

John Benjamins Publishing Company



This is a contribution from Chinese Language and Discourse 8:1
© 2017. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/content/customers/rights>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Semantics and chunking in written and conversational discourses

A corpus study of two near-synonymous words in Mandarin

Danjie Su

University of California, Los Angeles

Although much has been written about the differences between written and conversational discourses, less work has been done on how these two discourse types differ in terms of chunking patterns. This study investigates the different meanings and chunking patterns two words have in Mandarin written and conversational discourses. To overcome the problem of comparability between written and conversational corpora, instead of using a single word, I use two near-synonymous Mandarin words, *zhihou* and *yihou*, both of which mean roughly 'after' or 'later,' and compare their meaning and chunking patterns in written and spoken corpora. The investigation regarding semantic distinctions revealed that in both writing and conversation, *zhihou* favors past and *yihou* favors future, and that in writing but not in conversation *zhihou* is more often used with immediate high transitivity actions and causal relations, whereas *yihou* is more often used with low transitivity states. Regarding chunking patterns, whereas conversation preserves different stages of chunking, written discourse mainly has the final clear-cut stage. This study demonstrates the importance of grounding grammatical investigations on discourse types and of the possible usefulness of using near-synonymous words or grammatical constructions as a way of getting round the problem of comparability.

Keywords: discourse type, writing, conversation, semantics, chunking, grammar, Mandarin, corpus

1. Introduction

Previous studies have shown that written discourse can be very different from conversational discourse (Biber 1988, 1999, Clancy 1982, Tao 1999). In this paper, I

wish to pursue this theme by focusing on the question of how particular words in Mandarin can have different semantic meanings and chunking patterns in written and conversational discourses.

One way to investigate this question might be to compare the meanings of the same word in written and conversational discourses. To do this, one would need to have two comparable balanced corpora: one written with balanced samplings in different written discourse types, and the other conversational with balanced samplings in different conversational discourse types. However, this approach is not practical as there are no such comparable corpora for Mandarin Chinese at the current stage. In fact, it is questionable that there could be two truly comparable corpora, one written and one conversational.

To solve this problem, instead of using a single word, this study chooses a pair of near-synonymous words and compares their semantic distinctions and chunking patterns in written and conversational corpora. The aim is to reveal the different semantic meanings and formulaic chunks the two words have in written and conversational discourse.

2. Theoretical preliminaries

This study takes as its starting point the idea that studies of grammatical patterns should take into consideration discourse types (Miller & Weinert 1998, Hopper 1998, Tao 1999, Jing-Schmidt & Tao 2009, Iwasaki 2015, etc.). In outlining *Multiple Grammars*, Iwasaki (2015) notes that lexical and grammatical items in one component grammar (such as written) are subject to expansion or reclassification into another component grammar (such as conversation).

For operational purposes, this study uses the term “discourse type,”¹ defined as a set of written or spoken language data that has been produced in a communicative medium, for a certain communicative purpose, and as a result, has certain conventionalized linguistic configurations. For instance, news editorials differ from telephone conversations in that each is a discourse type that has its own communicative medium, purpose, and most importantly, conventionalized linguistic configurations. The number of discourse types can be numerous when the factor of register is taken into consideration (Tao 1999). Nevertheless, at least two major discourse types, written and spoken, should be distinguished. On top of that, more

1. Various authors have discussed the differences among register, genre, and style. However, their differences are not the focus of the current study. Interested readers may refer to Jing-Schmidt and Tao (2009) for a conceptual definition of genre, or Biber and Conrad (2009) for a systematic account on the differences between register and genre.

specific written discourse types such as news and fiction as well as specific spoken discourse types such as telephone conversation and face-to-face conversation can be further distinguished.

The differences between written and spoken discourses have been consciously studied during the past decades (e.g., see Biber 1988 for an early account). Overall, it has been noticed that grammatical structures in spontaneous spoken language are different from those in written language (Miller & Weinert 1998: 1) and that some basic notions developed from written or constructed data often do not account well for conversational data (Tao 1996). With regard to the differences between written and spoken discourses, it has been found that spoken discourse has a flexible word order (Clancy 1982, Tao 1996) and features lower degrees of transitivity (Thompson & Hopper 2001) and higher degrees of epistemic stance (Thompson 2002). In recent years, with the awareness of the differences among discourse types being raised, we have witnessed fruitful empirical investigations based on data of specified discourse types (Tao & Meyer 2006, Fox & Thompson 2007, Sohn 2010, among others). The effectiveness of these early investigations calls for further exploration of a more realistic picture of grammar based on type-specified discourse data.

Some studies have shown that grammatical units can take different configurations in written and spoken discourses. For example, Pawley and Syder (1983: 554–556) find that some grammatical units in English have different shapes and properties in conversation than in writing. Tao and Meyer (2006) find that grammatical operation ‘gapping’ appears in writing but not in conversation. Jing-Schmidt and Tao (2009) compare the use of *ba*-constructions and *jiang*-constructions with the use of written and spoken corpora. They find that these two constructions share the basic meaning of disposal but differ from each other in terms of subjectivity and emotionality. They also find that spoken discourse prefers *ba* disposal constructions over *jiang* disposal constructions. Because of the rare use of *jiang* in the spoken corpus (11 “quasi”² cases of written usage in the one-million-word corpus), it was not possible for Jing-Schmidt and Tao (2009) to conduct a comparison of the near-synonymous *ba* and *jiang* constructions in the spoken data. In this connection, Biber (1999) compares the distribution patterns and usages of the English *that*-clauses (e.g., *I hope that...*) and *to*-clauses (e.g., *I hope to...*) in conversation, fiction, newspaper language, and academic prose. A total of four related grammatical constructions were investigated using corpus data: *that*-clauses headed by verbs and by nouns, and *to*-clauses headed by verbs and by nouns. Biber

2. According to Jing-Schmidt and Tao (2009), “all 11 of the instances of *jiang* constructions are found in the same source: Beijing Ren (Zhang & Sang 1986), which is a collection of narrative oral histories edited by two writers and published in print.”

(1999) finds that the characteristic uses of each type are conditioned by register. He thus argues that “most functional descriptions of a grammatical feature will not be valid for the language as whole,” and functional studies on grammar “must consider the patterns of use in several registers.”

The above studies provide important and insightful findings on configurations of syntactic constructions in written and spoken discourses. However, it still remains critical to examine configurations of lexical items in written and conversational discourses to gain further understanding on how these two major discourse types differ from each other.

With regard to chunking, previous studies have noted that formulaic chunks can have different degrees of fixedness. It is difficult to detect formulaic forms (Wray 2002). One of the reasons is that even core idioms display some degree of variability (Grant 2003). For example, the English fixed expression *Have a nice day* may be modified to *Have a really nice day* or *Have a great day* (Van Lancker, 2004: 13; see also Bladas 2012). Coulmas (1981: 2–3) notes that a great many of highly conventionalized pre-patterned expressions are “either on the brink of lexicalization or have turned into fixed idiomatic units of the lexicon already.” There are different kinds of formulaic chunks such as conversational routines and idiomatic verbal phrases. Bladas’s (2012) analysis of Catalan formulaic language indicates that prototypical conversational routines (such as *bon profit* ‘enjoy your meal’) undergo increase of subjectivity, whereas idiomatic verbal phrases (such as *fer campana* ‘to play truant, lit. to make bell’) do not. By carefully examining utterances English speakers produced in audio- and video-taped conversations, Fox and Thompson (2007) come to the conclusion that speakers have different syntactic organizations – some syntactic structures are entirely pre-stored, others partially pre-stored and partially composed, others not at all pre-stored. This raises interesting questions regarding chunking patterns in conversation. Although these previous studies on formulaic language have provided valuable insights into the complexity of chunking in spoken discourse, very little work has been done to compare the chunking pattern in conversational discourse with that in written discourse. It is hoped that the current study will help fill this gap.

3. Method

3.1 Overall research design

This paper argues that a way to solve the problem of not having comparable corpora is to, instead of using a single word, use a pair of near-synonymous words or grammatical constructions. The research design of this study is to first compare

the semantic distinctions of the two near-synonymous words in written data and in conversational data respectively, and then to compare the formulaic chunks these two words form in written data with that in conversational data.

3.2 The pair of near-synonymous words chosen: *zhihou* and *yihou*

The pair of near-synonymous words chosen is 之后 *zhihou* and 以后 *yihou* in Mandarin Chinese. These two words have meanings similar to the English words “later, after, afterwards, thereafter” (Hui 2002).

- (1) 一 周 之 后
 yi zhou **zhihou**
 one week later
 ‘One week later.’
- (2) 一 周 以 后
 yi zhou **yihou**
 one week later
 ‘One week later.’

These two words were chosen because when used to refer to time, they are considered near-synonyms in most Mandarin dictionaries (Table 1), including the largest one – 汉语大词典 *Hanyu Da Cidian* ‘A Grand Chinese Dictionary.’ For expository convenience, in Table 1 *zhihou* is abbreviated as Z, and *yihou* as Y.

The original Chinese definitions are given in Table 1 to show that in most dictionaries Z is used to define Y and vice versa, indicating that the two words are near-synonymous.

Most existing studies on *zhihou* and *yihou* focus on their parts of speech and syntactic environments. Although very few studies (but see Yi 1997, Zhao 2014) have been dedicated to revealing the difference between these two near-synonymous words, there have been some debates over the subtle semantic distinctions between the two words, which will be reviewed presently.

Some previous studies claim that these two words differ in the sense that *yihou* can only refer to temporal relations, but *zhihou* can describe both temporal and spatial relations (Li 2004). The examples cited in previous studies, which are imaginary, mostly illustrate the usage of *zhihou* to describe concrete spatial relations. Example (3) is such an imaginary example from Li (2004):

Table 1. Dictionary definitions of *zhihou* and *yihou*

Year	Dictionary	Definition
2007	当代汉语词典 (Gong 2007) Contemporary Chinese Dictionary	Z means 某个时间以后(Y)。 Y means 现在或某一时间之后(Z)。
1999	汉语大词典 A Grand Chinese Dictionary	Z means 某个时间后面。上文所说的事情以后(Y)。 Y means 比现在或某一时间晚的时期。
2004	现代汉语规范词典 (Li 2004) A Standard Dictionary of Modern Chinese	Z means 某个时间后面。上文所说的事情以后(Y)。 Y means 现在或所说某时之后(Z)的时期。
2004	新华汉语词典 Xinhua Chinese Dictionary	Z means 某个时间后面。上文所说的事情以后(Y)。 Y means 从现在或所说某时算起往后的时期。
2002	现代汉语辞海 Modern Chinese Cihai	Z means 某一时间后面。 Y means 紧接在某一确定时间之后(Z)。
2000	HSK 汉语 8000 词词典 HSK Guideline: 8000 Words	Y means 比现在或某个时间晚的时间。 Z: Z 和 Y 区别在于 Z 不能单独用。
2000	HSK 汉语水平考试词典 (Shao 2000) Dictionary for HSK	Y means 比现在或某个时间晚的时间。 Z means 某个动作或某个时间的后面; 动作发生在上文所说的事情以后(Y); 顺序在后。

- (3) 大厅 之后 才 是 饭厅
 dating **zhihou** cai shi fanting
 living room behind EM COP dining room
 ‘The dining room is behind the living room.’³

In Example (3), *zhihou* is used to describe a concrete spatial relation between the living room and the dining room.

