

# A Corpus-Based Study of Counterfactuals in Mandarin\*

Qian Yong

*Jinan University*

Language and Linguistics  
17(6) 891–915  
© The Author(s) 2016  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1606822X16660505  
lin.sagepub.com



This research aims to discover the working mechanisms for counterfactuals in Mandarin. Following statistical analyses, several key findings emerged. First, besides pragmatic influences, Mandarin counterfactuals are found to be positively correlated with negation, emphatic modals, optative mood, past-oriented temporality, first person pronoun, demonstratives, and so on. The strength of pragmatic influences is subject to the variation of features present in the sentences. Second, different hypothetical conjunctions work quite differently in their ability to generate counterfactuality. Third, some significant differences have been found both in the frequencies of counterfactuals and in marking strategies between Mandarin and English. My empirical findings were confirmed by a search through six corpora, namely the Lancaster Corpus of Mandarin Chinese, The University of California at Los Angeles (UCLA) Corpus of Written Chinese, Texts of Recent Chinese (TORCH), CLOB corpus (Brown family, British English), CROWN corpus (Brown family, American English) and The English–Chinese Bilingual Parallel Corpus.

Key words: CFI-principle, chi-square test, corpus, counterfactual, phi-coefficient correlation

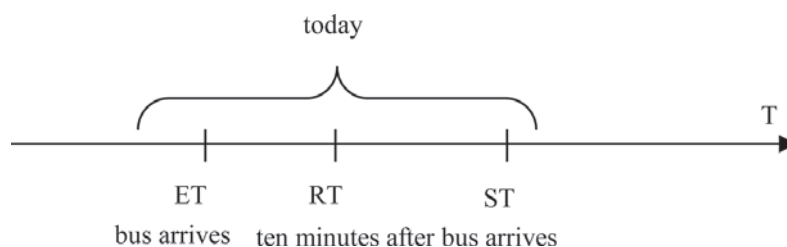
## 1. Introduction

Counterfactuals were not recognized as an independent category until Li (1924[1992]) first acknowledged the concept of hypothetical conditionals as an extended category of conditionals. Similarly, Lü (1942[1982]) and Wang (1985) also realized the necessity of separating conditionals contrary to known facts from general hypothetical conditionals. The issue of different treatments towards counterfactuality in Chinese from other world languages was first proposed by Bloom (1981). He further remarked that because Chinese does not have a specialized schema for counterfactual expression, Chinese people are less competent or efficient with regard to counterfactual reasoning. However, experiments by the likes of Au (1983), Liu (1985) and Yeh & Gentner (2005) reveal that Chinese speakers have almost no difficulty in understanding counterfactual logic. Some

---

\* This paper is part of a preliminary version of 「違實句的形態類型及漢語違實句」 which was presented at the joint conference of the 26th North American Conference on Chinese Linguistics (NACCL-26) and the 22nd Conference of the International Association of Chinese Linguistics (IACL-22), University of Maryland, 2014. This paper is supported by the Fundamental Research Funds for the Central Universities (grant no. 15JNYH007) – Constructing a Semantic Knowledge Network of English Words: A Multimodal Perspective. I also thank the anonymous reviewers of this paper and Professor Elizabeth Zeitoun for their insightful and supportive comments. Special thanks go to Professor Matthias Gerner for his supervision during the whole process of my Ph.D. research.

native Chinese speakers, such as Chen (1988) and Wu (1989), argue that Bloom's theoretical assertions about the lack of counterfactual schemes in Mandarin are highly problematic. Wierzbicka (1997) notes the distinction can be made quite clearly through time adverbials and the modal *buhui* (不會 'would not'). Similarly, Tien (1994) discusses the clearly distinguished counterfactual conditional in Taiwanese Min. Although many attempts to discover the markings of counterfactuals in Mandarin have been made, Jiang (2000) rejects all of them as dedicated counterfactuality markers on the logical grounds that they are neither *sufficient* nor *necessary* conditions for counterfactuals. However, Jiang does not negate the formal properties of counterfactuals in Mandarin, instead claiming that most counterfactuals in Mandarin can be detected through some grammatical, lexical, or logical features, such as fake tense (backshifting of tense), fake aspect (aspect coercion), sentence final *le* (了 perfective), modal adverbial *zhende* (真的 'really'), negation of protasis and contraposition (if  $p \rightarrow q$ , then  $\neg q \rightarrow \neg p$ ). However, Jiang's discussion of fake tense/aspect in Mandarin seems to be problematic in light of a further contrastive study between counterfactual temporal markers in Indo-European languages and Mandarin. For example, *zao* (早 'early') is taken by Jiang (2000) as a fake past tense because, when occurring in counterfactuals, it loses its temporal deictic function as it can refer to as distant a remote past as *dangshi* (當時 'at that time'), or as close a near past as *zuotian* (昨天 'yesterday'). For example, consider 昨天要是早知道會弄成這樣，我就會多帶點兒錢去 (*If I had known this earlier yesterday, I would have brought more money*). It seems that the pluperfect reference of *zao* plus the past tense reference is fake since the above sentence is a counterfactual with a simple past reference. However, a closer look reveals that *zao* does not have any function of tense; for example 今天巴士早到了十分鐘 (*The bus was ten minutes earlier today*). Here *zao* is employed to indicate the relation between ET and RT as Figure 1 shows:



**Figure 1:** Temporal reference of *jintian bashi zao dao le shi fenzhong*

The application of *zao* is independent of the relation between ET and ST. It can even occur in a sentence with future reference <ST, ET, RT>, as in 明天巴士會早到十分鐘 (*The bus will be ten minutes earlier tomorrow*). Therefore, *zao* is not a tense operator and the combination of *zao* and past temporal words will not lead to pluperfect understanding either in or out of conditionals.

Chen (1988) and Jiang (2000) highlight aspect prohibition in Mandarin counterfactuals alluding to the example 昨天要是下一場雨，莊稼就不會乾死了 (*If it had rained yesterday, the crops would not have died*). They argue that the prohibition on aspect marking in the above sentence is a reflex of fake aspect in Chinese counterfactuals. However, even if aspect prohibition is needed for marking counterfactuals, it does not function as a fake aspect like the transposed use of imperfective to perfective in Romance languages.

Feng & Yi (2006) carried out a corpus study to calculate the percentages of counterfactual responses for markers mentioned by Wu (1994)<sup>1</sup> and in other earlier research. Using these markers as key words, three native Chinese speakers were asked to pick out 200 counterfactuals containing these markers either from the Center for Chinese Linguistics PKU (CCL) corpus or from internet search engines. However, the authors did not specify how the Chinese speakers should select the 200 counterfactuals. If they used the markers as key words and picked out the required number of counterfactuals, they would possibly consciously monitor the percentage of counterfactuals triggered by each marker. Wang (2012) went further by conducting a comprehensive study regarding the formal features of counterfactuals in Mandarin. However, Wang's work may have been largely based on an intuitive approach, whereby researchers invent purer examples instantly for analyses but which may not truly reflect the real picture of natural language. In the following sections, I shall try to develop a model of counterfactuals in Mandarin by applying both empirical and theoretical analyses.

## 2. Methods

Studies of Mandarin counterfactuals cannot be conducted without recourse to attested language data. Several corpora will be introduced here in both developing and testing all the previous studies concerning Mandarin counterfactuals. Prior to conducting corpus analyses, I need to narrow my research scope as it is almost impossible to make an exhaustive study of Mandarin counterfactuals in mass corpora considering their great variety and variability. Therefore, the question concerning what the typical Mandarin counterfactual is needs to be solved beforehand. In early research, counterfactuals were often regarded as a special subclass of conditionals in Mandarin. This was also true for other languages where counterfactuals were often termed as subjunctive conditionals in contrast with indicative conditionals. The close relationship between counterfactuals and conditionals is assumed in the widely accepted definition which states that 'a counterfactual statement is one which presupposes that the antecedent is false' (Barwise 1986). More than occasionally, counterfactuality is realized under the framework of conditionals with hypothetical conjunctions in Mandarin. This study therefore aims to make a quantitative assessment towards the correlation between the potential markers and counterfactual reading under the framework of hypothetical conjunctions.

