RESEARCH ARTICLE

Conditionals in Mandarin: A Corpus-based Study from A Semantic Perspective

Bei Zhao
Department of French, University College Cork, Cork, Ireland
Corresponding Author: Bei Zhao, E-mail: 121100966@umail.ucc.ie

ABSTRACT
This corpus-based study delves into the expression of conditional constructions in Mandarin Chinese, employing a detailed analysis of both written and spoken data derived from four commonly used Mandarin corpora. Utilizing a mixed qualitative method, the research focuses on four widely-used conditional markers: rúguǒ, jiǎru, yàoshi, and yàobushì. The investigation aims to understand the syntactic function of these markers in initiating basic conditional statements, as well as to explore their varied use across different genres. The study reveals that while these markers share a common role in forming conditional sentences, they exhibit significant differences in terms of frequency of use and semantic nuances. This research contributes to a deeper understanding of Mandarin syntax and semantics, highlighting the complexity and diversity of language use in different communicative contexts. The findings offer insights into the intricate ways in which Mandarin speakers express conditionality, enhancing the comprehension of this aspect of the language.

KEYWORDS
Mandarin conditionals, Factual conditionals, Future (predictive) conditionals, Imaginative conditionals, Corpus Linguistics

ARTICLE INFORMATION
ACCEPTED: 12 December 2023 PUBLISHED: 01 January 2024 DOI: 10.32996/bjal.2024.4.1.1

1. Introduction
Research focused on conditional sentences has never ceased. Stalnaker (1968: 41) agreed that a conditional sentence encodes “a proposition which is a function of two other propositions, yet not one which is a truth function of those propositions”, expressing “the dependence of one set of circumstances on another” (Quirk & Greenbaum, 1973: 323). It contains two clauses, a subordinate clause and a main clause; the former represents the condition, always initiating with the if, so it can be called as the ‘if clause’ while the latter introduces the result or outcome.

Palmer (1978) proposed that in English and many other languages, there are two kinds of conditionals—real and unreal. The difference lies in whether the utterance contains the speakers’ individual opinions about the propositions that they present. If not, it is a real condition. In unreal conditionals, Palmer agreed that the propositions always imply the speakers’ doubts about the propositions. Diverse verb tenses and aspects in the main and subordinate clauses are used as appropriate. Quirk et al. (1985) divided the conditionals into direct and indirect conditionals, which laid the foundation for Dancygier’s (1998) classification of conditionals, including predictive, non-predictive, and generic conditionals. Eastwood (1994: 333) mentioned four categories of conditional sentences based on the degree of possibility implied by each conditional and its corresponding verb tense in each clause. That is, Zero Conditionals, Conditional Type 1, Conditional Type 2, and Conditional Type 3, which is one of the most prevalent taxonomies of conditional sentences. Larsen-Freeman and Celce-Murcia (2016: 575) once pointed out that conditional sentences are sometimes only divided into three types in the light of the time of the event: future conditionals, present conditionals, and past conditionals. Later, they classified the conditional sentences into nine types from a semantic perspective, which is the framework of the current research and is going to be illustrated in detail in Section 2.1.

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.
This corpus-based research generally aims to recognize and analyze the conditional constructions in Mandarin. The research questions are put forward as follows:

(a) Among the four Mandarin conditional markers (rúguǒ, jiǎrú, yàoshi, and yàobushì), which one is the most frequently used?
(b) What are the genre distinctions of using different markers?
(c) In what type of conditional sentences occur the most in written and spoken Mandarin corpora?

2. Categories of conditionals

2.1 Theoretical Framework

The theoretical framework used in this research follows Larsen-Freeman and Celce-Murcia’s (2015: 579) semantic descriptions regarding conditionals, consisting of factual conditionals, future (predictive) conditionals, and imaginative (subjunctive) conditionals, as shown in Figure 1.