Whereas some argue that *yihou* can only refer to temporal relations, others argue that *yihou* can also describe spatial relations, i.e., *yihou* is a time-spatial expression (Yi 1997, Zhao 2014). Example (4) is from Yi (1997):

3. Gloss abbreviations:

2SG	2nd person singular	EM	emphasis marker
3SG	3rd person singular	PRT	particle
NEG	negator	PASS	passive
BA	<i>ba</i> constructions	CLF	classifier
ASSO	associative marker	PFV	perfective
NMLZ	nominalizer/nominalization	COP	copular verb
RUC	resultative compound		

- (4) 第二排 以后 的 座位 是 留 给 二年级 的。
 di'erpai yihou de zuowei shi liu gei ernianji de
 second.row behind ASSO seat COP reserve for second.grade EM
 'The seats behind the second row are reserved for the second grade.'

In Example (4), *yihou* is used to describe a concrete spatial relation regarding the seats.

Some claim that unlike *yihou*, *zhihou* cannot be used for future time (Yi 1997, Zhou et al. 2004). *Zhihou* is also believed to be preferred in written language (Zhou et al. 2004). Although it is commonly accepted that *yihou* can be used in a free-standing position (e.g., Yi 1997, Zhou & Zhang 2005, Wang 2009), there are debates over whether *zhihou* can be free-standing (e.g., see HSK Guideline, 2000 and Yi 1997 for two opposing points of view.)

These existing contradictory views are based either on introspective intuition or some examples from written texts. Although the primary aim of this study is to compare the differences between written and conversational discourses in terms of semantic meanings and chunking patterns of lexical items, in the course of conducting such a study, the distinctions between these two words will also be revealed.

3.3 Corpora

This study started by coding a comprehensive set of semantic features of the immediate semantic and syntactic environments the two words are in. After the coding was done, a statistical analysis was run to determine with regard to what semantic features the two words differed from each other. This study is both corpus-based and corpus-driven. It is corpus-based in that it is based on a comprehensive coding system. It is corpus-driven in that it does not start with a hypothesis or a priori assumption, but rather lets the distinctive categories emerge through statistical analysis of all the categories captured in the coding system.

The methodology involves use of three corpora with specified discourse types: Two balanced written corpora – the Lancaster Corpus of Mandarin Chinese (LCMC) (McEnery & Xiao 2004) and the 1st Edition of UCLA Written Chinese Corpus (UCLA) (Tao & Xiao 2007), as well as a conversational corpus: the Mandarin telephone conversational corpus CallFriend (CallF) (Canavan & Zipperlen 1996).

The sizes are 1 million tokens for LCMC, 0.7 million for UCLA (1st Edition), and 0.2 million for CallF. LCMC and UCLA each contain five hundred 2,000-word samples of written texts and fifteen text categories (i.e., discourse types). LCMC collected samples in the early 1990s, and UCLA 1st Edition between 2000 and 2005. CallF is a collection of spontaneous telephone conversations among

friends and family members of native Mandarin speakers. The conversations were exchanged between people in China and people in North America.

In this study, the written corpora used are “balanced” with fifteen written discourse types, yet the conversational corpus consists only of one single discourse type – ordinary telephone conversation. Such a selection of data, however, does not affect the validity and reliability of this study. This is because this study is not comparing the usage of a single word in two different datasets, but is comparing the usage of two near-synonymous words in each dataset. In fact, I argue that comparing a pair of near-synonymous words in each dataset is a more reliable way to overcome the problem that hardly any written and conversational corpora are comparable.

Due to the high difficulty of building conversational corpora, at the current stage there are very few large-scale conversational corpora and barely any strictly balanced spoken Chinese corpora. Even a most comprehensive one, the Lancaster Los Angeles Spoken Chinese Corpus (by Xiao and Tao), is far from “balanced” as some of the written corpora such as the Lancaster Corpus of Mandarin Chinese (LCMC) and the 1st Edition of UCLA Written Chinese Corpus. At the current stage, there is not an existing balanced Chinese conversational corpus that is strictly comparable to an existing balanced written Chinese corpus. Therefore, it is questionable as to how valid it is to compare the usage of the same word in any existing written Chinese corpora with that in any existing spoken Chinese corpora. For example, if one were to compare the usages of the word X in a written corpus that contained a singular genre (such as news report) and in a telephone conversational corpus (such as CallFriend⁴), and one concluded that the word X was mostly used for past events in written discourse whereas mostly used for future events in spoken discourse, this claim could be well questioned. This is because news reports typically report past events, whereas in the telephone conversational corpus CallFriend, conversational partners often seize the first fifteen minutes of charge-free time to make future plans.

3.4 Coding of data

A total of 450 *zhihou* and 521 *yihou* tokens were retrieved from the two written corpora LCMC and UCLA, and 44 *zhihou* and 407⁵ *yihou* from the conversational

4. CallFriend is a collection of spontaneous telephone conversations among friends and family members of native Mandarin speakers. The conversations were exchanged between people in China and people in North America.

5. A total of 419 tokens of *yihou* were found in the CallFriend corpus, 12 of which were eliminated from the dataset used in this study, for the reason that they involved problematic hearing of the transcriber.

corpus CallF. Two corpus software tools were used: AntConc 3.3.0⁶ and UAM Corpus Tool.⁷ AntConc was used to retrieve collocation patterns, and UAM was used to code the sequences and analyze the distributions and semantic distinctions of the two words. The statistical analysis was conducted using UAM and Statistical Package for the Social Sciences (SPSS 16.0).

Following Sinclair's (2003: 12) coding of collocates, the words *zhihou* and *yihou* are called the "node (N)." The words to the left of the node word are called the N- sequence, and the words to the right are called the N+ sequence. Example (5) provides an illustration of what the N- sequence, the N, and the N+ sequence refer to.

- (5) N-sequence N N+sequence
 早餐 之后, 白素 便 问 秋林, 谁 叫 昆华。
 zaocan zhihou Baisu bian wen Qiulin, shui jiao Kunhua
 breakfast after Baisu then ask Qiulin who is.called Kunhua
 'Immediately after the breakfast, Baisu asked Qiulin who Kunhua was.'

In Example (5), the node (N) is *zhihou* 'after.' The N- sequence is *zaocan* 'breakfast,' and the N+ sequence includes the elements that follow immediately after *zhihou*: *Baisu bian wen Qiulin, shui jiao Kunhua* 'Baisu asked Qiulin who Kunhua was.'

Sequences containing the target words *zhihou* and *yihou* are coded on three dimensions: (1) certain semantic features of N- sequence, (2) the relation between N- sequence and N+ sequence, and (3) certain semantic and structural features of N+ sequence. Table 2 provides an illustration of the coding system, followed by a detailed description of the coding criteria.

The first coding dimension concerns the semantic features of the N- sequence. Distinctions were made regarding the following:

1. Whether this sequence describes a past event / situation (i.e., coded as "Past"; e.g., 她安顿他们吃饭之后 *ta andun tamen chifan zhihou* 'after she settled them to eat'), a future event / situation (i.e., coded as "Future"; e.g., 2021 年以后 *2021 nian yihou* 'after the year 2021'), or an event / situation in general (i.e., coded as "General"; e.g., 吃甜食之后 *chi tianshi zhihou* 'after one eats sweeties').
2. Whether the sequence denotes time (i.e., coded as "Time"; e.g., 6 点以后 *6 dian yihou* 'after 6:00 pm'), event(s) (i.e., coded as "Event"; e.g., 她安顿他们吃饭之后 *ta andun tamen chifan zhihou* 'after she settled them to eat'), or only contains a pronoun and the target word *zhihou* / *yihou* (i.e., coded as "Pronoun"; e.g., 这之后 *zhe zhihou* 'after this').

6. Corpus software developed by Laurence Anthony (2011).

7. Corpus software developed by Mick O'Donnell (2007).

Table 2. Coding of *zhihou* and *yihou* sequences

Coding	Features	N- sequence	N	N+ sequence
N- sequence	past_event	她安顿他们吃饭 settled them to eat	zhihou	便去服侍娘吃饭。
	past_time_period	几年 several years	yihou,	章珀尔生长子多佳，次子即弥伴。
	past_time_point	咱这是新鲜鸡头，爱买不买，要减价儿也得 6 点 6:00 pm	yihou,	愿等你就等。
	past_pronoun	这 this	zhihou,	陈赓被调到蒋介石身边做侍从参谋。
	past_zero	香港光复时，人口才 50 万， ϕ	yihou	陆续增加到 100 万。
	future_event	我说：“我已经想好了，我妻子来了 my wife comes	zhihou	我们一起照顾红老师。……”
	future_time_period	据我们的科学家预测，三个小时 three hours	zhihou,	地球就要爆炸，我得赶快离开这里。
	future_time_point	至少在 2021 年 the year 2021	yihou	太阳系公司法对股票持有人的权益是这样规定的。
	future_zero	马冰对记者说： ϕ	“zhihou	我们会尽快通知邵佳一提前回国办理相关的出国手续。”
	general	但如果在吃甜食 eat sweets	zhihou,	再吃些含有维生素 B1 较多的食品，便不会感到倦怠了。

Table 2. (continued)

Coding	Features	N- sequence	N	N+ sequence
Relation between N- sequence and N+ sequence	temporal _ general	切尼是在访问了南朝鲜、菲律宾 visit South Korean, Philippine	yihou	于 20 日抵达美军在日本的重要基地冲绳岛的。
	temporal _ causative	水石膏加热到 130-200°C being heated to 130-200 degrees	zhihou	便成为半水石膏。
	spatial	编辑们也可以在第一段 the first paragraph	zhihou	的任何一个部分结束文章而不至于担心漏掉什么精华。
	ranking	鱼在世界上总数仅排在细菌和昆虫 bacteria and insects	zhihou,	位列第三。
N+ sequence	immediate high transitivity	一阵暴打	zhihou,	菁菁的手掌被打肿了。Jingjing's palm was swollen (after a severe beating).
	low transitivity	肝炎病好了	yihou,	能结婚的。can get married (legally)
	modificatory		yihou	的日子, 还是常常我和外婆两人过生活。 days
	zero	初步有此报道是在当地时间 21 日早上 10 点	zhihou,	φ
Others (no ground for judging)				

3. Whether the sequence denotes a certain point in time (i.e., coded as “Point”; e.g., 6 点以后 6 *dian yihou* ‘after 6:00 pm’), or a time period (i.e., coded as “Period”; e.g., 几年以后 *ji nian yihou* ‘several years later’).
4. Whether the sequence only contains the target word *zhihou* / *yihou* itself and not any other words (i.e. coded as “Zero”; e.g., 以后 *yihou* ‘later’).

The second coding dimension concerns the relation between the N- sequence and the N+ sequence. Four types of relations were identified and coded as:

1. “Temporal_General,” in which the event denoted in the N- sequence happens prior to the event denoted in the N+ sequence, e.g., 在访问了南朝鲜、菲律宾以后 *zai fangwen le Nan Chaoxian, Feilubin yihou* ‘after visiting South Korean and Philippine.’
2. “Temporal_Causative,” in which the N- sequence denotes cause and the N+ sequence denotes effect, e.g., 水石膏加热到130–200°C之后 *shuishigao jiare dao 130–200°C zhihou* ‘upon being heated to 130–200 degrees.’
3. “Spatial,” in which spatial relations is denoted, e.g., 在一段之后 *zai di yi duan zhihou* ‘after the first paragraph.’
4. “Ranking,” in which the ranking information is denoted, e.g., 鱼在世界上总数仅排在细菌和昆虫之后 *yu zai shijie shang zongshu jin pai zai xijun he kunchong zhihou* ‘the total number of fish in the world only ranks after bacteria and insects.’