### 2.1 Corpora

I used three corpora in my study of Mandarin counterfactuals, including the Lancaster Corpus of Mandarin Chinese version 1 (LCMC), the University of California Los Angeles (UCLA) Corpus of Written Chinese, 2nd edition, and the Texts of Recent Chinese (TORCH) Corpus. These three corpora are compiled following the same format with a sampling frame of 15 genres. Among these three corpora, TORCH is a recently-built balanced corpus<sup>2</sup> as a match for CLOB (Brown family,

---

<sup>1</sup> Wu (1994) carried out an interesting study to find some strategies in counterfactual marking by asking native Chinese speakers to identify counterfactual sentences from newspapers and looking for recurring linguistic forms in the sample.

<sup>2</sup> The sampling frame of the TORCH corpus strictly follows the CLOB/CROWN, except for the minor variation that western and adventure fiction is replaced by martial arts fiction in TORCH.

British English)/CROWN (Brown family, American English), which provides a sound basis for further contrastive studies of counterfactuals between Mandarin and English. CLOB and CROWN, as two Brown family corpora, share exactly the same properties except that the former samples British English and the latter American English. Additionally, the English–Chinese Bilingual Parallel Corpus (<http://www.luweixmu.com/ec-corpus>) will be introduced to explore how counterfactuals in English are translated into Chinese. Some further information about these corpora is listed in Table 1:

**Table 1:** Basic statistics of six corpora

Corpus	Size	Built-up time
Lancaster Corpus of Mandarin Chinese version 1 ( <b>LCMC</b> )	1 million words	2003
The UCLA Corpus of Written Chinese 2nd edn ( <b>UCLA</b> )	1.12 million words	2012
Texts of Recent Chinese (Brown family, 2009, 2013 summer edn) ( <b>TORCH</b> )	1.29 million words	2013
CLOB Corpus (Brown family, British English) ( <b>CLOB</b> )	1.15 million words	2012
CROWN Corpus (Brown family, American English) ( <b>CROWN</b> )	1.15 million words	2012
英漢雙語平行語料庫 (The English–Chinese Bilingual Parallel Corpus)	4.10 million English words 7.15 million Chinese characters	2007

## 2.2 Data processing

Prior to conducting the corpus-based analyses, some commonly-used hypothetical conjunctions need to be determined based on their usage frequencies. To achieve this quantitative step, ‘c\_ (tag for conjunctions)’ is searched as a keyword in corpora, resulting in 13 different hypothetical conjunctions. These are 如果, 如若, 如, 倘使, 倘若, 若, 假如, 假使, 假若, 假設, 要是, 要不是, and 若非. Using these conjunctions as key words, we collected more than 3000 conditionals. Since Mandarin Chinese has very few dedicated counterfactual markers, the only method to pick out counterfactuals was to manually go through every example and decide individually whether the sentence is counterfactual or not. Some conditionals showed obvious counterfactual inferences, such as 如果我是外國人... (*If I were a foreigner...*). However, some conditionals showed ambiguous meanings with either counterfactual or non-counterfactual inferences possible, such as 如果電影院的觀眾都不認識我倒也罷了 (*If members of the audience in the cinema did not know me, it would be nothing*). The only way to judge whether it is a counterfactual was to go back to the context for clues about the factuality. Luckily, the sentence following shows that the speaker’s family was just sitting beside him. Therefore, it is a counterfactual with false presupposition. For some conditionals, the clues about the factuality may not be clear, such as 如果有一顆紅豆送給可欣那該多好啊! (*If only I could give Kexin a red bean*). The context does not show whether the speaker got a red bean or not. Therefore, it may not be appropriate to take it as a counterfactual. In this way, we collected 245 counterfactuals. All these counterfactuals show some common formal features. Further statistical tests were made towards the summarized formal features. The software SPSS was used to perform the statistical analyses.

### 3. Preliminary findings

#### 3.1 The influences of hypothetical conjunctions

Mandarin shows various uses of hypothetical conjunctions which are connected with hypotheticality at different levels. Chao (1968) ranks the hypothetical conjunctions in a hierarchy according to their counterfactuality-generating capability: 要是>要>假如>若是>倘若>假若>假使>倘使>設若. Lü (1942[1982]) holds a different opinion, arguing that conjunctions in classical Chinese like 使, 令, 設 and 向使 are closely connected to counterfactual reading, while the hypothetical conjunctions in modern Chinese are connected to counterfactuality to a lesser degree. Jiang (2000) supports Lü's proposal in arguing that however significant the correlation between hypothetical conjunctions and counterfactuality is, the appearance of hypothetical conjunctions cannot ensure a counterfactual reading. Therefore, we have no standard to measure the counterfactuality-generating ability of these conjunctions. I agree that hypothetical conjunctions are neither a sufficient nor necessary condition for counterfactual expressions, but I cannot deny the contributions of these conjunctions in counterfactual expression. To compare the degrees of probability connoted by these conjunctions, I made a distribution analysis of the selected 245 conditional counterfactuals after an initial concordancing of the 13 lexical items across the three corpora (Table 2).

**Table 2:** Proportions of counterfactuals under each hypothetical conjunction

Hypothetical conjunction	Proportions of counterfactuals (%)		
	LCMC	UCLA	TORCH
ruguo (如果)	5.3	4.8	6.4
ruruo (如若)	0.0	75.0	20.0
ru (如)	0.0	0.0	0.0
tangshi (倘使)	0.0	N/A	N/A
tangruo (倘若)	9.1	31.2	0.0
ruo (若)	13.1	1.6	10.5
jiaru (假如)	25.0	23.5	23.5
jiashi (假使)	100.0	N/A	N/A
jiaruo (假若)	100.0	0.0	0.0
jashe (假設)	33.3	10.0	14.3
yaoshi (要是)	21.1	32.3	15.4
ruofei (若非)	75.0	100.0	66.7
yaobushi (要不是)	100.0	100.0	100.0

Since different conjunctions in corpora have quite different usage frequencies, to make a comparison of counterfactuality-generating ability more meaningful the proportion of counterfactuals is used here. A comparison of the mean ranks for the proportions of counterfactuals under these conjunctions seems to show an obvious difference among them:<sup>3</sup>

<sup>3</sup> Note that *tangshi* (倘使) and *jiashi* (假使) are not recorded in either UCLA or TORCH.

**Table 3:** Mean ranks for the proportions of counterfactuals under 11 hypothetical conjunctions

	yaobushi	ruofei	jiaru	yaoshi	jashe	ruruo	jiaruo	ruo	tangruo	ruguo	ru
Mean rank	10.67	9.83	7.33	7.00	6.33	6.17	4.67	4.33	4.33	3.67	1.67

It should be noted that the hierarchy revealed in Table 3 may not be a strict one-to-one correspondence to counterfactuality-generating ability since counterfactuals in Mandarin are largely influenced by pragmatic inferences. But at least it presents us with a rough picture with *yaobushi*, *ruofei*, *jiaru*, *yaoshi* indicating events which are less probable than where *ruguo* and *ru* occur.

## 3.2 Frequency profiles of features included in the conditional counterfactuals

### 3.2.1 Negation

There are many ways to express negation in Mandarin, besides the commonly used adverbs *bu* (*hui/gai*) (不會/該) ‘will/should not’ and *mei* (*you*) (沒[有]) ‘do not have’; *bushi* (不是) ‘be-not’ comes out as a highly lexicalized negative conjunction from *bu+shi*. Additionally, the rhetorical question also has the illocutionary force of an assertion of the opposite polarity from what apparently was asked (Sadock 1971, 1974). Therefore, a positive rhetorical question may have the illocutionary force of a negative assertion. Based on my data, *meiyong* ... *bu* is the most common collocation, and they have similar frequency distributions.

