral texts emphasizing why written features are employed in spoken discourse. Section five summarizes the study and concludes the paper by pointing out the pedagogical and linguistics implications of the study.

[1] For Xue-dian, a mono-lingual dictionary, literal and colloquial pronunciations are distinguished. There is no mention of written and spoken usage as far as lexical items are concerned. However, in the notes for editing (p.3), it does mention that the meaning of a character is explained with clear, precise spoken style [yuti wen].

## 1. Some Preliminaries

In this section, I will clarify the terms used in this paper, and delineate the methodology adopted in the study. I distinguish mode and medium (see McCarthy & Carter 1994). The medium refers to the mechanism used to communicate the message, while the mode refers to the means used by the sender of the message. In contrast to the conventional spoken and written language dichotomy, I consider written and spoken mediums and investigate how linguistic features are used in both modes. The rationale behind using spoken/written medium and mode instead of spoken/written language is to minimize the confusion that might arise from the conventional understanding of the huge gap between the two languages. In this paper, spoken or written language will refer to the message that is conveyed through spoken or written medium.

### 1.1. The distinction between spoken and written language

Biber (1986) has identified three dimensions of linguistic variation in English texts. These dimensions are edited text versus interactive, abstract versus situated content, and reported versus immediate style. The dichotomous dimension as described by Biber is actually the distinction between spoken and written language. Spoken language is interactive, situated and immediate, while written language is edited, abstract and reported. The research of Chafe (1982) and others (Tannen 1982, Lakoff 1982) also shows that spoken language is interactive, involving the speaker, and written language is organized without the participant involvement. People speak to communicate, thus we would expect to find interaction or cooperation between/among participants. Writing is a unidirectional endeavor, lacking instantaneous interaction, and thus we would not expect to find the spirit of cooperation in a written text. This is the general distinction between spoken form and written form. Its interactive and participant-involvement nature guarantees that the spoken language will be composed of simple structures with general meaning (cf. Chafe 1982), while the edited and reported nature of the written language requires elaborated forms with specific meanings. All languages that have writing systems will exhibit such a distinction. The situation in Chinese is even more complicated. Chinese involves character writing and a large number of homophones. Given the nature of character writing and the existence of homophones, I hypothesize that the written form would be shorter (fewer characters to write) with more intense semantic implications (packing more meaning

into fewer characters) than the spoken form, and that spoken forms will be longer to avoid the confusion caused by homophones. There will be more negotiating or cooperative features in spoken discourse but less so in written texts. Following this line of thought, I examine the spoken discourse and written texts in order to identify the linguistic features of the written and spoken languages.

### 1.2. Methodology

Li & Thompson (1982) have conducted a study on the differences between classical and modern Chinese. They consider classical Chinese elegant and terse, a characteristic not found in the colloquial style. They believe that there is a considerable gap between the spoken and written languages of modern China, and attribute this gap to the heavy influence of the classical style on modern written Chinese (Li & Thompson 1982: 87). While I do not deny the influence of classical Chinese on modern written Chinese, I think an empirical support for such a statement is needed. I provide an empirical foundation (explicit study) by analyzing a large base of both the written and spoken languages in order to delineate the "literary" and "colloquial" features for dictionaries. I first examine recurrent patterns in spoken discourse and identify linguistic features that may be of "written" origin. I then check these features against the written corpus of the Academia Sinica to validate the hypothesis. The criteria for determining whether a feature is written or spoken depend mainly on frequency counts of a specific linguistic feature in the corpus. If a feature shows up in the spoken corpus more frequently than in the written one, I consider it a spoken feature. Similarly, a written feature will appear in the written corpus more frequently than in the spoken one. My base corpus came from two types of texts: spoken and written. The spoken data was derived from two different sources. The first is an interpreted version of President Clinton's inauguration speech. The second is the TV program Ban Wo Chengzhang 【伴我成長 Walk with me】, which is, to a considerable extent, devoted to interviews. The interpretation is produced under great time pressure, and thus a certain spontaneity can be assumed. The interview is produced in several layers. By layers I mean that some interviews are spontaneous, while some are planned and prepared. Spontaneous interviews are conducted with young children and adults who are not experts in any particular field. The planned interviews are done with special guests who are usually experts in the topic being discussed. For such a corpus, a balance between planned and unplanned language uses, and a balance between interactional and unidirectional involvement among

speakers are maintained[2]. The written data is taken from the electronic corpus of the Academia Sinica. The written corpus is a collection of article taken from the United Daily News (聯合報) and the China Times ( 中國時報).