The third coding dimension concerns the semantic features of the N+ sequence. Distinctions were coded into the following four categories:

1. “Immediate high transitivity,” e.g., 一阵暴打之后，菁菁的手掌被打肿了 *yi zhen baoda zhihou, Jingjing de shouzhang bei da zhong le*. ‘After a severe beating, Jingjing’s palm was swollen.’
2. “Low transitivity,” e.g., 肝炎病好了以后能结婚的 *Ganyan bing hao le yihou neng jiehun de*. ‘After one’s hepatitis is cured, one can get married (legally).’
3. “Modificatory⁸,” e.g., 以后的日子 *yihou de rizi* ‘the days afterwards.’
4. “Zero,” e.g., 初步有此报道是在当地时间 21 日早上 10 点之后 + \varnothing *chubu you ci baodao shi zai dangdi shijian 21 ri zaoshang 10 dian zhihou + \varnothing* ‘the very first report of this (incident) was after 10 o’clock in the morning on date 21st of the local time.’

Since the coding of immediacy and transitivity regarding the predicates in the N+ sequence is not as straightforward and has turned out to be important in revealing

8. Although “Modificatory” is a structural feature, and not a semantic feature, coding of this feature is necessary, for the reason that it marks the cases where *zhihou* and *yihou* cannot be coded on transitivity.

the distinctions of *zhihou* and *yihou* in written discourse, I will explain the related coding criteria in greater detail.

Immediacy is a feature regarding how immediately a second action occurs after the completion of the first action. This is a factor identified in the current study.

Using Hopper and Thompson's (1980) framework of transitivity, a distinction can be made between high transitivity and low transitivity with regard to the predicate of the N+ sequence. Note that the notion of transitivity in this widely adopted framework is a broad one that has various sub-categories such as "Kinesis," "Aspect," etc. Four categories (Table 3) of transitivity in Hopper and Thompson (1980) are of particular relevance to the current study.

Table 3. Illustration of four categories of transitivity in Hopper and Thompson (1980)

Categories	High in transitivity	Low in transitivity
Kinesis	action (<i>I hugged Sally</i>)	non-action (<i>I like Sally</i>)
Aspect	telic (<i>I ate it up</i>) (completed & bounded)	atelic (<i>I am eating it</i>)
Punctuality	punctual (<i>kick</i>)	non-punctual (<i>carry</i>)
Affirmation	affirmative (<i>I kicked him</i>)	negative (<i>I didn't kick him</i>)

According to Hopper and Thompson (1980), the distinction made in the category "Kinesis" is action versus non-action, namely, "actions can be transferred from one participant to another; states cannot. Thus something happens to Sally in *I hugged Sally*, but not in *I like Sally*." The distinction made in the category "Aspect" is telic versus atelic, namely, "an action viewed from its endpoint, i.e. a telic action, is more effectively transferred to a patient than one not provided with such an endpoint. In the telic sentence *I ate it up*, the activity is viewed as completed, and the transferral is carried out in its entirety; but in the atelic *I am eating it*, the transferral is only partially carried out." The distinction made in the category "Punctuality" is punctual versus non-punctual, namely, "actions carried out with no obvious transitional phase between inception and completion have a more marked effect on their patients than actions which are inherently on-going; contrast *kick* (punctual) with *carry* (non-punctual)." The distinction made in the category "Affirmation" is affirmative versus negative.

Using the notion of immediacy and Hopper and Thompson's (1980) categorization of transitivity, predicates of the N+ sequence were coded as "immediate high transitivity" or "low transitivity."

Predicates of the N+ sequence were coded as "immediate high transitivity" if at least one of the following criteria was met:

1. N+, which is transitive, immediately follows N-. A typical case is to co-occur with adverbs that denote immediate actions, i.e., a second transitive action

that occurs immediately upon the completion of the first action. Such adverbs include 便 *bian* ‘(immediately) then,’ 就 *jiu* ‘(immediately) then,’ 立刻 *like* ‘immediately,’ 立即 *liji* ‘immediately,’ 即 *ji* ‘immediately,’ etc. For example,

- (6) 他 决定 孩子 出世 之后 就 跟 李小兰 离婚。
 ta jue ding haizi chushi **zhihou** jiu gen Li Xiaolan lihun
 3SG decide child born after then with Xiaolan.Li divorce
 ‘He decided to divorce Xiaolan Li immediately upon the birth of the child.’

2. Punctual verbs and verb phrases: actions carried out with no obvious transitional phase between inception and completion (Hopper & Thompson 1980). E.g. 问 *wen* ‘to ask’ (see Example 3).

- (3) 早餐 之后, 白素 便 问 秋林, 谁 叫 昆华。
 zaocan **zhihou** Baisu bian wen Qiulin, shui jiao Kunhua
 breakfast after Baisu then ask Qiulin who is.called Kunhua
 ‘Immediately after breakfast, Baisu asked Qiulin who Kunhua was.’

The coding of this type also includes resultative verb complements and perfective aspectual marker *le* in the predicate. An example of a resultative verb complement is provided in (7):

- (7) 一 阵 暴打 之后, 菁菁 的 手掌 被 打 肿
 yi zhen baoda **zhihou** Jingjing de shouzhang bei da zhong
 one CLF severe.beating after Jingjing ASSO palm PASS hit swollen
 了。
 le
 PFV
 After a severe beating, Jingjing’s palm was swollen.

In Example (7), the verb phrase *da zhong le* is a resultative complement that indicates the result of the action: swollen. The resultative complement marks the completion of an action and therefore is coded as high in transitivity.

An example of the perfective aspectual marker 了 *le* is provided in (8):

- (8) 岑不凡 入座 之后, 重重 拍 了 一下 桌子。
 Cen Bufan ruzuo **zhihou** zhongzhong pai le yixia zhuozi
 Bufan.Cen take.the.seat after heavily bang PFV once table
 After taking the seat, Bufan Cen banged the table heavily.

In Example (8), the perfective aspectual markers *le* indicates the completion of the action – banging the table.

3. Telic verbs: a completed action or activity that is bounded and has an ending point. For example,

- (9) 水石膏 加热 到 130–200°C 之后 便 成为
 shuishigao jiare dao zhihou bian chengwei
 hydrated.gypsum be.heated to after then become
 半水石膏。
 banshuishigao
 semi-hydrated.gypsum
 ‘Upon being heated to 130–200°C, hydrated gypsum turns into semi-hydrated gypsum.’

In Example (9), *chengwei* ‘to become’ indicates a complete and bounded activity that has an ending point – turning into another substance.

4. Causative predicates with a verb as the causative manner: Predicates that denote a direct impact on the patient caused by the agent. For example,

- (10) 看 完 之后, 淑姬 把 画面 定格 在 惠婷、邦智 身
 kan wan zhihou Shuji ba huamian dingge zai Huiting Bangzhi shen
 watch finish after Shuji BA scene stop at Huiting Bangzhi body
 上
 shang
 on
 ‘After watching (the disk), Shuji stopped (the film) at a scene with Huiting and Bangzhi.’

In Example (10), the causative marker 把 *ba* ‘to cause’ denotes a direct and significant (Su, forthcoming) impact on the patient (i.e. the film or the disk) caused by the agent Shuji, namely, the film was stopped at a particular scene. This causative sentence contains a verb 定格 *dingge* ‘to stop’ as the causative manner.

On the other hand, predicates of the N+ sequence were coded as “low transitivity” if at least one of the following criteria was met.

1. Non-action: “Actions can be transferred from one participant to another; states cannot” (Hopper & Thompson 1980). These kinds of predicates can have at least one of the following features.

- i. Denotes a state, rather than an action. For example,

- (11) 自此以后, 我就更加心灰意冷了。
 ‘Ever since then, I became even more dejected.’

- ii. Non-action verbs such as 知道 *zhidao* ‘to know,’ 明白 *mingbai* ‘understand.’ The reason why non-action verbs are categorized as “low transitivity” is because non-action verbs are not bounded events and the status they denote often lasts for a relatively long period of time. For example,

(12)多年以后，我也终于明白了自己的脑海里为什么总保存着那些虚假的爱情故事。

‘Many years later, I finally **understood** why I always had those unrealistic love stories in my mind.’

iii. Indicating capability; concurring with potential adverbs 能 *neng* ‘can,’ 可以 *keyi* ‘can,’ 会 *hui* ‘would.’ For example,

(13)肝炎病好了以后**能结婚的**。

‘After one’s hepatitis is cured, one **can get married (legally).**’

2. Non-punctual actions, which are inherently “on-going” (Hopper & Thompson 1980). These kinds of predicates can have at least one of the following features.

i. The action lasts for a long period of time. For example,

(14)三中全会以后，我们用了十一年的时间，才使粮食从六千亿斤增长并稳定在八千亿斤的水平上。

‘Since the 3rd Plenary Session, it has taken us eleven years to grow the grain production from three hundred billion to four hundred billion.’

(15)回到日本以后，就**从事进口欧洲影片的工作**。

‘After returning to Japan, (he) **pursued a job importing movies from Europe.**’

ii. Concurring with adverbs that denote gradual tendency, e.g., 逐渐 *zhujian* ‘gradually,’ 渐渐 *jianjian* ‘gradually,’ 越来越 *yuelaiyue* ‘(getting) more and more.’ For example,

(16)从此以后，母熊**逐渐放松了对猎人的警惕**。

‘Ever since then, the female bear **gradually became less wary of the hunter.**’

iii. Exhibiting stability and regularity instead of having a single occurrence; concurring with frequency adverbs such as 通常 *tongchang* ‘usually,’ 每 *mei* ‘every time; whenever,’ 频繁 *pingfan* ‘frequently,’ 常常 *changchang* ‘usually,’ 频频 *pingping* ‘frequently,’ 总 *zong* ‘always,’ etc. For example,

(17)自那以后，**每到正月十五，家家户户都挂起了红灯**。

‘Ever since then, **whenever the Lantern Festival (i.e., 15th of the first lunar month) came, every house hung red lanterns.**’

iv. Continuing actions; concurring with adverbs that denote continuity, such as 依然 *yiran* ‘still.’ For example,

(18)这以后，党**依然让他执掌兵权**。

‘After this, the Party still allowed him to **maintain military power**.’

- v. Progressing activities; concurring with the progressive marker 着 *zhe*. For example,

(19) 从那时以后，苦根天天盼着买牛这天的来到。

‘Since that time, each day, Kugen **was looking forward to** the day when he could buy a cow.’

3. Negative predicates (Hopper & Thompson 1980): The reason negative predicates are categorized as low transitivity is because negative events are non-bounded. For example,

(20) 我知道，我以后都不会再见他了，永远不会了。

‘I know that from now on, I **will never** see him again.’

4. Interrogative predicates: The reason interrogative predicates are categorized as low transitivity is because they are non-bounded. For example,

(21) 现在还能维持，以后怎么样不知道。

‘Now (we) can still sustain. (We) don’t know **how (it would be)** in the future.’