A repeated phi-coefficient test will uncover the degree to which each negative expression correlates with counterfactuality.

**Table 4:** The frequency of non-counterfactuals and counterfactuals with and without *bushi* (不是) (under conditional framework)

	Non-counterfactuals	Counterfactuals
<i>bushi</i>	6	75
Without <i>bushi</i>	2778	170

**Table 5:** The frequency of non-counterfactuals and counterfactuals with and without *meiyong* (沒有) (under conditional framework)

	Non-counterfactuals	Counterfactuals
<i>meiyong</i>	94	45
Without <i>meiyong</i>	2690	200

**Table 6:** The frequency of non-counterfactuals and counterfactuals with and without *bu* (不) (under conditional framework)

	Non-counterfactuals	Counterfactuals
<i>bu</i>	415	93
Without <i>bu</i>	2369	152



**Table 7:** The frequency of non-counterfactuals and counterfactuals with and without rhetorical question (under conditional framework)

	Non-counterfactuals	Counterfactuals
Rhetorical question	23	25
Without rhetorical question	2761	220

The  $\phi$  value of *bushi*, *meiyou*, *bu* and the rhetorical question is calculated as 0.514, 0.195, 0.168, and 0.198, respectively, based on the data from Tables 4–7. Significant positive correlation even at the 0.1% level can therefore be concluded between each negative expression and counterfactuality.

*Bushi* is the one most likely to generate counterfactuality ( $\phi = 0.514$ ) under conditional conjunctions. This, perhaps, could be explained by its leftward position in the sentence. *Bushi*, usually only preceded by conjunctions, appears at the left periphery of the sentence. More than occasionally, *bushi* could be moved to the leftmost location with elided conjunctions, like 不是你出手大方，他早就餓死了 (*Had you not been so generous, he would have starved to death long ago*). This may shed some light on the universal left peripheral position for counterfactual markings. In some languages, the protasis of the conditional counterfactuals is marked by inversion of the finite verb, like *had I known*. Evidence in Turkish also reports a left movement of the past tense when it is counterfactually interpreted: *-se* (past) ... *-di* (conditional) rather than *-di* (conditional) ... *-se* (past) is used for counterfactuals. Another interesting finding concerns the priority of *meiyou* ( $\phi = 0.195$ ) over *bu* (*hui/gai*) ( $\phi = 0.168$ ) in generating the counterfactuality. My data reveal that *meiyou* always occurs in the protasis of the counterfactual and *bu* (*hui/gai*) in the apodosis. There are also many cases with the collocation of *meiyou* and *bu* (*hui/gai*) in conditional counterfactuals. Negation in the protasis always presupposes a corresponding positive proposition (Ziegeler 2000:40), while negation in the apodosis, which does not have the function of presupposition, presents a relevant consequence of the hypothesis in the protasis, as seen in the example 如果沒有這些鼓勵，我想我是支持不到今天的 (LCMC) (*Without this encouragement, I would not be here today*). Negation in the protasis presupposes the fact that I have already got this encouragement and negation in the apodosis introduces the consequence brought by the protasis. Obviously, the negation in the protasis is the main trigger for the counterfactual reading, which is in accord with our data, where only a few cases are found with negative apodosis in the absence of negative protasis. Interestingly, for all the counterfactuals with negative protasis, the negators tend to have a left-periphery location. For the 45 cases with *meiyou* in protasis, 22 of them have *meiyou* located in the second to the leftmost position (only preceded by a conjunction). Like *bushi*, *meiyou* can also be located in the initial position of counterfactuals like 沒有中國過去三十年的改革開放，就沒有今日的輝煌 (*Without reform and opening up during the past 30 years, China would not have today's performance*).

### 3.2.2 Emphatic modal adverbs

Another specific formal feature of these conditional counterfactuals across the three corpora is the application of modal adverbs such as *zhende* (真的 ‘really’) in the protasis to emphasize the

speaker's hypothesis, for example 如果植物利用太陽光的效率真的是百分之百…自然，高等植物葉子利用太陽光的效率不可能是百分之百 (TORCH) (*If all the sunshine were absorbed by plants ... actually, it is not possible for higher plants*).

Upon close observation of selected conditional counterfactuals, I find at least three different varieties of emphatic modal adverb with quite similar meanings to *zhende*, and calculate the frequency for each of them (Table 8):

**Table 8:** Varieties of modal adverbs in counterfactuals and their frequencies

真(的/是)	果真	當真	Total
25	2	1	28

The frequent appearance of emphatic modal adverbs in counterfactuals could be explained by their function of strengthening the hypothesis in the protasis. Compare the following examples:

- (1) A: 他昨天在研討會上發言了嗎? (*Did he make a speech at yesterday's seminar?*)  
 B: 發了。 (*Yes, he did.*)  
 A: 如果他昨天發了言，今天也就會發言。 (*If he did, he will do it today.*)

Or

- A: 如果他昨天真的發了言，今天也就會發言。 (*If he did/had done it, he will/would do it today.*)

In (1), the conditional structure used by A after B's confirmative answer shows A's distrust. And a subtle change with the insertion of *zhende* may emphasize A's distrust. Therefore in the latter sentence with *zhende*, the speaker describes the condition much closer to non-factual, if not counterfactual, than in the former sentence.

Similar to what we have already analyzed, the insertion of *zhen/zhende* may emphasize the speaker's uncertainty as to the realization of the event specified in the protasis, and it may not impose a counterfactual understanding, as in 如果金滿堂真的丟了硬碟，裡面有見不得人的秘密，他應該也不會報案 (TORCH) (*If Jinmantang really lost his hard drive, he would not report it*). According to context, the event of *Jinmantang losing hardware* is kept uncertain by the speaker. In order to ensure a counterfactual reading, the emphatic modal adverbs need to co-occur with other features like past tense reference, negative expression, first person pronoun, and so on. An independent use of emphatic modal adverb may lead to ambiguous understandings.

Despite the obvious lack of thoroughness of emphatic modal adverbs in marking counterfactuality, the phi-coefficient test based on the data in Table 9 still shows some positive correlation between them.

**Table 9:** The frequency of non-counterfactuals and counterfactuals with and without emphatic modal adverb (under conditional framework)

	Non-counterfactuals	Counterfactuals
Emphatic modal adverb	82	28
Without emphatic modal adverb	2702	217



The  $\phi$  value is 0.124, which is significant at even 0.1%. Although the degree of positive correlation is weaker compared with negation, we cannot deny the positive contribution of emphatic modal adverb to counterfactual understanding.

### 3.2.3 Optative mood

Some languages have ways to mark a vaguer hope; for example, in English the verb *wish* rather than *hope* could express a remote hope which is almost impossible to realize, like *I wish you were here*. Similarly in Mandarin, some counterfactuals can be realized in the framework of optative mood by indicating a nearly impossible wish. But unlike fusional or agglutinative languages, Mandarin has no inflectional morphology, but employs lexemes like *hao* (好 ‘good’) in the apodosis. Table 10 shows the various ways of expressing a remote wish in Mandarin by applying *hao* in the apodosis:

**Table 10:** Varieties of optative mood in counterfactuals and their frequencies

…就好了	…(該)多好	Total
6	6	12

The two varieties equal to English *that would be good* share almost the same meaning and usage. Sometimes it is quite difficult to determine whether it is a counterfactual or a hypothetical wish without pragmatic hints, for example 如果有一顆紅豆送給可欣那該多好啊! (UCLA) (*If only I could give Kexin a red bean*). According to context, I do not have clear clues about the realizability of the event related to *giving Kexin a red bean*, therefore it cannot be taken as a counterfactual. Obviously, the remote wish is applied to indicate an event which is far away from reality but which may not be necessarily opposite-to-factual. According to the data, there are only 12 cases in the corpora which use the word ‘good’ as the apodosis. A question may then arise as to whether it is a mere coincidence that good-wish apodosis appears in counterfactuals or if some correlation exists between the two. Likewise, a phi-coefficient test will be applied based on the data in Table 11.