## 2. Data analysis: lexical analysis

In this section, I will discuss the results of my data analysis. Later I will discuss the data on lexical and syntactic levels. I will discuss the recurrent patterns in the spoken data and use this as the point of departure for comparing linguistic features used in the written and spoken mediums. Features that are frequently exhibited in the written corpus would be termed "written features," and features that are frequently employed in the spoken corpus will be termed "spoken features." These features will be discussed on three major levels: lexical, syntactic and discourse. On the lexical level I will discuss lexical classes and lexical specificity. On the syntactic level, I will discuss variations in syntactic structures in the spoken and written languages. On the discourse level, then, I will identify co-occurring linguistic features in spoken discourse.

In the following, I will discuss lexical choices in spoken and written discourse. This discussion will have three parameters: the length of a word, the frequency of usage, and semantic density.

### 2.1. Monosyllable vs. Disyllable: The Use of *gai* and *yinggai,* xu and *bixu, ji* and *yiji*

As we have mentioned earlier, the written language, compared to the spoken language. tends to utilize short forms. The data in our spoken and written corpora confirm the hypothesis that one-character words are more frequently used in the written medium than they are in the spoken. For example, we have found the use of *yinggai* (應該) and *gai* (該) in both modes. In a 10,000-character spoken corpus, *yinggai* appears 10 times and *yet gai* appears only twice. Similarly, *bixu* (必需) and *xu* (需) are completely synonymous, but *bixu* is invariably used in the spoken corpus. *Xu* was never used in the spoken corpus. In searching the written corpus at the Academia Sinica, the proportional use of the two-character word *yinggai* to the one-character word *gai* is almost 1:10. There are 2,057 instances of *yinggai* and 22,638 instances of *gai* in a corpus of 20,698,116 characters. There are 2,945 instances of

[2] Further investigation needs to be done to ensure the balance of corpus. For our present purpose, the balance as mentioned is sufficient.

*xu*, but *bixu* appears in less than 40 instances in the written corpus. There are 40,043 instances of *ji* 及 'and' but 3,768 instances of *yiji* 以及 in the written corpus. Comparing the written medium and spoken mediums, we can conclude that the number of characters is one of the criteria in determining whether a linguistic form is a written or a spoken feature.

## 2.2. Frequency of Use

I found some frequently-used expressions in our spoken corpus; unfortunately I could derive no comparable data from the written corpus since the features are not tagged in at the moment. I would, however, cite one expression as an example. Observe the use of *suo* before a verb, as in the following examples:

1. *xianzai-de* *qingshaonian* *tamen* *suo* *jieshou* *wailai-de*
   now-DE[3] youth they SUO accept foreign-DE

   *xunxi* *chongji* *tai* *duo.*
   information impact too much

   'The youth today have experienced too much impact from foreign information.'

2. *Touguo* *women* *suo* *shuo-de* *hua* *women shide*
   through we SUO speak-DE word we made

   *chuntian* *tizao* *laidao.*
   spring early come

   'Through what we have said, we have made the spring come early.'

*Suo*, as shown in the example, can be deleted without losing any meaning. The data also shows that the use of *suo* is not prevalent across all speakers. Two interpreters use *suo* in their speech; and *suo* is also a marked feature in the speech of one of the interviewees. For the interviewee that often used *suo*, I also found that he often used *suo wei de* 'what we called' in his speech. Since the use of *suo* or *suo wei de* is not a

[3] The following abbreviations are used in the example sentences: DE, possessive or subordinate marker *de*; SUO, relative marker *suo*; ASP, aspect; MW, measure word.

common speech pattern with all speakers, it is safe to say at the moment that *suo* may be a residue of classical Chinese, thus, a residue of written language. Given the limitations of this paper, I will not pursue the historical development of *suo* but will reserve it for further research.

### 2.3. The use of four-character idioms or sayings

It is generally believed that four-character idioms are a residue of classical Chinese. They pack a semantically dense content into a few characters, and thus, should be considered a written feature. Considering our definition of written features, we can also identify four-character idioms as a written feature, since this dense packing of information into a relatively few characters is an identifying feature of written language. I found a high frequency of four-character idiom use in the spoken context. Examining the frequency counts in 1445,7534 (952,9233 words) characters news corpus I found that, contrary to what we might expect, the use of four-character idioms is not that frequent. The greatest number of occurrences, according to the statistics (CKIP Technical Report No. 93-02), is 87【層出不窮】, this is a 0.0001% chance of appearance. Looking back at our spoken discourse, the four-character idiom is frequently used. In the spoken corpus of 10,000 characters, I found 37 instances of four-character idiom use. This is a rather high frequency of use.