4. Findings

4.1 Findings regarding the two words

The overall findings concerning the distinctions between *zhihou* and *yihou* in the written and conversational corpora are: (1) In both writing and conversation *zhihou* favors past and *yihou* favors future; in writing but not in conversation *zhihou* is more often used with immediate high transitivity actions and causal relations, whereas *yihou* is more often used with low transitivity states. (2) The two words have different formulaic chunks and chunking patterns in written and conversational discourses. Details of these findings will be discussed in the following sections.

4.2 Previous claims about the two words revisited

With regard to the dispute on whether *yihou* can only describe temporal relations (Li 2004) or can also describe spatial relations (Yi 1997), my findings support the claim that *yihou* is a temporal expression (Li 2004). All occurrences of *yihou* in written and conversational corpora are time expressions.

Table 4. Free-standing *zhihou* in written and conversational corpora

N- sequence	Written <i>zhihou</i> (n = 450)		Conversation <i>zhihou</i> (n = 44)	
past _ zero	4.0%	18	0.0%	0
future _ zero	1.6%	7	2.3%	1
total	5.6%	25	2.3%	1

The previous claim that *zhihou* cannot be used in a free-standing way (HSK Guideline, 2000) was proven invalid. 5.6% of *zhihou* in the written corpora and 2.3% in the conversational corpus are free-standing (Table 4).

The previous claim that *zhihou* cannot be used for future time was also proven invalid. In the written data, although the majority of *zhihou* tokens are associated with past events or time, 10% of *zhihou* are used with future events or time. That number rises to 29.5% for the conversational data (Table 5).

Table 5. *Zhihou* indicating future in written and conversational corpora

N- sequence	Written <i>zhihou</i> (n = 450)		Conversation <i>zhihou</i> (n = 44)	
future _ event	5.1%	23	18.2%	8
future _ time _ period	2.4%	11	4.5%	2
future _ time _ point	0.9%	4	4.5%	2
future _ zero	1.6%	7	2.3%	1
total	10%	45	29.5%	13

The previous claim that *zhihou* is a time-spatial expression (Li 2004) is supported by this study. However, among all the occurrences of *zhihou* in the three corpora, only four (0.28%) are used to express spatial relations. None of them refers to a concrete spatial location. The four cases are:

- (22) 在第一段 之后

(LCMC)

zai diyiduan *zhihou*

‘After the first paragraph’
- (23) 黑暗 之后 的东西

(UCLA)

heian *zhihou* de dongxi

‘Things behind the dark’
- (24) 在这段 文字 之后

(LCMC)

zai zhe duan wenzi *zhihou*

‘After this paragraph’

- (25) 紧跟 潮流 之后 (UCLA)
 jingen chaoliu *zhihou*
 'Keep up with the latest trends'

The conclusion is that *zhihou* is predominately used as a time expression; only in very few occasions is it used to describe space. In such cases, instead of describing a concrete location, *zhihou* is usually used metaphorically.

The previous claim that *zhihou* is preferred in written language (Zhou et al. 2004) is not supported in the current study. My data suggests that modern Chinese written discourse does not favor *zhihou* or *yihou*. However, the two words have some other differences: *Zhihou* has more features of classical Chinese, and *yihou* is predominately preferred in conversational discourse.

4.3 Overall corpus distribution

An examination of some major Chinese corpora demonstrates that *yihou* occurs more often in conversational discourse (Table 6). This does not mean that *zhihou* occurs more frequently than *yihou* in written texts. In fact, each word appears as frequently as the other in Modern Chinese texts.

Table 6. Occurrences of *zhihou* and *yihou* in various written and spoken corpora

Discourse Type	Corpora	<i>zhihou</i>	<i>yihou</i>
Classical written	National (classical) ^a	73%	27%
	CCL by PKU (classical) ^b	66%	34%
Modern Written	Academia Sinica (written) ^c	61%	39%
	UCLA	50%	
	LCMC	44%	56%
Modern Written for spoken	CUC (broadcast, film & TV) ^d	55%	45%
	Academia Sinica (film & TV)	51%	49%
Modern Spoken	Taiwan NCCU ^e	58%	42%
	Academia Sinica (spoken)	24%	76%
	CallFriend	10%	90%
	BJKY (Beijing spoken) ^f	9%	91%

^a <http://www.cncorpus.org/ACindex.aspx>

^b http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp

^c <http://app.sinica.edu.tw/kiwi/mkiwi/>

^d <http://ling.cuc.edu.cn/RawPub/>

^e <http://140.119.172.200/chinese/cmain.php>

^f http://yykxy.bcu.edu.cn/art/2014/5/12/art_6233_1072951.html

Turning now to the written corpora, a statistical analysis reveals that *zhihou* and *yihou* are preferred in different written discourse types (Table 7).

Table 7. Frequencies across discourse types in written corpora LCMC and UCLA

Discourse Types	<i>zhihou</i> (n = 450)		<i>yihou</i> (n = 521)		ChiSqu	Signif.
(A) News Reportage	13.3%	60	3.6%	19	30.21	(p < 0.02)
(B) News Editorials	2.9%	13	2.7%	14	0.03	
(C) News Reviews	3.6%	16	3.6%	19	0.01	
(D) Religion	3.6%	16	5.8%	30	2.62	
(E) Skills, Trades and Hobbies	7.1%	32	6.1%	32	0.36	
(F) Popular Lore	8.2%	37	10.2%	53	0.91	
(G) Essays and Biographies	16.7%	75	22.5%	117	5.17	
(H) Reports and Official Documents	0.9%	4	1.3%	7	0.45	
(J) Science (academic prose)	5.3%	24	10.7%	56	1.49	
(K) General Fiction	6.4%	29	4.6%	24	1.03	
(L) Mystery and Detective Fiction	9.1%	41	9.4%	49	0.03	
(M) Science Fiction	4.7%	21	2.7%	14	2.70	
(N) Adventure Stories	9.1%	41	2.7%	14	18.58	(p < 0.02)
(P) Romantic Fiction	7.6%	34	10.7%	56	2.96	
(R) Humour	1.6%	7	3.3%	17	2.94	

A Chi-Square test shows that *zhihou* is significantly preferred in two discourse types: (A) News Reportage and (N) Adventure Stories (martial arts fictions). This kind of skewed distribution is related to the semantic distinctions between the two words. The reason why *zhihou* occurs much more often than *yihou* in (A) News Reportage is relatively transparent. News Reportage reports latest events. Therefore, current and recent activities are the major topics of this discourse type. The function of reporting immediate sequential events makes *zhihou* more suitable. There are two reasons why *zhihou* occurs much more frequently in (N) Adventure Stories: (1) Physical activity: Adventure and martial arts fiction are featured with a large number of physical actions. The tendency for *zhihou* to trigger an immediate high transitivity action (which will be discussed in detail later in this paper) fits this type of text. (2) Verbal antiquity: Adventure and martial arts fiction reflect the idiosyncratic literary tradition of heroic legends that can be traced back to the early Tang dynasty in the 7th century, a tradition that is linguistically marked with verbal antiquity. Words that are characterized as antiquated are associated with classic writings and serve the style of martial arts fiction better (Jing-Schmidt & Tao 2009).

5. Semantic distinctions in writing and conversation

5.1 *Zhihou* favors past and *yihou* favors future in writing and conversation

In both written and conversational discourse, there is a semantic distinction between the two words – *Zhihou* is more likely to occur with past events, whereas *yihou* is more likely to occur with future events, and in such use *yihou* is usually free-standing (Table 8 and 9). This difference is statistically significant ($p < 0.02$).

Table 8. Features in the N– sequence in written corpora LCMC and UCLA

N– sequence	<i>zhihou</i> (n = 450)		<i>yihou</i> (n = 521)		ChiSqu	Signif.
past_event	62.7%	282	28.2%	147	116.21	($p < 0.02$)
past_time_period	7.3%	33	6.3%	33	0.38	
past_time_point	1.1%	5	6.9%	36	20.08	
past_pronoun	1.6%	7	6.9%	36	16.35	
past_zero	4.0%	18	11.3%	59	17.74	
future_event	5.1%	23	3.3%	17	2.09	
future_time_period	2.4%	11	1.2%	6	2.35	
future_time_point	0.9%	4	1.5%	8	0.83	
future_zero	1.6%	7	24.8%	129	107.94	($p < 0.02$)
general	9.1%	41	6.7%	35	1.92	
others (no ground for judging)	4.2%	19	2.9%	15	1.29	

Table 9. Features in the N– sequence of in conversational corpus CallF

N– sequence	<i>zhihou</i> (n = 44)		<i>yihou</i> (n = 407)		ChiSqu	Signif.
past_event	50.0%	22	24.1%	98	13.66	($p < 0.02$)
past_time_period	0.0%	0	2.7%	11	1.22	
past_time_point	0.0%	0	0.0%	0	0.00	
past_pronoun	0.0%	0	0.7%	3	0.33	
past_zero	0.0%	0	1.7%	7	0.77	
future_event	18.2%	8	26.3%	107	1.37	
future_time_period	4.5%	2	1.7%	7	1.62	
future_time_point	4.5%	2	2.0%	8	1.22	
future_zero	2.3%	1	32.2%	131	17.16	($p < 0.02$)
general	15.9%	7	5.4%	22	7.28	
truncated	2.3%	1	0.2%	1	3.70	
r(first part of repair/repetition)	2.3%	1	2.7%	11	0.03	
others (no ground for judging)	0.0%	0	0.2%	1	0.11	

Combining the categories “past_zero” and “future_zero,” over one third (36.1%) of *yihou* tokens in the written data are free-standing, whereas only 5.6% of *zhihou* in the written data are free-standing. In the conversational data, over one third (33.9%) of *yihou* tokens are free-standing, whereas only 2.3% of *zhihou* are free-standing. Below is an example in which *yihou* stands by itself to refer to a future situation.

- (26) 专家 说,
zhuanjia shuo
expert said
'Expert(s) said,'
我国 现在 对 虚拟人 的 研究 还 刚刚 开始,
woguo xianzai dui xu'niren de yanjiu hai ganggang kaishi
our.country now on robots ASSO research still just began
'The research on robots in our country has just begun.'
以后 研究 逐步 深入,
yihou yanjiu zhubu shenru
in.the.future research gradually deepen
'In the future, with the research deepened,'
许多 现在 的 幻想 都会 一一 实现 的。
xuduo xianzai de huanxiang dou hui yiyi shixian de
many now ASSO fantasy all will one.after.another fulfill EM
'Many dreams (i.e., now-seem-to-be fantasy) will be fulfilled one after another.'

In Example (26), *yihou*, which means ‘in the future,’ stands by itself with no preceding N– sequence.

Combing all the categories related to past events and all those related to future events, the data indicate that the written corpora LCMC and UCLA tend to report more past events (Table 10), whereas the conversational corpus CallFriend tends to report more future events (Table 11).