**Table 11:** The frequency of non-counterfactuals and counterfactuals with and without optative mood (under conditional framework)

	Non-counterfactuals	Counterfactuals
Optative mood	5	12
Without optative mood	2779	233

The  $\phi$  value is 0.172, which is significant at the 0.1% level, therefore we can also conclude a positive correlation between the good-wish apodosis and a counterfactual reading in Mandarin.

### 3.2.4 Tense reference

There is sufficient cross-linguistic evidence to support past and perfect morphology as a marker for counterfactuality. As in Mandarin, where past tense reference has not been grammaticalized into a modal with fake temporal reference, there is no one-to-one relationship between a past tense reference and a counterfactual utterance. But even in a language lacking the fake past as a counterfactual marker, we can still assume a close relationship between past events and counterfactual reading since past tense is the product of factivity (Lyons 1977:819–820), and a hypothesis towards a factual known event will naturally produce a counterfactual.

Much has been discussed about the frequent appearance of past tense in counterfactuals; however, it is not easy to detect the past tense reference in Mandarin since it is not inflected as a specialized morpheme (Huang 1987; Li & Thompson 1990; Pan & Hu 2001). Temporal location can be realized through other devices such as temporal adverbs like *zhiqian* (之前 ‘before’), *gangcai* (剛才 ‘just now’) or lexical items like *zuotian* (昨天 ‘yesterday’), *qunian* (去年 ‘last year’) or perfective *le* (了). Additionally, present tense like *xianzai* (現在 ‘now’) in protasis works like a recent past indicator, like 如果我和姚佳現在抱在一起發抖...就會被自己沒來由的想像嚇死 (TORCH) (*If Yaojia and I had trembled in embrace now ... we would have died of fear from our unreasonable imagination*). The event that ‘Yaojia and I’ did not embrace each other has just happened, and the speaker makes a hypothesis towards the recently-happened fact. Therefore, present tense here works like a recent past trigger. Present tense reference in the apodosis can also have a function of indicating the past, since the logically causal relationship between protasis and apodosis is, although not always, reflected in the temporal order, with the protasis preceding the apodosis. Hence, the past location of the event in protasis can be inferred from the present tense reference in apodosis. Based on the analyses above, we therefore presume three major varieties of past tense in Mandarin, namely past temporal words, perfective aspect and present temporal words (Table 12).

**Table 12:** Varieties of past tense indicators in counterfactuals and their frequencies

Past temporal words									Perfective aspect			Present temporal words			Total
Zao (xie/zao/dianr) 早 (些/早/點兒)	dangshi 當時	dangchu 當初	ceng 曾	shixian 事先	guoqu 過去	qunian 去年	yiqian 以前	cong Xiao 從小	le1 了1	le2 了2	yi 已	xianzai 現在	jintian 今天	zhijin 至今	
22	11	5	3	2	2	1	1	1	24	58	7	14	11	1	163

To be noted, some varieties like *zao* (*xie/zao/dianr*) and *le* do not necessarily impose a past location, and the final past reading is generated as a result of interaction between lexicon and context.

However obvious the correlation between past tense reference and counterfactuals in Mandarin is, a phi-coefficient test will be introduced to verify our intuition based on the data in Table 13:

**Table 13:** The frequency of non-counterfactuals and counterfactuals with and without past tense (under conditional framework)

	Non-counterfactuals	Counterfactuals
Past tense	5	72 <sup>4</sup>
Without past tense	2779	173

The  $\phi$  value is 0.506 at significance level 0.1%, which satisfies our intuition that past tense is in a highly positive correlation with counterfactuality.

### 3.2.5 Context-sensitive indexicals

#### 3.2.5.1 First person pronoun

Ziegeler (2000:38) notes that the appearance of first person is quite ubiquitous in counterfactuals, and is likely to create a higher level of counterfactuality than a hypothetical conditional with a second person or third person subject. She further explains that it is because first person is deictically closer to the speaker's immediate domain of reference, and in the case of first person subjects, the subject and the speaker are the same. This intimacy, therefore, is in the best possible position for a factually-based prediction about the past, like 如果他/你/我昨天去了學校，就會看見學校門口的雕像 (*If he/you/I went/had gone to school yesterday, he/you/I would see/have seen the statue near the school entrance*). The third person pronoun in the above sentence generates an open interpretation, since the speaker is not familiar with *his* situation. The second person pronoun, closer to the speaker's domain, will generate an ambiguous interpretation, either being open or counterfactual, depending on whether the event in protasis has already been presupposed to be false, whereas the first person pronoun will undoubtedly produce a counterfactual reading. Unless suffering from a memory loss (and this is another case), the speaker surely knows his own situation. Any hypothesis based on a known reality will denote a counterfactual understanding. To further support our argument, we classify the subjects of counterfactual protasis into five types according to the degree of intimacy, and calculate the occurrence frequency for each type. Our calculation is based on selected counterfactuals across the three corpora (Table 14):

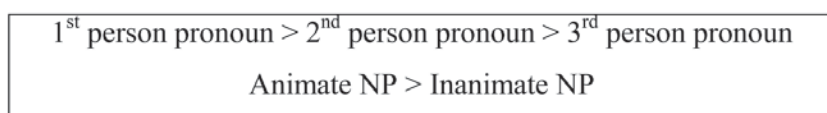
**Table 14:** Occurrence frequency of different subjects in counterfactual protasis

Subject types	1 <sup>st</sup> person pronoun	2 <sup>nd</sup> person pronoun	3 <sup>rd</sup> person pronoun	Animate NP	Inanimate NP
Occurrence frequency	52	19	17	26	14

NP: noun phrase.

<sup>4</sup> The frequencies of counterfactuals with past tense are much lower than the total frequencies of past tense reference outlined in Table 12. This is because, in most cases, two or more past indicators work together to ensure a past location of an event in one counterfactual sentence.

According to our calculation, the first person pronoun occurs most frequently with an overwhelming priority, which coincides with our hypothesis earlier. Moreover, animate NP's are used more frequently than inanimate NP's. A further hierarchy is established according to the statistics in Figure 2:



**Figure 2:** Hierarchy of counterfactuality-generating ability based on intimacy of subject

Those categories that are higher (i.e. further left) on the hierarchy are more intimate to the speaker's cognitive domain than those that are lower (i.e. further right). And another hierarchy of counterfactuality-generating ability mirrors this direction: the left one has greater ability to generate counterfactuality than the right one. Likewise, a phi-correlation coefficient test is introduced here based on the data in Table 15:

**Table 15:** The frequency of non-counterfactuals and counterfactuals with and without first person subject (under conditional framework)

	Non-counterfactuals	Counterfactuals
First person subject	180	40 <sup>5</sup>
Without first person subject	2604	205

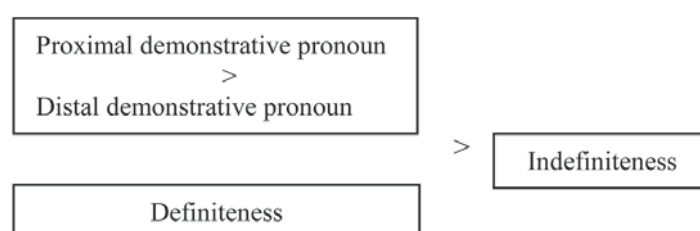
The  $\phi$  value is 0.096, which indicates that there is a significant positive correlation between first person subject and counterfactuality at 0.1%.