As Larsen-Freeman and Celce-Murcia stated, conditional sentences can be semantically divided into three general types with a few subtypes. Examples that use four different subordinators have been given below in (1)-(9). The first factual conditional is sometimes called the zero conditional, characterized by the simple present tense in both the main clause and if-clause [If + simple present, subject + simple present]. The generic timeless factual refers to scientific facts such as common sense, natural laws, and general knowledge, as in (1). The habitual timeless factual is often used to express personal habits that people always know, as in (2). The time-bound factual sentences can convey implicit inferences or explicit inferences, as in (3) and (4). Sentences (5) and (6) represent future conditionals, known as the first conditional ([If + simple present/should/happen to, subject + will/be going to/may/might/should]), like future predictions, plans, promises, offers, and contingencies with strong or weakened condition or result. The explicit implication of future time is embodied in the result clause, while the use of modals such as may, might, and should can indicate the degree of weakness. The last type is the imaginative conditional that consists of hypothetical and counterfactual conditionals whose syntax structure is [If + were to/simple past/past perfect, subject + would/could/might/would + perfect aspect]. The former describes an event that is theoretically possible but has little chance of actually happening, and the latter expresses impossible events or comments about past circumstances. In the light of the tense of the if clause, either the simple past or the past perfect, the imaginative conditionals can be classified into the second or the third conditional.
(1) If you add sugar to hot water, the sugar dissolves.
(2) If I drink too much tea or coffee during the day, I cannot fall asleep at night.
(3) If it’s 40 degrees outside, the ice must have melted.
(4) If he was involved in that accident, he must have been scared.
(5) If it does not rain tomorrow, I will go on a picnic.
(6) If it rains tomorrow, I may not go out.
(7) If I were to have a long holiday next year, I would visit Europe. [Counterfactual Future]
(8) If I were you, I would buy that bag. [Counterfactual Present]
(9) If you had worked harder, you would have passed your exam. [Counterfactual Past]

2.2 Conditional Variants in Mandarin

In relation to English sentences, where the adverb *if* is typically employed to initiate the subordinate sentence, Mandarin similarly makes use of a few words, either positive or negative, to mark the prerequisite part of the whole conditional sentence; they are *rúguǒ*, *jiǎrú*, *yàoshì*, and *yàobushì*. They can be regarded as connectors whose parts of speech may be conjunctions, adverbs, and modal auxiliaries with a linking function. Unlike English clauses in which there is strict parallelism of tenses, aspects, and modals, Mandarin has relatively flexible grammatical rules because it is sometimes treated as a tenseless language in which “no obligatory inflection marking overt absolute tense” appears (Lin, 2012). Additionally, some Chinese linguists have approved that conditional sentences in Mandarin can be judged by the context rather than specific conditional markers (Xing, 2001; Lv, 2004; Wang, 2010). This case is not going to be discussed in the current research.

In Larsen-Freeman and Celce-Murcia’s (2016: 580) statements, there are four subtypes of factual conditionals. Among them, generic factual conditionals refer to some true and unchanging relationships which especially appear in scientific writing; habitual factual conditionals are literally related to personal habits rather than physical laws, which are usually habitually true. Two examples are listed to provide an illustration as below:

(10) *rúguǒ*  *értóng*  *shòudào*  *guòduōde*  *guāngxiàn*  *zhàoshè*  *tuìhēijīsù*  *de*  
    “If children are exposed to too much light, the production of melatonin will decrease.”

(11) *rúguǒ*  *cháng*  *shíjiān*  *chùyú*  *jǐnzhāng*  *fùxí*  *de*  *jiēduàn*  
    “If examinees are in the stage of intense review for a long time, it will make them even more nervous.”

Distinguishable from the previous two timeless factual conditionals, the inference conditionals, implicit and explicit, are time-bound or location-bound, expressed in the *if* clause, exemplified by sentences (12) and (13), respectively. Research related to the inference conditionals is quite limited, and the verb tense and aspect topic mentioned by Larsen-Freeman and Celce-Murcia (2016: 580-581) has little to do with the current study, which is not going to be analyzed in detail.

(12) *qiánmiànde*  *wēijī*  *hái*  *méiyǒu*  *guòqù*  *rúguǒ*  *zài*  *yīnwéi*  
    “If the previous crisis still passed again because of the裁员闹出新的事故，自己当何以自处？”

---

All the English sentences are researcher own and all the Mandarin sentences are extracted from the corpora.
layoff make new accident I should what do
“The previous crisis has not passed; if a new accident because of layoffs, what should I do?”

(13) 如果 您 能 科学地 教育 孩子， 您的 孩子
rúguǒ nín néng kēxué dì jiàoyù hái zì ninde hái zì
if you can scientifically educate children your children
一定会 回报 您的 爱。
yídìng huíbào nínde ài
must will return your love
“If you can educate your children scientifically, they must return your love.”

Sentence (12) presents an implicit inference conditional, showing a specific time-bound relationship through the “previous crisis” and “new accident”. Sentence (13) is an example of explicit inference conditional with “an inferential modal, typically must or should in the main clause” (Larsen-Freeman and Celce-Murcia, 2016: 581); here is a yídìng (must).

As the name of ‘future (predictive) conditionals’ shows, this type of conditional sentences encodes future events which have not happened yet. The strong or weakened conditionals mark different degrees of conditions or results. Compared to the strong type, the weakened ones tend to use “weaker modal predictions such as may, might, or should” (Larsen-Freeman & Celce-Murcia, 2016: 581).