Table 10. Features in the N– sequence in written corpora LCMC and UCLA

N– sequence	<i>zhihou</i> (n = 450)		<i>yihou</i> (n = 521)	
All past	76.7%	345	59.6%	311
All future	10.0%	45	30.8%	160

Table 11. Features in the N– sequence in conversational corpus CallF

N– sequence	<i>zhihou</i> (n = 44)		<i>yihou</i> (n = 407)	
All past	50.0%	22	29.2%	113
All future	29.5%	11	62.2%	247

This indicates that these two kinds of corpora may not be comparable. If one conducts a comparison of a single word or grammatical construction in these two kinds of corpora, one cannot tell whether the difference is related only to that single word or grammatical construction in question or is a reflection of the difference between the two corpora. This supports the argument of this study that instead of using a single word or grammatical construction to compare two discourse types, using a pair of near-synonymous words or grammatical constructions is more reliable.

5.2 *Zhihou* indicates causal relations in writing but not in conversation

In the written corpora, *zhihou* is more likely to describe causal relations than *yihou* (Table 12). This semantic distinction is not found in the conversational data.

Table 12. Relation between N– sequence and N+ sequence in written corpora

Relation between N– and N+	<i>zhihou</i> (<i>n</i> = 450)		<i>yihou</i> (<i>n</i> = 521)		ChiSqu	Signif.
temporal_general	81.3%	366	95.4%	497	48.29	(<i>p</i> < 0.02)
temporal_causative	16.0%	72	4.6%	24	35.18	(<i>p</i> < 0.02)
spatial	0.9%	4	0.0%	0	4.65	
ranking	1.3%	6	0.0%	0	6.99	
Others(no ground for judging)	0.4%	2	0.0%	0	2.32	

As Table 12 shows, 16.0% of the time *zhihou* is used to describe causal relations, in comparison to only 4.6% of *yihou* usages. A Chi-Square test shows that the difference is significant (*p* < 0.02).

In the written data, when used to describe causal relations, *zhihou* mainly takes the form of “A *zhihou* B” in which A indicates the cause and B the effect. For example,

(27)

A (cause)				B (effect)			
水石膏	加热	到	130–200°C	之后	便	成为	半水石膏。
shuishigao	jiare	dao		zhihou	bian	chengwei	banshuishigao
hydrated.	be.heated to			after	then	become	semi-hydrated.gypsum
gypsum							

‘Upon being heated to 130–200°C, hydrated gypsum turns into semi-hydrated gypsum.’

In Example (27), being heated to 130–200°C would cause gypsum to turn into semi-hydrated gypsum. *Zhihou* serves as a temporal link of the cause and effect.

(28)

A (cause)					
当	劳动生产力	向前	发展,	认识能力	得到 提高 之后,
dang	laodong shengchanli	xiang qian	fazhan	renshinengli	dedao tigao zhihou
when	labor.productivity	forward	develop	cognitive.ability	get advance after

B (effect)					
必然	产生	两	方面	的	变化。
biran	chansheng	liang	fangmian	de	bianhua
certainly	produce	two	aspect	ASSO	changes

‘With the development of labor productivity and the advancement of cognitive ability, two changes are inevitable.’

In Example (28), development of labor productivity and advancement of cognitive ability are the causes that lead to the changes. *Zhihou* not only serves as a temporal device to connect two successive events, but also a logical linking of cause and effect.

The link between temporal relations and causal relations can also be found in other languages such as English (e.g., *since* can denote both temporal relations and causal relations).

The semantic distinction where *zhihou* is more associated with causal relations than *yihou* is not found in the conversational data (Table 13). An Exact Pearson Chi-Square test shows that the differences between the two words regarding the relation between the N- sequence and the N+ sequence in the conversational corpus are not significant (value=.243, df = 1, *p* = .785).

Table 13. Relation between N- sequence and N+ sequence in conversational corpus CallF

Relation between N- and N+	<i>zhihou</i> (n = 44)		<i>yihou</i> (n = 407)	
temporal_general	88.6	38	90.7	369
temporal_causative	0.0	0	0.0	0
Spatial	0.0	0	0.0	0
Ranking	0.0	0	0.0	0
Others(truncated, no ground for judging)	11.4	5	9.3	38

5.3 *Zhihou* favors immediate high transitivity actions and *yihou* favors low transitivity states in writing but not in conversation

In written discourse, *zhihou* and *yihou* have opposite properties in the N+ sequence (Table 14). Whereas *zhihou* is more often followed by a predicate that denotes an immediate high transitivity action, *yihou* is more often followed by a predicate that denotes a low transitivity action or state. A Chi-Square test shows that this difference is significant ($p < 0.02$).

Table 14. Features in the N+ sequence in written corpora LCMC and UCLA

N+ sequence	<i>zhihou</i> (n = 450)		<i>yihou</i> (n = 521)		ChiSqu	Signif.
immediate high transitivity	50.7%	228	15.9%	83	133.82	($p < 0.02$)
low transitivity	33.8%	152	63.9%	333	87.72	($p < 0.02$)
modificatory	8.9%	40	12.1%	63	2.61	
zero	3.1%	14	2.9%	15	0.04	
Others (no ground for judging)	3.6%	16	5.2%	27	1.51	

In the written corpora, *zhihou* is mostly (50.7%) associated with immediate actions that are high in transitivity. “Immediate” means that the action or state denoted in the N+ sequence takes place immediately upon the completion of whatever described in the N– sequence. For example,

- (29) 他 决定 孩子 出世 之后 就 跟 李小兰 离婚。
 ta jueding haizi chushi **zhihou** jiu gen Li Xiaolan lihun
 3sg decide child born after then with Xiaolan.Li divorce
 ‘He decided to divorce Xiaolan Li immediately upon the birth of the child.’

In Example (29), 就跟李小兰离婚 *jiu gen Li Xiaolan lihun* ‘to immediately divorce Xiaolan Li’ in the N+ sequence is an immediate high transitivity action performed upon the completion of the event described in the N– sequence –孩子出世 *haizi chushi* ‘the child is born.’

In the written corpora, *yihou* is mostly (63.9%) associated with low transitivity states or actions. Readers can refer to Section 3.4 Coding of data for a large number of examples in which *yihou* is used with low transitivity states or actions.

Example (30) shows the difference between *zhihou* and *yihou* with regard to immediacy and transitivity.

- (30) 这样 过 了 几 年 的 洞 后 生 活 之 后,
 zheyang guo le ji nian de dong hou shenghuo zhihou
 this spend PFV several year ASSO cave behind life after
 母熊 生 了 一 个 小 崽, 从 此 以 后, 母 熊
 muxiong sheng le yi ge xiaozai, congciyihou muxiong
 female.bear give.birth.to PFV one CLF baby ever.since.then female.bear
 逐 渐 放 松 了 对 猎 人 的 警 惕。
 zhujian fangsong le dui lieren de jingti.
 gradually let.down PFV towards hunter ASSO guard
 ‘After (i.e., *zhihou*) several years of living in the cave, the female bear gave birth to a baby bear. Ever since then (i.e., *congciyihou*), the female bear gradually became less wary of the hunter.’

In Example (30), *zhihou* is used with the verb 生 *sheng* ‘to give birth to,’ which is high in transitivity, whereas *yihou* is followed by a predicate 逐渐放松 *zhujian fangsong* ‘gradually let her guard down,’ which is low in transitivity and immediacy.

When *zhihou* and *yihou* are both used with a proximate demonstrative word 这 *zhe* ‘this,’ *zhihou* is more likely to be used with verbs of high transitivity, whereas *yihou* is more often followed by a low transitivity predicate. Compare Example (31) with (32):

- (31) 这 之后, 陈赓 被 调 到 蒋介石 身 边
 zhe zhihou Chen Geng bei diao dao Jiang Jieshi shenbian
 this after Geng.Chen PASS transfer to Kai-shek.Chiang at.his.sides
 做 侍从参谋。
 zuo shicongcanmou
 work as aide
 ‘After (i.e., *zhihou*) this, Chen Geng was transferred to work as an aide to Chiang Kai-shek.’
- (32) 这 以后, 党 依然 让 他 执掌 兵 权。
 zhe yihou dang yiran rang ta zhezhang bingquan
 this after Party still let 3SG hold military.power
 ‘After (i.e., *yihou*) this, the Party still allowed him to maintain military power.’

In Example (31), *zhihou* is used with a high transitivity verb 调 *diao* ‘to transfer somebody from a position/place to another’. In Example (32), *yihou* is used with a low transitivity predicate 依然让他执掌兵权 *yiran rang ta zhizhang bingquan* ‘still let him maintain military power.’

The feature of *zhihou* being more associated with causal relations correlates with the feature of this word to introduce an immediate action high in transitivity.

This is because: (1) an action that immediately follows a prior event is more likely to be a result of it; and at the same time, (2) two events that occur with immediate temporal relations are more likely to be perceived as cause and effect.

The distinction regarding immediacy and transitivity between *zhihou* and *yihou* is not found in the conversational data (Table 15). An Exact Pearson Chi-Square test shows that the differences between the two words are not significant (value=.673, df = 3, $p = .957$).

Table 15. Features in the N+ sequence in conversational corpus CallF

CallF	<i>zhihou</i> (n = 44)		<i>yihou</i> (n = 407)	
immediate high transitivity	29.5%	13	29.9%	122
low transitivity	52.3%	23	51.4%	209
truncated	9.1%	4	9.3%	38
modificatory	0%	0	1.5%	6
zero	4.5%	2	3.9%	16
first part of repeat/repair	2.3%	1	2.7%	11
others (no ground for judging)	0%	0	0%	0
n/a (what the speaker means is unintelligible)	2.3%	1	1.2%	5

- (33) B: 冬天 干 了 之后 呢, 它 到 来年 春天
 dongtian gan le **zhihou** ne ta dao lainian chuntian
 winter dry PFV after, PRT 3SG until the.following.year spring
 还是 旱。
haishi han
 remain droughty
 ‘After (i.e., *zhihou*) (it gets) dry in the winter, it **remains droughty** through the next spring.’

Example (33) is a case of *zhihou* used with a low transitivity predicate in conversational discourse. 还是旱 *haishi han* ‘remains droughty’ in the N+ sequence indicates a state and is low in transitivity.

- (34) B: 就 可以到 商店 去买 那个 高级 的 来用,
 jiu keyi dao shangdian qu mai nage gaoji de lai yong,
 then can go store to buy that good-quality NMLZ to use
 用了 以后, 就 退。
 yong le **yihou**, jiu tui.
 use PFV after, then return.
 ‘(You) can go to the store to buy that good-quality one. **Once** (i.e., *yihou*) you have used it, **return it**.’

Example (34) is a case of *yihou* used with immediate high transitivity actions in conversation. 退 *tui* ‘to return (to the store)’ in the N+ sequence is an immediate action that is of high transitivity. The concurrence with 就 *jiu* ‘then; right after’ reinforces the sense of immediacy.

6. Chunking in writing and conversation

The findings regarding the chunking patterns of *zhihou* and *yihou* are as follows: (1) The two words have different formulaic chunks in the written and conversational corpora. (2) The formulaic chunks that the two words form in the conversational corpus have different kinds of fixedness and have some fuzzy cases that seem to be bridging between two different chunks. The finding that chunks in conversation have different kinds of fixedness is consistent with some conclusions in previous studies (such as Fox and Thompson 2007). In contrast, the formulaic chunks the two words form in the written corpora are more fixed and clear-cut.

These findings suggest that in conversational discourse, chunking may exhibit different stages of fixedness, whereas in written discourse it may manifest itself at the final stage, namely, more or less fully fixed in form.