### 3.2.5.2 Demonstrative pronoun

The influence of intimacy on enhancing the hypotheticality can also be found in the demonstrative pronouns, where the proximal ('this', 'these') is more inclined to express a counterfactual reading than the distal ('that', 'those'). This is because the proximal demonstrative pronoun pulls the distance between the event and the speaker closer, and the event in the protasis is more likely to be presupposed to be false. Consider the example 要是我拿不到那/這張票，我就不能去看演出了 (*If I had not got this ticket/If I did not get that ticket, I would not see the performance*). The application of *zhe* (這 'this') elicits a counterfactual sense while *na* (那 'that') shows an ambiguous interpretation, either being counterfactual or factual. The effect of intimacy can be further proved by the differences between definite nouns and indefinite nouns: 要是我拿不到票，我就不能去看演出了 (*If I cannot get the ticket, I will not be able to see the performance*). Since indefinite nouns have arbitrary

<sup>5</sup> The number of counterfactuals with first person subject is less than the occurrence frequency of first person subject outlined in Table 14 because there are many counterfactuals with reduplicative uses of first person subject both in protasis and in apodosis.

reference which may fall outside the speaker's knowledge, it is highly unlikely to be able to make a counterfactual presupposition based on this unknown situation. Therefore, the above sentence with an unmarked indefinite object will convey a totally open understanding. Likewise, another hierarchy concerning the counterfactuality-generating ability caused by the definiteness and indefiniteness can be established as follows (Figure 3):



**Figure 3:** Hierarchy of hypotheticality caused by definiteness and indefiniteness

**Table 16:** Occurrence frequency of definite and indefinite NP in counterfactuals

Proximal demonstrative pronoun	Distal demonstrative pronoun	Indefinite NP
48	24	12

Table 16 seems to verify our hypothesis about the counterfactuals favoring definiteness more than indefiniteness. Interestingly, the similarity in distribution between negative expression and demonstrative pronoun can perhaps be explained by the frequent collocations of negative expressions and the demonstrative pronoun in counterfactuals. A question may then arise as to why negative expressions tend to co-occur with a demonstrative pronoun in counterfactuals. Before answering this question, we need to return to the anaphoric function of demonstrative pronouns in Mandarin. When employing *zhe/na* before an NP, the referent of the NP should exist at least theoretically (being referential). Since negation of an expression does not change its presuppositions, we then have a negative expression with positive presupposition. Observe the following examples:

- (2) 中國不會有這/那樣的才子 (*China would not have this/that genius again*)  
 >> This/That kind of talent exists somewhere.
- (3) ...他就不會有這/那樣的提案 (*He would not have this/that proposal*)  
 >> he had this/that proposal.

As an apodosis of a conditional, the sentence presupposes a positive fact in contrast with the literal negative meaning. It is then not difficult to understand why the collocation of negative expression and demonstrative pronoun produces a contrary-to-fact hypothesis.

**Table 17:** The frequency of non-counterfactuals and counterfactuals with and without proximal demonstrative pronoun (under conditional framework)

	Non-counterfactuals	Counterfactuals
Proximal demonstrative pronoun	178	48
Without proximal demonstrative pronoun	2606	197

Based on the data in Table 17, we have calculated the  $\phi$  value as 0.137, which means that the proximal demonstrative pronoun is positively correlated with counterfactuals in Mandarin at the 0.1% level of significance.

**Table 18:** The frequency of non-counterfactuals and counterfactuals with and without distal demonstrative pronoun (under conditional framework)

	Non-counterfactuals	Counterfactuals
Distal demonstrative pronoun	103	24
Without distal demonstrative pronoun	2681	221

The  $\phi$  value based on the data in Table 18 is 0.083, which is significant at the 0.1% level. It seems to verify our hypothesis that the distal demonstrative pronoun is less likely to produce counterfactuality than the proximal demonstrative pronoun, although both of them bear some positive correlation with counterfactuals.

### 3.3 Contextual inference and background knowledge

Apart from the lexical items and morpho-syntactical properties, counterfactuals also allow the hearer to infer the negation of reality from various discourse clues and world knowledge about a situation.

For a language lacking a specialized marking strategy, context may play an essential role in generating a counterfactual reading. Jiang (2000) takes the pragmatic shortcut as an important trigger for counterfactuality in Mandarin, which can work even without assistance from counterfactual features, like 如果電影院的觀眾都不認識我倒也罷了 (TORCH) (*If members of the audience in the cinema did not know me, it would not matter*). Sentences with the modal particle *bale* (罷了 ‘that is all’) at the end are always followed by a *but*-sentence with contrastive factuality. By returning to the context of this sentence in the corpus, we find the following sentence: 關鍵是我的身邊還坐著兩位最熟悉我的人—我的妻子和女兒 (*The key thing is that sitting beside me were two of the people closest to me—my wife and my daughter*). It is obvious that this sentence refutes the hypothesis in the protasis and provides strong evidence to determine the counterfactuality of the sentence.

Further evidence of contextual dependence is that the counterfactuality of a certain sentence is liable to evaporate in certain contexts, either in the immediate linguistic context or in the less immediate discourse context, or in circumstances where assumptions about the uncertainty of realizability of the event outlined in the protasis are made. As noted earlier, a protasis with past tense reference is generally understood counterfactually, as in 如果你剛才沒叫住他，他現在就已經在現場了 (*If you did/had not stop/stopped him, he would be/have been there*). This sentence



would generally be understood counterfactually out of context. However, the counterfactuality can be cancelled by introducing an additional clause before it without contradicting the utterance, like 我不知道你剛才有沒有來這/叫住他，但是... (*I don't know whether you were here/stopped him just now, but ...*). Ziegeler (2000:24) claimed that if an additional sentence presumes no factual knowledge about the propositions presented in either clause of the conditional, then it will act to cancel the implicature of the counterfactuality provided by the content of the protasis.

A common instance of the pragmatic counterfactuals is the stative situation with concept exchange, as shown in 如果在平時... (*If it were as usual ...*) and 我要是你... (*If I were you ...*). The concept exchange, in an obvious violation of world knowledge, is a quite common semantic type of counterfactual in the world's languages. For example in Spanish, an older Peninsular formula *Yo que tú/usted* (*If I were you*) is still used in modern Spain, and another borrowed word from Catalan to express the role-play is also widespread: *Yo de ti/usted* (*If I were you*). There are also many examples which are uttered in obvious contradiction to our general world knowledge, like 假如有條侏羅紀的蛇頸龍爬行到了現代，大概也是這樣子 (LCMC) (*If a plesiosaurus from the Jurassic period were alive in modern times, it would be like this*). It is obvious that pragmatics have a great impact on the interpretation of a counterfactual in Mandarin. To quantify the importance of pragmatic influences on counterfactual reading, we make the following calculation (Table 19) of the sentences influenced by pragmatic inference.

**Table 19:** The number of sentences influenced by pragmatic inference across the three corpora

Contextual inference	Background knowledge	Total
20(8.2%)	54(22.0%)	74(30.2%)

## 4. Further findings

### 4.1 Patterns of counterfactual features: CFI-principle

Through analyses of the counterfactual features in Mandarin, several points are concluded as follows: (1) Mandarin does not have specialized linguistic markings for counterfactuals; (2) there are many features which are found to have a positive correlation with counterfactuality; however, they are not highly predictive of a counterfactual interpretation; (3) pragmatic inference seems to play a crucial role for counterfactual reading in Mandarin. Therefore, we may ask how Chinese people communicate counterfactual thoughts without any consistent linguistic marking. Feng & Yi (2006) and Wang (2012) argue that it is based on an interaction between counterfactual features and other variables such as semantics and contexts. What I shall present in the following passage is that the evaluation of the truth supporting a hypothetical utterance can in fact be taken as a measurable factor.