### 2.4. The Use of Resultative Verbs

I mentioned earlier that we tend not to incorporate too much information in one spoken exchange and that the semantic density is usually low in spoken discourse (Chafe 1995). Resultative verb compounds are used to distinguish subtle semantic difference. Given the low semantic density of spoken language, it is only natural that we should find few RVCs in a spoken database. In searching a 6,000 character spoken corpus, I found 40 RVCs, of which the second elements *dao* (22 times), *diao* (5 times) and *wan* (5 times) constituted 80% of the result element. The percentage of resultative verb compounds is not as frequently used as we would have assumed in the spoken corpus. RVCs, no doubt, are an important category in Mandarin grammar[4] and are also part of native speakers' linguistic knowledge. However, in actual speech performance, we rarely found varied uses of RVCs. The most frequent use of second verb is *dao*, 'reach,' which marks the ending point of the preceeding verb. Frequently-

[4] This is evidenced by articles, thesis written on this topic. Some of the references include Hashimoto (1964), Thompson (1973), Chang (1989, 1991), Li (1990), Ross (1990), Tang (1992), just to name a few.

used first verbs include *shuo* 'speak,' *kan* 'look, see,' and *ting* 'listen to.' In searching for the first 5,000 frequently-used words in a written corpus of journal Chinese of 14,457,534 words, I found 50 resultative verb compounds with various resultative endings. The most frequently-used resultative endings (in order of appearance) are: *chu* 'out', *de* 'obtain', *dao* 'reach'. The highest appearance rate for resultative verbs is 0.15%. The average appearance rate is approximately 1%, which is about the same as the appearance rate in the spoken corpus. Thus, it appears that the resultative verb compound is not crucial in distinguishing written feature from spoken feature. Given the semantic specificity, I classify the use of RVCs as a written feature. It is speculated that the appearance of resultative verb compounds may vary with types of discourse, and the topic under discussion. Further research needs to be done to substantiate the claim that the use of resultative verb compounds is a written feature. Previous corpus-based research has shown that RVCs are an important class of verbs. Comparable to resultative verbs, our data show that syntactic resultatives constitute a large portion of the corpus. For example:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3. | *shuo* | *de* | *hen* | *shangxin...* | *xiao* | *de* | *hen* | *kaixin* |
| | speak | DE | very | sad | laugh | DE | very | happy |

'...Speak so that she is very sad, ...laugh to the extent of happiness.'

Based on the frequency count of features in spoken and written texts, we may conclude at this point that syllable length (in the case of Mandarin, the number of characters), and semantic density are two features that distinguish the spoken and written languages. In the spoken discourse, I found a high frequency of disyllabic word and four-character idiom use. In written corpus, on the other hand, I found a high frequency of monosyllabic word use.

## 3. Syntactic analysis

In this section, I will discuss various syntactic constructions to aid further understanding of linguistic features used in written medium.

### 3.1. Use of Nominalization Phrases

Oral communication tends to be spontaneous and immediate while writing is

planned and organized. The spontaneity and immediacy of oral communication is reflected in its linguistic forms. Previous research (Biber 1988, Chafe 1982) has shown that spoken language linguistic forms are short, simple, and contain less information. Conversely, written language linguistic forms are long, elaborate, and pack more information into one sentence. Still there are some nominal phrases with high information density in our spoken corpus. Consider the following examples:

4. *Dao-le* *women* *gai* *dapo* *buhaode* *jiu* *xiguande*
   come-ASP we should break not good old habit

   *shiji* *le.*
   time ASP

   'It is time for us to break the old habits that are not good.'

5. *Zhende* *shi* *you* *yi* *zhong* *henhaode* *gen* *duifang*
   really is have one MW very good with counterpart

   *jiechu-de* *yi* *zhong* *suoweide* *qiangpo-de* *goutong.*
   contact-DE one MW so-called force-DE communication

   'There is really a very good way of communicating with the counterpart, so-called forced communication.'