6.1 Different chunks of the same word in writing and conversation

Although there is no fixed way to determine what can be counted as formulaic chunks, this study basically follows Bybee’s (2010) views on chunking. Bybee (2010: 56) observes that chunking leads to many characteristics of linguistic structure, including the “grouping of meaning with particular morphosyntactic constructions and with context, which at once maintains specific meanings for specific contexts and also allows new meanings to be established through inference from context.”

Applying this criterion, at least one formulaic chunk in written discourse and five formulaic chunks in conversational discourse can be identified.

The multi-word unit 从此以后 *congciyihou* ‘ever since then,’ which appears solely in my written data but not in my conversational data, is a formulaic chunk associated with dramatized context. Despite the very few occurrences (8 out of 521 *yihou* tokens, 1.5%), *congciyihou* is considered a formulaic chunk because of its strong association with exceptional contexts and the fixedness of its morphosyntactic makeup.

- (35) 这样 过 了 几 年 的 洞 后 生 活 之 后,
 zheyang guo le ji nian de dong hou shenghuo zhihou
 this spend PFV several year ASSO cave behind life after
 母熊 生 了 一 个 小 崽, 从 此 以 后, 母 熊
 muxiong sheng le yi ge xiaozai congciyihou muxiong
 female.bear give.birth.to PFV one CLF baby, ever.since.then female.bear
 逐 渐 放 松 了 对 猎 人 的 警 惕。
 zhujian fangsong le dui lieren de jingti.
 gradually let.down PFV towards hunter ASSO guard.
 ‘After (i.e., *zhihou*) several years of living in the cave, the female bear gave birth to a baby bear. Ever since then (i.e., *congciyihou*), the female bear gradually became less wary of the hunter.’

In Example (35), the previous event in the N– sequence is ‘giving birth to a baby bear.’ The subsequent situation following *congciyihou* is an extraordinary circumstance with exceptional connotations – in this case life threatening, namely, the bear letting down her guard against the hunter.

An examination of the texts containing *congciyihou* shows that this string has a strong tendency to occur in imaginary and dramatic narrations. No occurrence is detected in official, scientific, objective, informational descriptions, non-dramatic narrations, or construed conversations in the written corpora.

The function of *yihou* in introducing a low transitivity action or state is in accord with the connotation of *congciyihou* as having a long-term significant consequence. Although in theory there can be another multi-word unit 从此之后 *congciizhihou*, no such case is identified in the three corpora I examined⁹. The reason might be that the meaning ‘ever since then’ is not consistent with the notion of immediacy *zhihou* indicates.

In the written discourse, there are also some other formulaic chunks regarding *yihou* that are similar to 从此以后 *congciyihou*, such as 从那以后 *congnyayihou* 从今以后 *congjinyihou*. As for *zhihou*, only one instance of 从那之后 *congnaizhihou* was identified, and none of the other chunks was found.

Turning to the chunks *zhihou* and *yihou* form in conversational discourse, the chunk identified in written discourse (i.e., *congciyihou*) is not found in the conversational data. Instead, conversational discourse has its unique formulaic chunks (Table 16).

9. Instances of 从此之后 *congciizhihou* ‘ever since then’ were identified in hyper large databases, such as the online Chinese search engine baidu.com. But *congciyihou* far outnumbers *congci-zhihou* in these databases.

Table 16. Frequencies of chunks regarding *zhihou* and *yihou* in conversational corpus

<i>zhihou</i> (n = 44)	Percentage	Count	<i>yihou</i> (n = 407)	Percentage	Count
完了之后 <i>wanlezhihou</i>	11.4%	5	完了以后 <i>wanleyihou</i>	3.1%	13
之后就 <i>zhihoujiu</i>	9%	4	以后就 <i>yihoujiu</i>	7.2%	30
			以后再 <i>yihouzai</i>	8.6%	36

As Table 16 shows, at least five formulaic chunks 完了之后 *wanlezhihou* ‘afterwards,’ 完了以后 *wanleyihou* ‘afterwards,’ 之后就 *zhihoujiu* ‘right after,’ 以后就 *yihoujiu* ‘right after,’ and 以后再 *yihouzai* ‘later’ are identified in the conversational data.

In the interest of space, here I will only briefly illustrate the use of *wanlezhihou* ‘afterwards’ and *wanleyihou* ‘afterwards.’ There is no detectable semantic distinction between these two words. These two words are often used as a discourse parenthetical (Heine 2010), marking a transition to the next move in storytelling. For example,

- (36)

1.

B:

你妈提出来让她,

ni ma ti chulai rang ta

2SG mom offer RUC let 3SG

‘Your mom offered to let her...’
2.

她说在这过夜吧,

ta shuo zai zhe guoye ba

3SG say at here stay.overnight PRT

‘She (your mom) said please stay here overnight.’
3. →

完了以后她说好啊,

wanleyihou ta shuo hao a

afterwards 3SG say good PRT

‘And she said okay.’
4.

A:

哦

o

PRT

‘Oh’
5. →

B:

完了以后呢,她给她的

wanleyihou ne ta gei ta de

afterwards PRT 3SG give 3SG ASSO

‘Afterwards, she gave her’

一个同伴打了个

yi ge tongban da le ge

one CLF friend call PFV CLF

‘a friend (a phone call).’

In Example (36), there are two instances of 完了以后 *wanleyihou* (in lines 3 and 5), both of which mark a transition to the next move in storytelling.

In conversational discourse, the speaker may abandon the use of a single word (i.e., *zhihou* or *yihou*) in favor of the use of a formulaic chunk (i.e., 完了之后 *wanlezhihou* or 完了以后 *wanleyihou*). Example (37) is a self-repair sequence of this kind.

- (37) 1. B: 我 跟 他 说说,
 wo gen ta shuoshuo
 1SG with 3SG speak
 ‘(Let) me speak with him a bit.’
 2. 就 是 呃 档案 办,
 jiu shi e dang'an ban
 EM is uh dossier process
 ‘That is, process the dossier,’
 3. → 办 完 以后,
 ban wan yihou
 process finish after
 ‘after (it) has been processed,’
 4. → 完了之后 就 把 这 档案 拿,
 wanlezhihou jiu ba zhe dang'an na
 afterwards EM BA this dossier take
 ‘after that, just take the dossier,’
 5. 如 果 能 拿 回 家,
 ruguo neng na huijia
 if can take home
 ‘if (it) can be taken home,’
 6. 拿 回 家.
 na huijia
 take home
 ‘take (it) home.’

In this instance of self-repair, the speaker first uses the single word *yihou* ‘after,’ following a resultative complement 办完 *ban wan* ‘finish processing’ (line 3). The speaker then immediately abandons this phrase, and adopts the use of a formulaic chunk 完了之后 *wanlezhihou* ‘afterwards’ (line 4). It is observed that in this case 完了之后 *wanlezhihou* marks a transition to the next move, namely, taking the dossier home. 完了之后 *wanlezhihou* serves the function of transitioning between different steps, and therefore, it is preferred in this context where the speaker is describing the different steps in an instruction.

6.2 Conversation has different stages of chunking, whereas written discourse mainly has the final clear-cut stage

Based on findings on the formulaic chunks the two words form in written and conversational data, this study argues that conversation may have different stages of chunking, whereas written discourse may mainly have the final clear-cut stage.

Two kinds of chunks can be distinguished: Partially pre-stored and entirely pre-stored. Partially pre-stored chunks have at least one open slot for the speaker to fill in information that fits the syntactic and semantic categories the open slot specifies. For example, the lexical string “-¹⁰ 以后, 再看 -” “- *yihou*(,) *zai kan* -” ‘after -, then see -’ has multiple open slots. Entirely pre-stored chunks have no open slots and function independently as a fixed unit. For example, 以后再说 *yihouzaishuo* ‘put the matter aside for later’ has no open slots within the lexical string.

This study argues that chunking in conversational discourse has different stages. This conclusion is based on the finding that the same string of lexical items can be observed to form both a partially pre-stored chunk and an entirely pre-stored chunk, or to form several different partially pre-stored chunks. For example, the same string of lexical items 以后 *yihou* ‘after’ and 吧 *ba* ‘(an utterance final particle)’ is found to form two kinds of partially pre-stored chunks in the conversational data: (1) “- *yihou ba*” ‘after -,’ and (2) “*yihou ba*, -” ‘In the future, -.’ It can also form an entirely pre-stored chunk, which is *yihouba* ‘later.’ The examples below show that chunking in conversational discourse has different stages.

(1) Partially pre-stored

- 以后, 再 - 看 -
- *yihou*, *zai* - *kan* -
- ‘after -, then - see -’

- (38) B:我打算到了那个太原 以后 ,再在那看时间合适时间.再在太原买回来的票
 ‘B: I plan to arrive in Taiyuan,’ ‘after’ ‘then see in Taiyuan when a good time (to come back) would be and then buy the return ticket.’

(2) Partially pre-stored

- 以后 (,) 再 看
- *yihou* (,) *zai kan*
- ‘after -(,) then see (what to do).’

10. The dash - indicates an open slot.

- (45) A:什么时候有时间过这边儿来,我 以后再说,
 们要是办那个办好的话.B:啊, 'B: let's talk about it
 'A: Come here when you are free. If later.'
 we can finish doing that thing,
 B: Ah,' 不早的了,等你爸爸退休
 了,再说吧,你们先好好的,
 'B: ((here is an ambitious
 sequence)). Wait until your
 father retires, and then we
 can talk about this. You
 guys take good care of your-
 selves first.
- (46) B1:当然如果看看有没有合适的人 以后再说,
 带.A:哦 'A: let's talk about it
 'B1: Of course see if there is some- later.'
 one suitable to bring...
 A: Oh, 她妈妈愿意干了,到时候呵
 到时候再说.
 'A: (Wait until) her mom
 is willing to do it. Then we
 will talk about it'

(5) Partially pre-stored

以后再 看 / 说 + 吧 / 了

yihouzai kan / shuo + ba / le

'(Let's) put the matter aside for later.'

- (47) A:那以 以后再看吧?
 'A: In that case, then' 'see it in the future'
- (48) B1:{sigh}算嘞, 以后再说吧. A:啊?B1:不要紧诶,以
 'B1:{sigh} Never 'Let's talk about it 后再呃,
 mind.' later.' 'A: Ah?
 B1: Don't worry.
 Later.'
- (49) A:啊,那算了,算了, 以后再说吧. A:好不好?B:嗯,嗯.A:
 'A: Ah, never mind, 'Let's talk about it 诶,你还有什么事啊
 never mind.' later.' 'A: Okay?
 B: Okay.
 A: What else do you
 want?'
- (50) A:哦,那么有一个什- 以后再说吧. B:还有什么?
 吧,这个.B:嗯.嗯.A: 'let's talk about it 'B: What else?'
 哦,这事 later.'
 'A: Oh, then there is a
 what, this.
 B: Yeah, yeah.
 A: Oh, this matter'

- (51) A:就这样,反 以后再说了, B: [[distortion]] 嗯,
 'A: That's it.' 'Let's talk about it 嗯, 嗯, 那
 later.' 'B: [[distortion]]
 Okay, okay, okay,
 then'
- (6) Partially pre-stored
 – 以后 吧
 – *yihou ba*
 'after –'
- (52) 那样你你赶紧弄完了 以后吧, B: 嗯, 嗯
 'Then (wait until) you quickly finish it' 'after' 'B: Okay, okay'
- (53) B: 嗯, 妈妈他们是从, 寒假 以后吧, 从((二月)).
 'B: Yeah. Mom and they are (after) 'after' 'from ((February))'
 winter vacation
- (54) 她们住到八月十号 以后吧. A1: 那家里有电话吗?
 'They will stay until August 10' 'after' 'A1: In that case, is there a phone at
 home?'
- (55) B1: 这个我看到那个莽弟回来 以后吧. A: 呃.
 'B1: This, I think we should just wait 'after' 'A: Yeah.'
 until Mangdi is back'
- (7) Partially pre-stored
 以后 吧, –
yihou ba, –
 'In the future, –'
- (56) A: 呃, 你要这个事是越越早越好, 要 以后吧, B: 嗯, A: 这个再再再发财也不
 是时时间久而久之, 'in the future' 容易了, 你知道吗?
 'A: The sooner you do this thing, the bet- 'B: Yeah.
 ter. If you postpone it for a long time' A: it would not be as easy (as it
 is now) to make a fortune. You
 know?'
- (8) Entirely pre-stored
 以后吧
yihouba
 'later'

This chunk is often used to avoid direct rejection or making a decision at the moment of speaking. It is an entirely pre-stored chunk for the reason that it has no open slots.