As analyzed earlier, the derivation of counterfactuality in Mandarin is related to the presence or absence of a number of features. Therefore, in the absence of dedicated counterfactual markers, the counterfactual features which are often associated with environments with a high information density function as the potential catalyst for inferring the counterfactual sense. The more features present, the greater the probability it will be understood as counterfactual. Therefore, the

continuum nature of Comrie (1986)'s description may correspond to variations in the number of features present in the sentences. The cluster of features together may contribute to an overall optimal situation for the interpretation of counterfactual notions, as in: 如果我們當時真的沒有愛過就好了 (*If we really had not loved at that time, it would have been good*). Here, the first person pronoun-*women* (我們), past tense indicator *dangshi* (當時), emphatic modal adverb *zhende* (真的), negation *meiyou* (沒有) and optative mood . . . *jiuhaole* (就好了) all work together to enhance a counterfactual reading, which therefore cannot easily be cancelled. However, we can barely find these idealized patterns in the actual use of Mandarin, where in most cases two or three features can ensure a counterfactual reading. To more accurately estimate the number of features which are expected to occur in counterfactual sentences, we make the following statistical analyses about the number of features and their corresponding frequencies. The features discussed so far are the following: a. Negation; b. Emphatic modal adverb; c. Optative mood; d. Past tense reference; e1. First person pronoun; e2. Demonstrative pronoun; f1. Contextual inference; f2. Background knowledge.

**Table 20:** The frequency of counterfactuals with a certain pattern of features<sup>6</sup>

f1	f2	d	a	e1	b+c	d+a	e1+d	e1+f2	a+a	e1+f1	b+f2	a+f2	e1+a	f1+c	a+b	e2+d	f2+d	a+f1	a+e2										
8	25	9	16	3	1	11	12	1	11	3	2	12	2	1	4	9	1	1	17										
e1+a+a		f1+c+e2		b+e2+a		e1+b+a		d+e1+e2		e1+f2+a		e1+d+f1		a+e2+d		a+a+d		e1+d+c		a+a+a		f2+d+c		f2+a+a					
4		1		3		5		2		2		1		3		17		1		3		1		2					
a+a+b			a+a+e2			f1+b+c			f1+e2+c			f2+e2+a			a+a+c			d+a+a+a			a+e2+e1+d			a+e2+a+e2			a+e2+a+e2+d		
10			19			3			2			8			2			2			2			2			1		

Table 20 presents us with different patterns of feature collocation which can be divided into five parts, namely sentences with one feature, two features, three features, four features and five features. The aggregate frequency of each category is given as follows (Table 21):

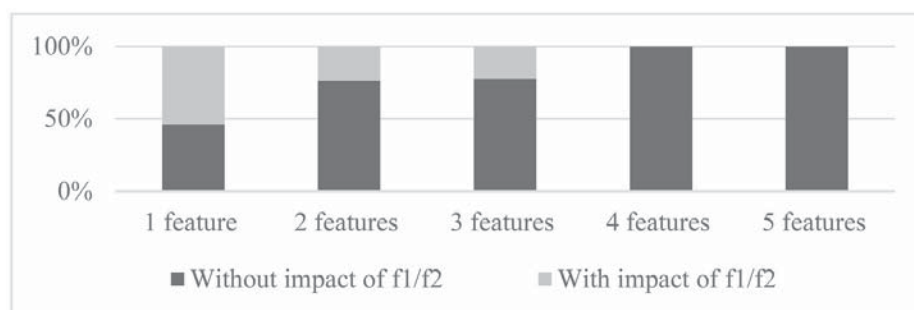
**Table 21:** The aggregate frequency of counterfactuals with a certain number of features

1 feature	2 features	3 features	4 features	5 features	Total
61	88	89	6	1	245

According to the data, only 2.9% of the counterfactuals have more than three features co-occurring simultaneously. The data of the present study show that the average number of features applied in the counterfactuals is 2.17, and the mode is 3, which means that most counterfactuals have

<sup>6</sup> The double or triple use of negators in one sentence is symbolized as 'a+a' or 'a+a+a', since they mutually complement the word meaning, for example, double negation is equal to affirmation. However, there are many cases with more than one past tense indicator in one sentence, and we use only one 'd' here because of the non-complementary meaning of the reduplicative uses of the indicators, for example, double uses of past indicators are not interpreted as pluperfect.

three features co-occurring together. Another finding concerns the changes of proportion of f1/f2 (contextual inference and background knowledge) with the increasing number of features (Figure 4):



**Figure 4:** Frequency of counterfactuals with and without pragmatic inference

Counterfactuals with only one feature are largely influenced by the context. As previously analyzed, emphatic modal adverbs or optative mood can never ensure a counterfactual reading for feature 1 sentences. At this stage, f1/f2 plays an essential role in 54.10% of the counterfactuals. With features increasing to two, the proportion of f1/f2 witnesses a dramatic decline to 23.60%, where most sentences can express counterfactuality independent of the pragmatic influences. Then a subsequent slight decrease of f1/f2 impacting strength to 22.73% of the pragmatic factors is found in counterfactuals with three features. For counterfactuals with four or five features, f1/f2 influence is negligibly minimal.

Therefore, we may conclude that a single grammatical or lexical device is available to indicate that the utterance is irrealis, but it is not a reliable indicator for disambiguating counterfactuals from open conditionals. World knowledge and extra-textual information may determine the factuality of the premise upon which a counterfactual hypothesis is constructed, especially when there aren't sufficient features in the sentences. It is also found that the more features present, the more contextually independent the counterfactual is. Ziegeler (2000:41) points out that the counterfactual implicatures, usually associated with environments with high information density, can be predicted by the presence of a certain number of features. She takes this phenomenon further in the form of a single principle labeled the Counterfactual Implicature (CFI) Principle:

**The CFI-Principle:**

‘The strength of an implicature is directly proportional to the specificity conditions in which it is located.’

The Principle is especially applicable to languages lacking dedicated counterfactual markings, where the strength of an implicature is positively proportional to the number of features in the sentence. Therefore the features working to increase the information specificity of an utterance in which the counterfactuality is strengthened may be better termed as counterfactuality enhancers. Different from markers, enhancers are neither necessary nor sufficient conditions for indicating counterfactuality. For languages with counterfactuality enhancers, it is possible to measurably predict the licensing of counterfactual inferences from the presence of certain patterns of features.

## 4.2 Contrastive analyses between counterfactuals in Mandarin and English

### 4.2.1 Differences in the aggregated frequency of counterfactuals among Mandarin, British English, and American English

The materials used for the contrastive study consist of three comparable balanced corpora: TORCH (Mandarin), CLOB (British English), and CROWN (American English). These three corpora are homogenous in terms of the variety of written registers, total number of words, and lexical variety (measured by Type-Token Ratio). Following the corpus-based contrastive inter-language analyses, the collocations of ‘if’ and POS. form, VBDR (were), VHD (had), VDD (did) are concordanced respectively. Also, the concordanced outputs are subsequently checked by manual analyses so as to prevent any noise entering into the data. All of these selected counterfactuals were calculated under each register. For each register, Mandarin shows an obviously lower number of uses of counterfactuals (as seen in Table 22).

**Table 22:** The register distribution of counterfactuals in TORCH, CLOB, and CROWN

	TORCH	CLOB	CROWN
A_reportage	4	17	15
B_editorials	0	10	17
C_reviews	0	7	9
D_religion	3	3	4
E_skills_hobbies	0	7	10
F_popular_lore	12	12	20
G_biography_memoirs	13	47	53
H_reports_gov_docs	0	5	5
J_academic_writing	3	15	6
K_general_fiction	22	24	18
L_mystery_detective	12	33	17
M_sci_fiction	2	16	3
N_adventure_fiction	18	21	14
P_romance_stories	4	14	40
R_humor	0	10	11

### 4.2.2 Mandarin translation of English counterfactual markers

The discussion in this section is based on the English–Chinese parallel corpus described in §2.1. As noted in the last section, counterfactuals are more frequently used in English than in Mandarin. In English, there are special schemes to mark counterfactuals like fake uses of past or pluperfect, conditional mood ‘would’ in the apodosis and so on. However, in Mandarin we have also discovered several features that are relevant for counterfactual expressions; nonetheless, they are not markers since they are not highly predictive of counterfactual utterance. A further question arises as to whether these features constitute a necessary part of translation for counterfactuals from English to Mandarin. To answer this question, I made a concordance of counterfactuals in the

corpus by doing a search with ‘if ... would’ as key words. Further manual work was still needed in order to avoid any isomorphs.