(14) 如果 不 注意 健康 饮食 的话 你的 身材 同样 会 走形。
rúguǒ bù zhùyì jiànkāng yǐnshí dehuà nǐde shēncái tóngyàng huì zǒuxíng
if NEG attention healthy diet PRT your figure also will out of shape
“You will also get out of shape if you don’t pay attention to healthy diet.”

(15) 如果 米 特别的 白， 就 可能 是 毒 大米。
rúguǒ mǐ tèbié de bái jiù kěnéng shì dù dàmǐ
if rice particularly white PRT may be toxic rice
“If the rice is particularly white, it may be toxic rice.”

Jiang (2019) defined that “counterfactual conditionals (hereafter CF) are conditional sentences containing antecedents which are known to be contrary to fact or impossible to be true”. Karawani (2014: 5) agreed that world languages universally have the ability to convey the CF meaning by using “a dedicated or specialized CF morpheme” or “combinations of particular morphemes”, and Mandarin is no exception. In Mandarin, a few special lexicalized chunks are often employed to lead CF contexts. Furthermore, Jiang (2019) adopted Comrie’s (1985) opinion and pointed out that Mandarin makes use of relative tense to “indicate prior or posterior temporal relationships”, which can be marked by aspect markers such as the verb suffix le to demonstrate the completion of the action. The use of le can both indicate the hypothetical event in the present time and mark the accomplishment of an event, namely, the past time. Other markers, including temporal references like zǎo (early), negators like yàoburán (had it not been the case), and complements like …dehuà (in the case of…), also work.

(16) 如果 我是 外国人， 能 用 外语 给我指 路吗?
rúguǒ wǒ shì wàiguórén néng yòng wài yǔ géi wǒ zhǐ lù ma?
if I BE foreigner can use foreign language for me show way ²PRT

２Lists of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>classifier</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PERF</td>
<td>perfect tense</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>PRT</td>
<td>particle</td>
</tr>
<tr>
<td>ASP</td>
<td>aspect marker</td>
</tr>
</tbody>
</table>
“If I were a foreigner, could you show me the way in a foreign language?”

The *rúguò* in sentence (16) leads to an impossible CF meaning because the property of the subject in the *if* clause can never be changed. Similarly, a situation contrary to the facts of the past or future can also be conveyed by using different chunks in sentences (17) and (18).

(17) 一宁空道：你虽然救了我，但是却杀了这些人...
 Ningkong say you although save ASP me but but kill ASP these many people

一白衣人道：笑话，这些强盗刚刚在关口杀了几百个兵丁，要是让他们冲到城里，更不

báiyīrén dào: xiàohuà, zhèxiē qiángdào gānggāng zài guānkǒu shā le
Baiyi say joke these bandit just at pass kill ASP

hundred CL soldier if let them rush to city more NEG

zhī yào sǐ jǐqiān jǐwàn rén!
know will die thousand ten thousand people

“一宁空说：虽然你救了我，但是却杀了这些人...

一宁空说：Although you saved me, but you killed so many people...

一 白衣人说：笑 下， 这些 强盗 刚刚 在 关口 杀了 几百 个 兵丁， 要是 让 他们 冲 到 城里， 更 不

jiān rú yàoshì  tián le  róng guó  shā le  jǐbǎi  gè  bīngdīng,  yào shì  ràng  tāmen  chōnɡ dào  chénɡlǐ,  gènɡ  bù
hundred CL soldier if let them rush to city more NEG

zhī yào sǐ jǐqiān jǐwàn rén!
know will die thousand ten thousand people

“一宁空：What a Joke! These bandits have just killed hundreds of soldiers at the pass. If they had rushed to the city, thousands of people would have died!”

(18) 印尼有 17000 个岛屿，假如海平面上升一米，印尼有 CL island if sea level arise one meter

jiāng yǒu 2000 个岛屿会消失

jiāng yǒu CL island will disappear

“Indonesia has 17,000 islands, and 2,000 of them would disappear if sea levels were to arise by one meter.”

The *if* clauses in the two sentences above, beginning with *yàoshì* and *jiān rú*, describe imaginative contexts that oppose the facts of the past and the future, respectively. In sentence (11), through the indication of the aspect marker *le*, it is made known that the bandits have actually been killed before the conversation happens, making the assumption that “if the bandits rush to the city,” it is impossible to take place. According to Larsen-Freeman and Celce-Murcia (2015: 585), imaginative conditional sentences contrary to the future time often express “unrealized or unlikely yet theoretically possible event”. The event in sentence (12) is this case because, despite the fact that the sea level has not risen while the islands have not disappeared, the hypothetical circumstance is possible to happen in the future.