- (57) 家里安个电话.A:管怎么着,((这个))对吧?B:那以后吧, A:反正-
'Let's install a phone at home. 'later' 'A: Anyhow'
A: whatever, this, right?
B: That

Turning now to the chunking patterns of the two words *zhihou* and *yihou* in written discourse, based on the findings, it is argued that written discourse mainly has the final clear-cut stage. The Chinese four character idioms (i.e., 成语 *chengyu*), which are usually used in writing, are a typical example of chunking at the final clear-cut stage. Most chunks at the final clear-cut stage are entirely pre-stored. They have no open slots and are functioning independently on their own. However, chunks at the final clear-cut stage can also be partially pre-stored. For example, “自从-以后” “*zicong – yihou*,” ‘since -’ found in the written data is a partially pre-stored chunk.

One of the differences between chunking in written discourse and conversational discourse is that chunks in written discourse are mainly at the final clear-cut stage, which means that a chunk is either a partially pre-stored chunk or an entirely pre-stored chunk. Whereas a lexical string in conversation can be both a partially pre-stored chunk and an entirely pre-stored chunk, or can be several partially pre-stored chunks, a lexical string in written discourse may only form one chunk, either partially pre-stored or entirely pre-stored. Below I will use the chunk 从此以后 *congciyihou*¹¹ found in the written data as an example to illustrate this point.

Congciyihou is an entirely pre-stored chunk. All concordance lines of *congciyihou* in the two written corpora share some common characteristics (Table 17).

Table 17. All tokens of *congciyihou* in the written corpora LCMC and UCLA

	Left context	<i>congciyihou</i>	Right context
(58)	这样过了几年的洞后生活之后，母熊生了一个小崽， 'After several years of living in the cave, the female bear gave birth to a baby bear.'	从此以后， 'Ever since then,'	母熊逐渐放松了对猎人的警惕。 the female bear gradually became less wary of the hunter.'
(59)	船行海上，人不能足踏实地，在浮浮沉沉中总感到恐惧惊慌。而华山的峻峭险峰悬空鸟道，甚至山风游云也令她留恋不已。而对茫茫大海，只怕 (Summary: The girl is travelling with her newly wedded husband. She is leaving this place forever.)	从此以后， 'Ever since then,'	那莲花峰落雁峰公主峰，连同义父师兄师姐们的情谊，都将成为她记忆中的一尘一埃了啊。 (Summary: for the rest of her life, the girl can no longer come back to reunite with her adoptive father and her closed friends.)

11. This chunk was not identified in the conversational data.

Table 17. (continued)

Left context	<i>congciyihou</i>	Right context
(60) 司空化羽道：“这……有是有些，但自有了你，我就与她们断了往来。”华忆红面泛青光，冷笑道：“哼，我告诉你小子， (Summary: The girl is demanding absolute loyalty from her lover by threatening to kill any woman he would dare flirt with.)	从此以后， 'from now on,'	你若再与哪个女人眉来眼去，我就杀了她！ 'if you dare flirt with any woman again, I will kill her!'
(61) 丈夫的不信任伤透了她的心，她哭了，一向疼爱他的丈夫见到眼泪也心软了，便向她道歉，他们暂时和好了，但 'Her husband's lack of trust broke her heart. She cried. Seeing her crying, her husband, who always loved her, apologized to her. They made up temporarily. However,'	从此以后， 'ever since then,'	她的感情深处便留下了一道可怕的阴影。 'it has been left in the deep place of her heart with a horrible shadow.'
(62) 我从没有这样激动的情绪，因此我经历了以前从未经历过， 'I had never experienced such kind of excitement. (That time) I experienced what I had never experienced before,'	从此以后， 'and ever since then,'	再也未能经历到的这样的狂喜。 'the rapture that I never experienced for a second time'
(63) 天知道他和琼之间那道不可逾越的鸿沟在一夜之间消失了， 'Tian knew that the previous impassable gulf between Qiong and him no longer existed.'	从此以后， 'Ever since then,'	天将琼的幻影从他美妙的梦境里永远地驱逐了出去！ 'Qiong no longer was in Tian's fantasies.'
(64) 原来，乾隆为了讨得尼姑的欢心，下了一道圣旨，在从蓝靛厂到南海淀的大路上修起了一条买卖街，从苏杭一带迁来了商贾、店铺，一时间把一个荒凉的地方变成了江南。 (Summary: In order to please a woman who was a nun, a famous king in Chinese history turned a desolate place into a prosperous shopping street that modeled the style of the nun's hometown.)	从此以后， 'Ever since then,'	他就可以经常携带那尼姑来此欣赏苏杭风光，以解尼姑思乡之愁。 (Summary: The king could often take the nun to this place to appreciate the sceneries of her hometown and make her less homesick.)

continued

Table 17. (continued)

Left context	<i>congciyihou</i>	Right context
(65) 我又一次在深夜走在大街上，在转角处又看见那个水果摊，还是那对母女，我想去买一些水果，谁知走近一看，却又什么也没有。 (Summary: A person went out at midnight and came across a mother and her daughter who turned out to be ghosts.)	从此以后， 'Ever since then,'	我再也没有看见过这个水果摊了。 (Summary: The person never again saw that mother and her daughter selling fruit on that street.)

First of all, these tokens mostly (87.5%) appear as a freestanding clause followed by a punctuation mark. This suggests that the string *congciyihou* is functioning as a fixed lexical unit.

Another common feature is that the demonstrative pronoun 此 *ci* ‘this’ functions as an anaphor referring to the turning point of a story, which is elaborated in the left context. In (58), the antecedent event is ‘giving birth to a baby bear.’ In (59), the antecedent event is ‘leaving a place (forever) with one’s newly wedded husband.’ In (60), the antecedent event is ‘having met one’s sweetheart in life (a lady who later became his wife).’ In (61), the antecedent event is a heartbreaking problem between a woman and her husband. In (62), the antecedent event is ‘experiencing an intense emotion that one has never experienced before.’ In (63), the antecedent event is that a man finally had a chance to spent one very special night with a woman he secretly loved, a night that completely changed the relationship between him and the woman and overcome the previous impassable gulf between them. In (64), the antecedent event is that in order to please a woman (a nun), a famous king in Chinese history turned a desolate place into a prosperous shopping street that modeled the style of the nun’s hometown. In (65), the antecedent event is that a person went out at midnight and came across two ghosts. All these instances have a common feature – highly dramatic. No mundane events are found to be the antecedent event of *congciyihou* in the two written corpora (LCMC and UCLA).

Turning to the right context that follows *congciyihou*, again, one can easily notice a striking common feature – the following events are all extraordinary circumstances with an exceptional dramatic connotation (such as life threatening or forever farewell) indicating a long-term significant consequence. In (58), the following event is that the bear let down her guard against the hunter. In (59), a woman who has left her home place with her new husband can no longer come back to reunite with her adoptive father and close friends for the rest of her life. In (60), a wife is threatening her husband that she would kill him if he dares flirt with other women again. In (61), a dreadful shadow has gripped a woman’s heart

and darkened her emotional world. In (62), a man never goes into a state of ecstasy as intense as the one he has experienced before. In (63), spending a night with a man took away all the mystery from the lady and made the man no longer have any feelings for her. In (64), a famous Chinese king has finally achieved his aim to please a nun that strongly attracts him. In (65), the person no longer had the second chance to see the ghosts that she met before.

Based on all of these incidents, it is clear that the chunk *congciyihou* found in the written corpora is strongly associated with dramatic events that have a long-term significant impact. This example shows that chunks in the written discourse tend to be highly clear-cut.

To summarize, this section discusses the finding that conversational and written discourses have different chunking patterns. Chunking in conversation has different stages. That is, the same string of lexical items can form both a partially pre-stored chunk and an entirely pre-stored chunk, or several different partially pre-stored chunks. On the other hand, chunks in written discourse occur mainly at the final clear-cut stage, which means that a chunk is either a partially pre-stored chunk or an entirely pre-stored chunk.

7. Conclusions and implications

This study investigates the different semantic meanings and chunking patterns that words have in written and conversational discourses. Unlike previous studies that use a single word or grammatical construction, this study chooses two near-synonymous words in Mandarin, *zhihou* and *yihou*, and examines their semantic distinctions and formulaic chunks in written and conversational corpora respectively.

My findings regarding semantic distinctions are: (1) In both writing and conversation *zhihou* favors past and *yihou* favors future; (2) in writing but not in conversation *zhihou* is more often used with immediate high transitivity actions and causal relations, whereas *yihou* is more often used with low transitivity states. Findings regarding chunking patterns are: Whereas conversation preserves different stages of chunking, written discourse mainly has the final clear-cut stage of chunking. The findings point to an important fact that different discourse types can have very different linguistic configurations.

This study demonstrates the importance of grounding grammatical investigations on discourse types. The idea of recognizing written and spoken discourses as two major components of language usage is not merely a matter of “style.” As Miller and Weinert (1998: 4–5) observe: “The terms ‘spoken language’ and ‘written language’ do not refer merely to different mediums, but relate to partially

different systems of morphology, syntax, vocabulary, and the organization of texts. These facts are crucial for investigations of language but are generally ignored in theoretical syntax.” Likewise, the concept of a general grammar based on genre-unspecified constructed data is criticized in Emergent Grammar (Hopper 1987, 1988, 1998). Hopper (1998) notes that “language is not a general abstract possession that is uniform across the community. ...As more and more genres are gathered in, the core grammar common to all of them shrinks vanishingly, until it may be doubted whether there is, ultimately, such a thing as English grammar in any comprehensive sense.” Hopper (1998) suggests that written and spoken versions of the same language can be as distinct as two grammatically distinct languages. Tao (1999) also calls for a shift from “the grammar” to “grammars.” Tao (1999) discusses the significance of discourse-type based grammatical investigations and questions a general grammar existing independent of all discourse types. Tao points out that “a serious consideration of this question and systematic studies of discourse types could open new vistas to grammar.” In this connection, the model of Multiple Grammars (Iwasaki, 2015) even considers that the entire grammar consists of multiple spoken and written component grammars at the level of usage-based grammar. Iwasaki’s study outlines a general hypothesis that speaker’s grammatical knowledge is composed of spoken grammar and written grammar.