From the data, I find that counterfactuals were translated into Mandarin by applying different strategies, either by changing the past inflections into a general past temporal lexical item, as in the following:

- (4) He would have lost his position if you hadn’t backed him up.  
假如你當初 (‘at that time’) 沒有支持他，他可能已經失去職位了。

or by inserting adverbs like *zao* (早 ‘early’):

- (5) For Jane would have got Mr. Bingley, if she could.  
因為吉英要是能夠嫁給彬格萊先生，她早就嫁了。

or by applying sentence-final perfective aspect *le* (了):

- (6) If you had asked for directions, you wouldn’t have got lost.  
如果你問一下路，就不會走丟了。

or by combining the negator and conjunction into a compound conjunction:

- (7) **If not** so payable, they would have been authorized investments.  
若非如此付款則會是特准投資專案者。

or by introducing a left-periphery negator:

- (8) If it **weren’t** for the education by everyone and their efforts to save me ...  
如果不是大家的教育、挽救...

or by adding an emphatic modal adverb *-zhen* (真).

- (9) If I were to be a bird, I would fly to you at once.  
如果我真能變成一隻鳥，我就立即飛到你的身旁。

Some counterfactuals were translated simply by word-to-word mapping, but luckily, due to the assistance of context sensitive indexicals or pragmatic means, the counterfactuality is still preserved in the translated sentences:

- (10) But, **if I were you** ...  
要是我處在你的地位...

However, in most cases, counterfactuality was cancelled after being translated into Mandarin either by intentional transferring of negative counterfactuals to positive factuality:

- (11) If it were **not** for hope, the heart would break.  
人是靠希望活著的。(Literal meaning: Humans live on hope.)

or by unintentional translation without applying any counterfactual features in Mandarin.

- (12) This operation would be considerably faster if FindFast **were** installed.  
如果安裝了「檔檢索」，這項操作的速度就會顯著增快。

Since no correspondent subjunctive form could be used in Mandarin, the sentence is understood as an open conditional without any hint from the context.

## 5. Conclusions and some further thoughts

Our present research aims to discover the working mechanisms for counterfactuals in Mandarin and to compare the frequency and marking strategies between Mandarin and English. Following the statistical analyses, several key findings emerged. First, besides the contextual inference and background knowledge, Mandarin counterfactuals are found to be positively correlated with negation, emphatic modals, optative mood, past-oriented temporality, first person pronouns, and demonstratives. The degree of impact from contextual inference and background knowledge is subject to the variation of features present in the sentences. Second, the influence of hypothetical conjunctions on counterfactuals is revealed both in their different counterfactuality-generating abilities and their preferences in feature selection. Third, some significant differences have been found both in the frequencies of counterfactuals and in marking strategies between Mandarin and English.

The first finding of the study is the different features of counterfactuals; however, they are not working at the same level in indicating the counterfactuality. We have to make a distinction between, on the one hand, devices that a speaker uses to indicate or suggest that they intend to convey counterfactuality (irrealis operators), and on the other, various factors that make it more likely to the listener that a clause or sentence is false or known to be false (realis operators). Thus, the addition of a modal adverb (irrealis operator) can create a higher level of hypotheticality. But this is then different from the first person pronouns (realis operators) which involve the speaker and make it likely that the speaker has knowledge about its truth-value and thus make it more likely that it is in fact known to be false. Following this thought, all these features can be classified into two categories, namely irrealis operators and realis operators (Table 23).



**Table 23:** Classification of counterfactual features in Mandarin

Irrealis operators	Realis operators
Hypothetical conjunctions <sup>7</sup>	Past tense indicators
Emphatic modal adverbs	First person pronouns
Optative mood	Demonstrative pronouns
<i>bu</i> (不)	Pragmatic inference
	<i>meiyou</i> (沒有) <i>bushi</i> (不是)
<i>yaobushi</i> (要不是) <i>ruofei</i> (若非) hypothetical conjunction + <i>bushi</i> (不是)/ <i>meiyou</i> (沒有)	

*Bu* (不) is often employed to negate a non-past event, therefore it is an irrealis operator. However, *meiyou* (沒有) is often employed to negate an event which happened in the past, therefore it is a realis operator. The case of *bushi* (不是) is more complex. It is assumed that a factive operator of *bushi* (不是) (Su 2008) comes into effect when a counterfactual is uttered. Compare the following examples: 要是你有任何問題，你都可以來問我。(If you have any questions, you can come to ask me) and (\*) 要不是你有任何問題，你都可以來問我。(If you have any questions, you can come to ask me). Here, the negative polarity item *renhe* (任何 ‘any’) is licensed in the *yaoshi* (要是 ‘if’) sentence but not in the *yaobushi* (要不是 ‘if not’) sentence. According to Zwarts (1995) and Giannakidou (1998), it is assumed that a factive operator may work here, since it can block the negator *bu* (不 ‘not’) from licensing a negative polarity item as an upward-entailing operator.

An ideal counterfactual semantically must contain at least one irrealis operator and one realis operator. Therefore, the lexicalized *yaobushi* (要不是 ‘if not’) and *ruofei* (若非 ‘if not’), having both of the two components simultaneously, work as dedicated counterfactual markers in Mandarin. It should be noted that it seems that Mandarin and other languages share some common properties in applying past tense reference in counterfactuals; however, past tense works differently between Mandarin and other languages. In languages with a fake past, past tense is a dedicated counterfactual marker and its temporal reference is substituted by the modal use. Therefore, past tense can appear in a non-past environment to indicate a future counterfactual. However, in Mandarin, the temporal meaning of past tense indicators is maintained, and since the past termination of the event can indicate a greater potentiality for factual recall by the speaker, it contributes to the likelihood of a factual supporting premise on which to base a hypothetical expression. First person subjects who have first-hand information about the factual premise of the speaker’s utterance, demonstrative pronouns which presuppose the factual existence of the indicated event, and the contextual inference/extra-world knowledge about the facts supporting a hypothesis, all work together to enforce the factual basis. The interaction of formal features versus context and background knowledge (f1/f2) is shown in that if the features present are not forceful enough to provide clues to factuality, then f1/f2 is needed to increase the grounds for information specificity according to the CFI-Principle. On the other hand, if the features can provide an overall evaluation of factuality for the hearer, then f1/f2 is optional.

<sup>7</sup> Lexical stress on the hypothetical conjunction can enhance its role as a hypothetical component, thus leading to a stronger counterfactual reading.

Our second finding concerns the impact of a hypothetical conjunction on counterfactuals. According to our statistical analyses, we find that conjunctions like *jiaru* (假如) and *yaoshi* (要是) are more likely to evoke counterfactual interpretations than the others. It is hypothesized that a process of lexicalization may happen between *jia* (假 ‘fake’) and *ruguo* (如果 ‘if’). The evolution of conjunctions in Germanic languages, perhaps, can verify our hypothesis. According to Lehmann (2007), modifications were introduced to conjunctions for greater precision in Gothic, for example conjunctions and adverbs were added to the simple conjunction *ei* for different specific uses (like *fáurþizeī* ‘before’, *miþþanei* ‘while, as’, *sunsei* ‘as soon as’, *swāei* ‘so that’, *þadei* ‘where to’, *þandei* ‘because’, *þanei* ‘that’, *þarei* ‘where’, *þaprōei* ‘from where’, *þēei* ‘through that’). A similar modification process may happen on *jia* + *ruguo* to satisfy the specific needs to express events which are fake or highly hypothetical.