Again, our data show that the use of nominal phrases is user-dependent. Interpreters use this linguistic form frequently due to the influence of the original text[5]. A similar phenomenon can be observed in some of the interviewees included in this study. Interviewees who use nominal phrases are invited guests who have prepared some notes for the interview. Examine the following sentences 6-8.

6. *jiezhe* *weiji-de* *xiaochu* *lai* *jianli* *women*
   through crises-DE elimination LAI establish our

   *lishi-de jichu.*
   history foundation

[5] The orginal text (a speech) is written to be read; the influence of written feature is apparent.

'Through the elimination of crises, we establish the foundation for history.'

7. *Ta hui you hen duo de jiaolü, huangkong haipa.*
he will have very much DE anxiety apprehension fear

'He may have a lot of anxiety, uneasiness and fear.'

8. *Nei zhong xinli-shang de cuozhe, aonao huozhe*
that MW mental-on DE frustration remorse or

*bu-zhi-suo-cuo dou you yi-xie yingxiang.*
not knowing what to do all have some influence

'The mental frustration, regret and not-knowing-what-to-do would have some impact (on their understanding of why parents fight.)'

The present technical support is not sophisticated enough to search for nominal patterns and calculate the frequency in the written corpus. Based on some of the previous studies (Biber 1988, Tannen 1982a) and taken the fact that nominal phrases are difficult to process in real time situation, I consider the nominal phrase a written language influence, and classify it as a written feature. (Biber 1988 study of variation across speech and writing confirms that nominal phrases are frequently used in news writing.). Certain lexical items[6] (such as *goutong* 'communicate, communication', *fazhan* 'develop, development') rank high on the frequency count scale, and the high frequency may be the result of the nominal use of these nouns. Table 1 shows that the verb use of some lexical items prevails in the written corpus, but the noun use of these lexical items cannot be underestimated.

| lexical item | *zhuiqiu* 'to pursue, pursuit' | *yongbao* 'embrace' | *fazhan* 'to develop, development' | *xingcheng* 'to form, formation' | *touzi* 'to invest, investment' |
|---|---|---|---|---|---|
| as a verb | 432 | 63 | 7842 | 1853 | 10871 |
| as a noun | 19 | 6 | 1279 | 57 | 997 |

**Table 1 - Verb and Noun Use of Certain Lexical Items**

[6] These lexical items are cross-reference of nouns and verbs, and the frequency is the combination of instances of both categories.

### 3.2. Use of Verbs

Parallel to the use of nominal phrases, I also found the use of verbal phrases, in which a noun is used as a verb; an intransitive verb as a transitive verb. Here, categories nouns, verbs and adjectives are determined and defined primarily in terms of semantics, morphology and syntactic distribution (Tang 1992a). Dictionaries are also consulted for determining the category.

a) Nouns used as verbs:

9. *Mei-yi-dai* *meiguoren* *bixu* *yao* *dingyi* *shemo* *shi*
every generation American have to define what is

*zuowei* *yi* *ge* *Meiguoren* *de* *dingyi.*
be one MW American DE definition

'Every generation of Americans should have to define what is to be an American.'

Dictionaries[7], predominantly, give a noun category to the lexical item *dingyi*. In our written corpus, *dingyi* appears 96 times with 96 times used as a verb and not once as a noun.

b) intransitive verbs used as transitive:

In the following example, taken from the speech of one of the interviewees on the TV program *Walk with me*, the underlined lexical items are usually used as intransitive verbs or stative verbs[8] but are here used as transitive verbs.

10. *Yi* *ge* *ren* *juedui* *bukeneng* *tai* *guo*
one MW man absolutely not possible too much

*keke* *ziji* *zeguai* *ziji* *huozheshi* *erdu* *ziji.*

[7] See Pu-tong-hua Min-nan Fanyan Cidian (Xiamen University, 1982), among others.
[8] Generally, a stative verb can be preceded by a modifier hen 'very'.

unkind self blame self or vicious self
'One can never be too harsh (by restraining, blaming and abusing) on oneself.'

Interestingly enough, interpreters and invited guests on TV are found to employ nouns/adjectives as verbs more often than other people, and they are the same group of people that uses nominalization more often than others. My explanation for such a language use is that it is 'written-dependent', that is, the speaker formulates his/her message through written materials. In that regard, I consider the verbal use a written feature. However, this 'innovative' way of using language occurs in the situation where the message is conveyed through spoken language. Rarely does it occur in the written language. A lexical item would make its way into written language only when it has been used frequently enough in spoken language, the lexical item *dingyi* is the case in point.