This study demonstrates that different discourse types can have very different linguistic configurations – even when it comes to the subtle semantic distinction between near-synonymous words. By carefully comparing the chunking patterns of near-synonymous words, this study also reveals how chunking patterns in written discourse differs from that in conversational discourse.

This study may also have methodological implications for the study of grammar in discourse. Whereas previous studies usually compare a single word or grammatical construction in multiple corpora, this study shows how, instead of using a single word or grammatical construction, using near-synonymous words or grammatical constructions, can yield new findings. In this study, the written corpora are balanced with sizable data representing various written genres, but the spoken data is of a singular genre – ordinary telephone conversation. They may seem to be incomparable. Actually, they are incomparable. In fact, there can hardly be any balanced spoken corpora that are comparable with balanced written corpora. The methodology of this study is designed to provide a solution for this problem.

This study does not compare the word *zhihou* in written versus conversational data, nor does it compare the word *yihou* in written versus in conversational data. Doing this would be inappropriate, because the written data and conversational data are very different. In fact, I believe that even if the discourse types compared are both of conversational discourse, they can still be very different. For example,

telephone conversation among strangers can differ from face-to-face conversation among friends and family members. Telephone conversation between strangers might have more features of written discourse than face-to-face conversation in that: (1) Unlike face-to-face conversation where speech can be aided by gestures and other bodily semiotic recourses, telephone conversation relies mainly on linguistic channel and the communication is merely done by wording. (2) Compared to friends and family members, strangers share relatively limited common knowledge, a situation resembling the written discourse where the writer and the invisible readers share limited common knowledge. As a result, the speakers on the phone are more obliged to make their meaning precise, which may lead to the use of long and completed structures as usually seen in scientific writing. (3) Conversation with friends and acquaintances feature fewer action events and more emotional and epistemic exchange (Thompson & Hopper 2001). In comparison, talks among strangers may focus more on the exchange of information, and thus have more concrete semantic meaning.

Using the new method of comparing a pair of near-synonymous words, this study first compares the usages of the two words within the written database, which are comparable, because instances of the two words come from the same dataset. This study then goes on to compare the usages of these two words in the conversational dataset, which are again comparable. In this way, no matter how different the corpora are, the comparison is conducted within each dataset. Because of this, investigators would have no problem with the comparability of the corpora used. This is why I argue that a more reliable way to compare different discourse types would be to use pairs of near-synonymous words or grammatical constructions.

Finally, this study also shows an advantage of using conversational data for linguistic research. It is found that conversation preserves different stages of chunking, which can provide a window into the process of chunking. On the other hand, written discourse seems mainly to have the final fixed stage, rendering the process of linguistic change opaque and difficult for linguists to restore with synchronic data.

To summarize, this study argues that discourse type should be a factor in our attempts to understand grammar. A more precise and specific taxonomy of discourse types is not only beneficial but also indispensable to modern linguistic studies. The methodology of this study also provides a way to compare seemingly incomparable corpora.

Acknowledgements

I am very grateful to Hongyin Tao for his careful reading of the drafts of this paper and his enlightening comments. The insightful questions and comments from Shoichi Iwasaki were also very helpful in preparing this paper. I also want to thank the two anonymous reviewers for their valuable and constructive comments. The author alone takes responsibility for the paper's remaining shortcomings.

References

- Anthony, Laurence. 2011. AntConc (Version 3.3.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net>
- Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511621024
- Biber, Douglas. 1999. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies* 1.2: 131–150. doi: 10.1177/1461445699001002001
- Biber, Douglas and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge, New York, Melbourne: Cambridge University Press. doi: 10.1017/CBO9780511814358
- Bladas, Oscar. 2012. Conversational routines, formulaic language and subjectification. *Journal of Pragmatics* 44.8: 929–957. doi: 10.1016/j.pragma.2012.04.009
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526
- Canavan, A. and G. Zipperlen. 1996. *CALLFRIEND Mandarin Chinese-Mainland Dialect*. Linguistic Data Consortium, Philadelphia.
- Centre of HSK at Beijing Language and Culture University (北京语言文化大学汉语水平考试中心) (Ed.). 2000. *HSK 汉语8000词词典* [HSK Guideline: A Dictionary of 8000 Chinese Words], Beijing: Beijing Language and Culture University Press.
- Clancy, Patricia. 1982. Written and spoken style in Japanese narratives. *Spoken and Written Language: Exploring Orality and Literacy*, ed. by D. Tannen. Norwood, 55–75. Norwood, N.J.: Ablex Publish Corp.
- Coulmas, Florian (ed.). 1981. *Conversational Routines: Explorations in Standardized Communication Situations and Prepatterned Speech*. The Hague: Mouton.
- Fox, Barbara and Sandra Thompson. 2007. Relative clauses in English conversation: relativizers, frequency and the notion of construction. *Studies in Language* 31: 293–326. doi: 10.1075/sl.31.2.03fox
- Gong, Xuesheng (ed.) 2007. *当代汉语词典* [Contemporary Chinese Dictionary], 商务印书馆国际有限公司.
- Grant, Lynn. 2003. A corpus-based investigation of idiomatic multi-word units. Unpublished PhD Thesis. Victoria University of Wellington.
- Heine, Bernd. 2010. Elements of discourse grammar: an outline. Paper presented at Chonnam National University, Gwangju, 8–27–2010.

- Hopper, Paul. 1987. Emergent grammar. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. J. Aske, N. Beery, L. Michaelis and H. Filip. pp. 139–155. doi: 10.3765/bls.v13i0.1834
- Hopper, Paul. 1988. Emergent grammar and the a priori grammar. *Linguistics in Context: Connecting Observation and Understanding*, ed. by Deborah Tannen. Norwood, New Jersey: Ablex Publishing Company. pp. 117–134.
- Hopper, Paul. 1998. Emergent Grammar. *The New Psychology of Language: Cognitive and Functional Approaches to Linguistic Structure*, ed. by Michael Tomasello. Englewood Cliffs, NJ: Erlbaum, pp. 155–175.
- Hopper, Paul, and Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56.2: 251–299. doi: 10.1353/lan.1980.0017
- Hui, Yu (ed.) 2002. *A New Century Chinese-English Dictionary*, Foreign Language Teaching and Research Press.
- Iwasaki, Shoichi. Forthcoming. Model of Multiple Grammars.
- Jing-Schmidt, Zhuo, and Hongyin Tao. 2009. The Mandarin disposal constructions: usage and development. *Language and Linguistics* 10.1: 29–58.
- Li, Xingjian (ed.) 2004. 现代汉语规范词典 [A Standard Dictionary of Modern Chinese], Beijing: 外语教学与研究出版社 [Foreign Language Teaching and Research Press], 语文出版社 [Language and Literature Press].
- McEnery, Tony, and Richard Xiao. 2004. *The Lancaster Corpus of Mandarin Chinese*, Lancaster.
- Miller, Jim and Regina Weinert. 1998. *Spontaneous Spoken Language – Syntax and Discourse*. Oxford (England): Clarendon Press.
- O'Donnell, Mick. (2007) UAM CorpusTool [Computer Software]. *Wagsoft Linguistic Software*. <http://www.wagsoft.com>
- Pawley, Andrew and Frances Hodgetts Syder. 1983. Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics* 7.5: 551–579. doi: 10.1016/0378-2166(83)90081-4
- Shao, Jingmin (ed.). 2000. *HSK汉语水平考试词典*, 华东师范大学出版社.
- Sinclair, John. 2003. *Reading Concordances*. London: Pearson/Longman.
- Sohn, Sung-Ock. 2010. The role of frequency and prosody in the grammaticalization of Korean *-canh-*. *Formal Evidence in Grammaticalization Research*, ed. by An Van Linden, Jean-Christophe Verstraete, and Kristin Davidse, 245–273. John Benjamins. doi: 10.1075/tsl.94.09soh
- Su, Danjie. (Forthcoming). Significance as a Lens: Understanding the Mandarin Ba Construction through Discourse Adjacent Alternation, *Journal of Pragmatics*.
- Tao, Hongyin. 1996. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam/Philadelphia: John Benjamins. doi: 10.1075/sidag.5
- Tao, Hongyin. 1999. Discourse taxonomies and their grammatico-theoretical Implications. *Dangdai Yuyanxue* [Contemporary Linguistics], 1.3: 15–24.
- Tao, Hongyin and Charles F. Meyer. 2006. Gapped coordinations in English: Form, usage, and implications for linguistic theory. *Corpus Linguistics and Linguistic Theory* 2.2: 129–163.
- Tao, Hongyin, and Richard Xiao. 2007. *The UCLA Chinese Corpus* (1st Edition). UCREL, Lancaster.
- Thompson, Sandra. 2002. “Object complements” and conversation: Towards a realistic account. *Studies in Language* 26.1: 125–164. doi: 10.1075/sl.26.1.05tho

- Thompson, Sandra and Paul Hopper. 2001 Transitivity, clause structure, and argument structure: Evidence from conversation. *Frequency and the Emergence of Linguistic Structure*, ed. by Joan L. Bybee, and Paul J. Hopper, 27–60. Amsterdam: John Benjamins. doi: 10.1075/tsl.45.03tho
- Van Lancker, Diana. 2004. When novel sentences spoken or heard for the first time in the history of the universe are not enough: Towards a dual-process model of language. *International Journal of Language & Communication Disorders*, 39.1: 1–44.
- Wang, Song. 2009. “以后”相关问题再考察及留学生偏误分析. [Rethinking *yihou* and errors in second language acquisition], *Haiwai Huawen Jiaoyu* [Overseas Chinese Education], 51.2: 46–50.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511519772
- Yi, Ping. 1997. “以后、之后、后、后来”意义和用法比较 [A comparison study of the meanings and usages of *yihou*, *zhihou*, *hou*, and *houlai*], *Yuwen Zhishi*, [Knowledge of Language and Literature], 2: 12–14.
- Zhao, Jianhua. 2014. 现代汉语中单用的“之后”“以后”辨析 [The free-standing use of *zhihou* and *yihou* in Modern Chinese]. *Wenxue Jiaoyu* [Literature Education], 7. 48–49.
- Zhou, Yonghui, Sicheng He, & Yuanbi Fu. 2004. 同义词反义词词典 [A Dictionary of Synonyms and Antonyms]. 四川辞书出版社.
- Zhou, Qingyan & Jingjing Zhang. 2005. “以后”和“后来”[*Yihou* and *houlai*] *Zhanjiang Haiyang Daxue Xuebao* [Journal of Zhanjiang Ocean University], 25.2: 107–111.
- 汉语大词典编委会. 1999. 汉语大词典 [A Grand Chinese Dictionary], 汉语大词典出版社.
- 现代汉语辞海编委会. 2002. 现代汉语辞海 [Modern Chinese Cihai]. 印刷工业出版社.
- 新华汉语词典编委会. 2004. 新华汉语词典 [Xinhua Chinese Dictionary]. 商务印书馆国际有限公司.

Author's address

Danjie Su
Department of Asian Languages and Cultures
University of California, Los Angeles
Los Angeles, CA 90095-1540
USA

danjie.su@ucla.edu