Our last finding reveals significant differences between counterfactuals in Mandarin and English. The frequency-based data reveal an obvious difference in frequency between counterfactuals in Mandarin and English. A subsequent translation analysis verifies the difference in that counterfactuals are more commonly used in English than in Mandarin. For example, some counterfactuals are intentionally translated into non-hypothetical narrations. Additionally, counterfactuality cancellation may occur after translation since there is no specialized inflection like subjunctive mood or past tense in Mandarin. In English, a factual component like past tense or perfect has been grammaticalized into a dedicated counterfactual marker, therefore it is possible that counterfactuality could be uttered without any factual base, like *If it were to rain, we would have to cancel the match tomorrow*. However, it is almost impossible to translate counterfactuals without any factual basis into Mandarin. Without any additional explanation, the translated sentence 如果下雨，我們只好取消明天的比賽 is only an open conditional. This is because English and Mandarin represent two different stages in the evolution of counterfactual markings. In English, since a dedicated counterfactual marker has been formed, counterfactuals could be expressed with an unmarked factual basis as with the above sentence. However, this is not the case in Mandarin, where counterfactuals are still restricted to sentences with a factual basis. Even for compound conjunctions like *yaobushi* (要不是 ‘if-not-be’) and *ruofei* (若非 ‘if not’), a factive operator works when a counterfactual is uttered: 要是你有任何問題，你都可以來問我 (*If you have any questions, you can come to ask me*). \*要不是/若非你有任何問題，你都不可以來問我 (*\*If it is not the case that you have any questions, you cannot come to ask me*). Here, the negative polarity item *renhe* (任何 ‘any’) is licensed in the *yaoshi* (要是 ‘if’) clause but not in the *yaobushi* (要不是 ‘if-not-be’) clause. According to Zwarts (1995) and Giannakidou (1998), a factive operator, which would block the negator from licensing a negative polarity item, may work in the *yaobushi* clauses.

## References

- Au, Terry Kit-Fong. 1983. Chinese and English counterfactuals: The Sapir-Whorf hypothesis revisited. *Cognition* 15(1–3):155–187.
- Barwise, Jon. 1986. Conditionals and conditional information. *On Conditionals*, ed. by Elizabeth Closs Traugott, Alice ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson, 21–54. Cambridge: Cambridge University Press.

- Bloom, Alfred H. 1981. *The Linguistic Shaping of Thought: A Study in the Impact of Language on Thinking in China and the West*. Hillsdale: Lawrence Erlbaum Associates.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chen, Guohua. 1988. Yinghan Jiashe Tiaojianju Bijiao [A comparison between English and Chinese hypothetical conditionals]. *Waiyu Jiaoxue yu Yanjiu [Foreign Language Teaching and Research]* 1:10–18.
- Comrie, Bernard. 1986. Conditionals: A typology. *On Conditionals*, ed. by Elizabeth Closs Traugott, Alice ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson, 77–100. Cambridge: Cambridge University Press.
- Feng, Gary, & Li Yi. 2006. What if Chinese had linguistic markers for counterfactual conditionals? Language and thought revisited. *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006)*, ed. by Ron Sun, 1281–1286. Mahwah: Lawrence Erlbaum Associates.
- Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)Veridical Dependency*. Amsterdam: John Benjamins.
- Huang, C.-T. James. 1987. Remarks on empty categories in Chinese. *Linguistic Inquiry* 18.2:321–337.
- Jiang, Yan. 2000. Hanyu Tiaojianju de Weishi Jieshi [Counterfactual interpretations of Chinese conditionals]. *Yufa Yanjiu he Tansuo Shi [Studies and Explorations on Syntax (Chinese) Vol. 10]*, 257–279. Beijing: The Commercial Press.
- Lehmann, Winfred P. 2007. *A Grammar of Proto-Germanic*, ed. by Jonathan Slocum. Austin: Linguistics Research Center, University of Texas at Austin. Available at <http://www.utexas.edu/cola/centers/lrc/books/pgmc00.html> (accessed 22 July 2014).
- Li, Charles N., & Sandra A. Thompson. 1990. Chinese. *The World's Major Languages*, ed. by Bernard Comrie, 811–833. New York & Oxford: Oxford University Press.
- Li, Jinxi. 1924[1992]. *Xin Zhu Guoyu Wenfa [New Grammar of Chinese]*. Beijing: The Commercial Press.
- Liu, Lisa Garbern. 1985. Reasoning counterfactually in Chinese: Are there any obstacles? *Cognition* 21:239–270.
- Lü, Shuxiang. 1942[1982]. *Zhongguo Wenfa Yaolue [Essentials of Chinese Grammar]*. Beijing: The Commercial Press.
- Lyons, John. 1977. *Semantics*, Vol. 2. Cambridge & New York: Cambridge University Press.
- Pan, Haihua, & Jianhua Hu. 2001. OT fang'an yu zhaoyingyu de yueshu [OT models and anaphoric binding]. *Wai Guo Yu [Journal of Foreign Languages]* 1:24–31.
- Sadock, Jerrold M. 1971. Queclaratives. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*, ed. by Douglas Adams, Mary Ann Campbell, Victor Cohen, Julie Lovins, Edward Maxwell, Carolyn Nygren & John Reighard, 223–232. Chicago: Chicago Linguistic Society.
- Sadock, Jerrold M. 1974. *Towards a Linguistic Theory of Speech Acts*. New York: Academic Press.
- Su, Julia Yu-Ying. 2008. Deriving counterfactuality in Mandarin Chinese *yaobushi* conditionals. Paper presented at the TOM (Toronto-Ottawa-Montreal) Semantic Workshop 1, March 29–30, 2008. Toronto: University of Toronto.

- Tien, Adrian. 1994. *Semantic Primitives in Taiwanese Chinese*. Canberra: Australian National University BA Hons thesis.
- Wang, Li. 1985. *Zhongguo Xiandai Yufa [Modern Chinese Grammar]*. Beijing: The Commercial Press.
- Wang, Yuying. 2012. *The Ingredients of Counterfactuality in Mandarin Chinese*. Hong Kong: Hong Kong Polytechnic University dissertation.
- Wierzbicka, Anna. 1997. Conditionals and counterfactuals: conceptual primitives and linguistic universals. *On Conditionals Again*, ed. by Angeliki Athanasiadou & René Dirven, 15–60. Amsterdam & Philadelphia: John Benjamins.
- Wu, Hsin-Feng. 1994. *'If Triangles were Circles ...': A Study of Counterfactuals in Chinese and English*. Taipei: Crane.
- Wu, Zhaoyi. 1989. *Exploring Counterfactuals in English and Chinese*. Amherst: University of Massachusetts Amherst dissertation.
- Yeh, David, & Dedre Gentner. 2005. Reasoning counterfactually in Chinese: picking up the pieces. *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society (CogSci 2005)*, ed. by Bruno G. Bara, Larry Barsalou & Monica Bucciarelli, 2410–2415. Mahwah: Lawrence Erlbaum Associates.
- Ziegeler, Debra. 2000. *Hypothetical Modality: Grammaticalisation in an L2 Dialect* (Studies in Language Companion Series). Philadelphia: John Benjamins.
- Zwarts, Frans. 1995. Nonveridical contexts. *Linguistic Analysis* 25:268–312.

[Received 5 September 2014; revised 7 January 2015; accepted 9 January 2015]

College of Foreign Studies  
Jinan University  
No. 601 West Huangpu Road  
Tianhe District, Guangzhou City  
Guangdong Province 516032  
China  
qyong2@jnu.edu.cn

## 基於語料庫的漢語違實句研究

雍茜

暨南大學

本文的研究是爲了揭示漢語違實句的工作機制。通過對所得語料進行統計分析可以得出以下結論：（一）除了語用因素，漢語違實句與諸多詞匯語法特徵關聯，如否定、情態副詞、祈願語氣、過去時制、第一人稱代詞、指示代詞等，語用干擾度受這些特徵出現數量的制約；（二）不同假設連詞具有不同的違實生成能力；（三）漢語與英語在違實句的使用頻率和標記策略上有明顯差異。本文的語料來源於六個語料庫，分別是：蘭卡斯特漢語語料庫、UCLA 漢語書面語語料庫、最新漢語文本語料庫、布朗家族語料庫（英式英語）、布朗家族語料庫（美式英語）和英漢雙語平行語料庫。

關鍵詞：CFI 原則，卡方檢驗，語料庫，違實，phi 相關性檢驗