### 3.3. Use of Varied Syntactic Structures: Use of Causative Construction *shide* 'make...' and *rang* 'let' in the Corpus

In our spoken corpus, I found a very interesting phenomenon in the use of causative construction. In one of the two interpreted versions used in this study the pattern that includes *shi* occurs 16 times while *rang* occurs 5 times. The second interpreter used *shi 2* times and *rang 8* times. On the other hand, we did not find any *shi* pattern in the TV interviews, and yet the *rang* usage appeared 13 times and the *ling* 'cause' usage once. Closer examination of *shi* and *rang* shows that *shi* tends to indicate that an inanimate subject causes a third party to do something. Thus the object of *shi* in the interpreted versions varies and includes *women* 'we' or the third party (usually it is expressed with a lexical noun, examples will be provided later). The object of *rang* tends to be a pronoun with *women* 'we' constituting the highest frequency. *Rang* conveys an invitation to work on something rather than placing restrictions, especially when the object of *rang* is *women* 'we'. Examine the following examples:

11. *Women* *shide chuntian* *tizao laidao.*
We cause spring early come

We made the spring come early.'

12. *Shide* *suoyoude* *renmen* *nenggou* *dedao* *tamen*
cause all people can get they

*yinggai* *dedaode* *baochang.*
should get-de reward

'(We) enable all of the people to get what they deserve.'

13. *Rang* *women ba* *geren-de* *sili* *fangzai* *yibian.*
RANG we BA individual-de private profit put one side

'Let us (allow us) to put aside our individual profit.'

14. *Rang* *tamen* *buyao* *zheyang* *changchang* *chaojia.*
RANG they not like that often fight

'Allow them not to fight like that often.'

15. *Ai* *shi* *yao* *rang* *duifang* *zhidao* *erqie* *yao*
love is have RANG the other party know and have

*yong* *tonglixin.*
use same reasoning

'Love is to allow the other party to know (what you feel) and to feel the same.'

As confirmed by the search of *shi(de)* and *rang* structures in the CKIP corpus, *shi(de)* (使得) is used more frequently than *rang* (讓). In a corpus of 1,770,000 characters, *shi(de)* occurs 1560 times and *rang* 758 times. It would thus be safe to say that *shi* is used in the written language while *rang* is used in the spoken language.

### 3.4. Alternation Between the Use of Preverbal *zai* and Postverbal *zai*

Our spoken corpus shows a recurrent pattern of using prepositional phrases, but

a search of the written corpus shows an opposite pattern. That is, prepositional phrases are incorporated into the verb to form a complex verb. In our written data there are 61 instances of a prepositional phrase preceding the verb *chuxian* 'to appear 出現 'and 264 instances of 出現 being incorporated into the verb *zai*. A similar phenomenon is observed in verbs *duo* 'to hide' and *zuo* 'to sit.' There are 14 instances of *duo* being used in prepositional phrases and 123 instances of the incorporated *duo-zai* form. In 16 instances *zuo* is used in prepositional phrases and in 227 instances the incorporated *zuo-zai* form appears. The data indicates clearly that prepositional use is a spoken feature, while the incorporation of verbs is a written feature.

### 3.5. Adverbial Modification and Adjectives Used as an Attribute

16. *zemoyang lai* *jiankang-de* *chaojia huozhe* *bimian* *zengjia...*
how come healthy fight or avoid increase

'...how to fight or aviod (fighting) healthily in order to improve ....'

17. *He* *xiansheng* *jiao* *women yi ge* *henhao-de* *fangshi.*
name Mr. teach we one very good way

'Mr. He taught us one good way.'

18. *Wo* *meiyou* *henhao-de* *jingji* *nengli.*
I not have very good economic ability

My financial situation is not very good.'

We have found stative verbs used as predicates and attributes in our spoken corpus and the difference in distribution is not as great as we might have thought. There are 30 instances where stative verbs are used as a predicate and 18 instances where they are used as an attribute. The attribute use of adjectives encodes information earlier in the discourse, and should be considered as a mark of planned speech, possibly an influence of the written language, as the research of Chafe (1982) and Thompson (1988) shows. Interestingly, these attributive adjectives are used by special guests. This further supports Chafe and Thompson's study that adjectives used as attributes is

a written feature. In searching the written corpus, we must identify frequently-used adjectives in order to determine the instances of attributive or predicative usage. I chose *da* (大) 'big', *duo* (多) 'many', *rongyi* (容易) 'easy', *hao* (好) 'good' as examples. The following chart shows the occurrence frequency for these adjectives used as predicate or attribute in a written corpus of 1,200, 000 words.

| lexical item | *da* 'big' | *duo* 'many | *rongyi* 'easy' | *hao* 'easy' |
|---|---|---|---|---|
| as predicate | 1049 | 151 | 1 | 290 |
| as attribute | 307 | 85 | 0 | 108 |

**Table 2 - Frequency Count of Predicative and Attributive Use of Stative Verbs**

### 3.6. Discourse Features

Similarly, we found spoken features in the text. These features are associated with interactive or affective purposes and are constrained by production considerations. These spoken features, with the exception of the use of pronouns and modal verbs, do not appear in the written corpus and is a marked feature of the spoken corpus. They include the following:

(1) the use of pronoun, especially the use of the second person pronoun
(2) the use of modal verbs and auxiliary verbs to achieve the exhortative purpose
(3) fragmented, discontinuous flow of thought
(4) discourse markers to mark informational relation
(5) loose presentation, real-time production, affective functions of personal feelings or attitudes.

These features exhibit the real-time production when there is little opportunity to elaborate or modify. This is also shown in the English data of Biber (1988) and the findings of Poole and Field (1976) and Halliday (1979). Given the limitations of this research topic, I will not elaborate on these issues. Interested readers may refer to Tannen (1982) and Biber (1988) and the references they cite for a further discussion of the spoken features. For a discussion of Chinese discourse, refer to Biq (1995), Chui (1994) and Tao (1995).

## 4. Discussion and Implication

By comparing the spoken and written corpora, I have shown that the gap between spoken and written language is not necessarily great. Rather, there is a certain amount of interaction between written features and spoken features. I have established the criteria for distinguishing written and spoken features based on a frequency count and the length of linguistic units. In this section, I will try to provide explanations for the use of written features in spoken discourse and offer suggestions for corpus linguistics and the teaching of Chinese as second language.

Our spoken corpus is of two different types of discourse: one is directed speech, that is, the production of the speech is more or less controlled by the incoming message; the second, to some certain extent, is spontaneous speech. The written features employed in these two types of discourse can be influenced by two different elements. The former (interpretation) is shaped by the structure of English, while the latter (interview) is shaped by the expertise of the speaker. The common ground between the two is that both obtain knowledge from the written language. The interpretation must follow the original text closely in order to gain time, while the interview aims to get the message across. A certain amount of information transference occurs in these two speech situations. The easiest way to present the material or message would be to rely on the source information. This reliance will take two routes: meaning transfer and form transfer. The data shows that form transfer constitutes a significant portion of the data, showing that spoken and written features co-exist in spoken corpus. Following this study, research could be done on delineating parameters that affect the employment of written or spoken feature in a context.

The use of written features generally conveys high information density. Chafe and Danielewicz (1986) found that precise lexical selection is a very difficult task and is rarely found in speech. In simultaneous interpretation, where the interpreter is under time pressure to come up with a corresponding lexical selection that will convey a meaning appropriate to the context, the precise lexical selection by definition should be even more difficult to come by compared with normal conversation. However, we have found several incidences where precise lexical items were chosen. This can be attributed to the influence of the original text or a manifestation of written features in a spoken context.

The influence of written features in the spoken corpus is immediately obvious in the frequent use of four-character idioms and emerging RVCs. This is contrary to what we might have expected and is a very interesting case of interaction between spoken and written features. A written feature has made its way into the spoken

mode and emerges as a common expression. Thus, it has gradually becoming a spoken feature. Four-character idioms are derived from old texts, and yet we use them more frequently in the spoken than in the written. In the television interviews adults use four-character idioms more often than do teenagers[9].

The implications of this study are many. (1) Pedagogically, teachers of Mandarin Chinese as second language can appreciate how complicated day-to-day uses of language are and may learn from this study that they must be conversant with the actual use of the language and should design their syllabi and teaching materials accordingly. The teaching of Mandarin cannot be limited to the exposition of linguistic rules only. Actual language use must also be incorporated in the curriculum. The syllabus should be designed in such a way that frequently-used lexical items or syntactic structures are introduced earlier in the course. For example, the findings on the use of four-character idioms has great implications for the teaching of Chinese. In most of the textbook of Mandarin Chinese language, the use of four-character idioms is taught in the second or third-year. In fact, the kind of idioms used in ordinary daily conversation have never been taught. The high frequency of idioms in our spoken data requires that we rethink our current curriculum. Students need to be exposed to the frequent use of four-character idioms, assuming that most of them approach the language through spoken mode. Similarly, the use of RVCs or syntactic resultatives has implications for textbook compilation. We need not over-emphasize the importance of RVCs, and should not bombard students with all kinds of different RVCs in the beginning of the curriculum. On the contrary, we need to make sure that they are able to understand the most frequently used and most productive resultative endings. The finer aspects of resultative endings can be elaborated and taught at a later stage (see Light 1979). (2) Theoretically, the analysis and the results of this study show that there are interactive factors in any domain of the language and thus provide a piece of evidence for advocating linguistic theories that take on an interactive and modular view. (3) Pragmatically, the results provide evidence for corpus linguistics that a balanced corpus can be viewed as the balance between the spoken and written features in any text rather than the balance between texts rendered in spoken and written mediums. The corpus data is a very useful tool for language research and teaching. Teachers could use it in searching for linguistic features to explain semantic and structural information of a certain lexical item or syntactic pattern.

Due to the time and space limitations, I was not able to include a broad range of

[9] The observation needs to be further confirmed, since the teenagers' speech constitutes a very small amount of data. However, the acqusition facts show that children acquire idioms late in their language development.

texts in this study. Rather this is an attempt to show that the distinction between spoken and written language is not as great as is generally believed. Su (1995) study confirms that there is a certain amount of interaction between spoken and written features in modern fiction.

## 5. Conclusion

This study is the first attempt to observe the interaction between written and spoken features in different spoken contexts (mode). I would assert that the dichotomous distinction between written and spoken language, as is usually claimed for Chinese, is not valid. Rather, speakers employ written features in their speech, depending on the nature of the discourse and their expertise. The other logical consequence of this study is the implication for the corpus linguists. We may reconsider the general definition of balance as not a balance in terms of balancing different texts, but rather, a balance in terms of features. Previously, in order to observe the language and draw valid conclusions from the language data, a large and "balanced" corpus was required. Such a corpus should encompass texts of various genres and styles to be considered balanced. Now, we may build a "balanced" corpus in terms of balance between spoken and written features. The implications of this study indicate a clear path toward new pedagogical practice. That is, instructors of Chinese, when designing a language course, need to take actual language use, the distinction between written and spoken features in different contexts, into consideration and design their courses accordingly. The data in a large corpus allows teachers to search a large database for language use. This should be greatly encouraged. Learners can also benefit from the research results. Our data show that lexically, there is a principle of simplicity. I found a higher frequency of shorter lexical items in written texts. Syntactically, structures that denote cooperation are found frequently used in spoken discourse. At the same time, complex and planned structures do exist in spoken situations. That is, there are a fair amount of written features in the spoken context.

## References

Biber, D. 1985. Investigating macroscopic textual variation through mutli-feature/multi-dimensional analyses. Linguistics 23.337-360.

Biber, D. 1986. Spoken and written textual dimensions in English: resolving the contradictory findings. Language 62.384-414.

_____. 1986. On the investigation of spoken/written differences. Studia Linguistica 40.1-21.

Biber D. 1988. Variation across speech and writing. Cambridge: Cambridge University Press.

_____, and Edward Finegan. 1986. An initial typology of English text types. In Corpus Linguistics II. ed. by Jan Aarts and Willem Mijs. Amsterdam: Rodopi. 19-46.

Biq, Y. O. 1995. Chinese causal sequencing and yinwei in conversation and press reportage. Paper presented at Berkeley Linguistics Society.

Chafe, Wallace L. 1982. Integration and involvement in speaking, writing, and oral literature. In Tannen ed. Spoken and written language: exploring orality and literacy, 35-54.

_____. 1995. Discourse, consciousness and time. Chicago & London: The University of Chicago Press.

Chafe, Wallace L, and Jane Danielewicz. 1986. Properties of spoken and written language. In comprehending oral and written language, ed. by Rosalind Horowitz and S. J. Samuels. New York: Academic Press.

Chang, Claire Hsun-huei. 1989. Compounds in Mandarin Chinese: with special emphasis on resultative verb compounds. Linguistics Department Working Papers 20.59-84. Honolulu: University of Hawaii at Manoa.

Chang, Claire Hsun-huei. 1991. Interaction between syntax and morphology: a case study of Mandarin Chinese. Ph.D. Dissertation: University of Hawaii at Manoa.

Chui, Kawai. 1994. Information flow in Mandarin Chinese discourse. Ph.D. Dissertation: National Taiwan Normal University.

CKIP Technical Report No. 93-02. 1993. Corpus-based frequency count of words in journal Chinese: corpus-based research series no. 2. Taipei: Insitute of Information Science, Academia Sinica.

CKIP Technical Report No. 93-03. 1993. The most frequent verbs in journal Chinese and their classification: corpus-based research series no. 3. Taipei: Insitute of Information Science, Academia Sinica.

CKIP Technical Report No. 93-04. 1993. The most frequent nouns in journal Chinese and their classification: corpus-based research series no. 4. Taipei: Insitute of Information Science, Academia Sinica.

CKIP Technical Report No. 93-05. 1993. Part-of-speech in Mandarin Chinese (2nd ed): corpus-based research series no. 5. Taipei: Insitute of Information Science, Academia Sinica.

Halliday, Michael A. K. 1979. Differences between spoken and written language: some implications for literacy teaching. In Communication through reading: Proceedings of the 4th Australian Reading Conference, vol 2, ed. by Glenda Page, John Elkins, and Barrie O'Connor, pp. 37-52. Adelaide, S.A.: Australian Reading Association.

Hashimoto, Anne Y. 1964. Resultative verbs and other problems. POLA No. 8.

Lakoff, Robin Tolmach. 1982. Some of my favorite writers are literate: the mingling of oral and literate strategies in written communication. In Tannen ed. Spoken and written language: exploring orality and literacy, 239-260.

Li, Charles, and Sandra Thompson. 1982. The gulf betwen spoken and written language: a case study in Chinese. In Tannen ed. Spoken and written language: exploring orality and literacy, 77-88.

Li, Ya-fei. 1990. On V-V compounds in Chinese. Natural Language and Linguistic Theory 8.177-207.

McCarthy, Michael, and Ronald Carter. 1994. Language as discourse: perspectives for language teaching. New York: Longman Publishing.

Murry D. 1988. The context of oral and written language: a framework for mode and medium switching. Language in Society 17.3.351-373.

Poole, Millicent E., and T.W. Field. 1976. A comparison of oral and written code elaboration. Language and Speech 19.305-311.

Pu-tong-hua Min-nan Fanyan Cidian [Putonghua and Min dialect dictionary], Overseas ed. 1982. Ed. by Mandarin Dialect Research Group, Graduate Institute of Chinese Languages and Literatures at Xiamen University. Hong Kong: Shenghuo, Dushu, Xinzhi Sanlian Bookstore, Hong Kong Branch.

Ross, Claudia. 1990. Resultative verb compounds. Journal of Chinese Language Teachers Association 25.43-64.

Su, Lily I. W. 1995. Written or spoken?: some observation on the spoken text in modern fiction. Paper presented at the First International Conference on New Technologies in Teaching and Learning Chinese. April 26-28, 1995, San Francisco.

Tang, Ting-chi. 1992a. *Hanyu de cilei: huafen de yiju yu gongyong* [Part-of speech in Mandarin: criteria and function]. In Studies on Chinese morphology and syntax: 2, 59-92. Taipei: Students Book Co., Ltd.

Tang, Ting-chi. 1992b. The structure, function and origin of V-C compound verbs in Chinese. In Studies on Chinese morphology and syntax: 4, 95-164. Taipei: Student Book Co., Ltd.

Tannen, Debroah. ed. 1982a. Spoken and written language: exploring orality and literacy. Norwood, NJ: Ablex.

Tannen, Debroah. 1982b. Oral and literate strategies in spoken and written discourse. Language 58.1-20.

Tao, Hongyin 1993. Units in Mandarin: discourse and grammar. Ph.D. dissertation: University of California, Santa Barbara.

Thompson, Sandra. 1973. Resultative compounds in Mandarin Chinese: a case for lexical rules. Language 49.361-379.

Thompson, Sandra. 1988. A discourse approach to the cross-linguistic category 'adjective'. In Hawkins, J. ed. *Explanations for languae universals*, pp. 167-185. Oxford: Basil Blackwell.

Xue Dian [Learned Dictionary]. 1991. Taipei: San Min Bookstore Company.