3. Methodology

3.1 Data collection

The current study presents a few possible markers of conditionals in Mandarin, including *rúguò, jiān rú, yào shì, yào bushi*, and *zāozhídào* based on four Mandarin corpora—Lancaster Corpus of Mandarin Chinese version 2, BFSU ToRCH family Chinese corpora, the UCLA Corpus of Written Chinese (2nd edition), and the BFSU DiSCUSS Corpus.

The LCMC2 is an update of LCMC version 1, consisting of the same text samples as LCMC1, which was created by Richard Xiao and represented “written Chinese published in China in the early 1990s” (McEnery & Xiao, 2005). A few typo and segmentation corrections were made in the LCMC2. The corpus contains a massive amount of data in written and spoken discourse belonging

4 Hereafter referred to as ToRCH. [http://114.251.154.212/cqp/torchfamily/](http://114.251.154.212/cqp/torchfamily/)
to 15 hybrid genres, from reportage to fiction, which is identical to the types covered in the ToRCH and UCLA. The former is the acronym of ‘Texts of Recent Chinese Texts’, belonging to the Brown corpus family. The ToRCH corpus used in the current research is a combination of three previous versions—ToRCH 2009, 2014 and 2019. The data in the UCLA 2 were all “collected from written modern Chinese available from the internet”, whose genres were as consistent as possible with the Brown corpus model (Xiao, 2013). The UCLA 2 is of great significance in complementing the LCMC 2 in terms of considering the influence of the Internet on Mandarin. The DiSCUSS corpus presents diverse and balanced materials of spoken Chinese uttered in social contexts. In order to better analyze the conditionals in Mandarin, the DiSCUSS is applied here to supply the oral data, making a contrast between the use of conditionals in spoken and written Mandarin. General information about each corpus mentioned above is provided in Table 1 below.

<table>
<thead>
<tr>
<th>Corpora</th>
<th>LCMC 2</th>
<th>ToRCH</th>
<th>UCLA 2</th>
<th>DiSCUSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text types</td>
<td>Written</td>
<td>Written</td>
<td>Written</td>
<td>Spoken</td>
</tr>
<tr>
<td>Total words in all corpus texts</td>
<td>1,021,037</td>
<td>3,694,301</td>
<td>1,097,113</td>
<td>1,268,835</td>
</tr>
<tr>
<td>Total number of corpus texts</td>
<td>500</td>
<td>1,172</td>
<td>15</td>
<td>300</td>
</tr>
</tbody>
</table>
| Text genres
  - Press: reportage
  - Press: editorials
  - Press: reviews
  - Religion
  - Skills, trades, and hobbies
  - Popular lore
  - Biographies and essays
  - Miscellaneous: reports and official documents
  - Science: academic prose
  - General fiction
  - Mystery and detective fiction
  - Science fiction
  - Adventure and martial arts fiction
  - Romantic fiction
  - Humor
| Dialogue:
  - Direct Conversations
  - Telephone Calls
  - Classroom Lessons
  - Broadcast Discussions
  - Broadcast Interviews
  - Parliamentary Debates
  - Legal Cross-Examinations
  - Business Transactions
  - Monologue:
    - Spontaneous Commentaries
    - Unscripted Speeches
    - Demonstrations
    - Legal Presentations
    - Broadcast News
    - Broadcast Talks
    - Speeches (Not Broadcast)

Table 1 General information about LCMC2, ToRCH, UCLA2, and DISCUSS

3.2 Data Analysis
An analysis of the Mandarin conditional markers rúguǒ, jiārú, yàoshì, yàobushì, and zǎozhīdào proceeded in three steps. First, whole sentences containing the tokens of these markers were all extracted from the four corpora and categorized into a few lists in light of the types of conditionals and genres of texts in Mandarin. Second, all the sentences were interpreted to analyze the discourse function of each conditional marker. If the sentence where the marker is too short or ambiguous to convey the meaning, the context would be checked.

4. Results
According to Table 2 below, there are a total of 4974 tokens of rúguǒ, 212 of jiārú, 527 of yàoshì, and 37 of yàobushì identified in the four corpora. It is obvious that rúguǒ occurs much more frequently than the other three markers in each corpus, and yàobushì as the only marker that contains the negation meaning, is the least commonly used. The trend of using different markers is found to be similar in the written data and the spoken ones that rúguǒ significantly takes priority, and yàobushì is rarely employed. A small difference lies in the use of the other two markers—jiārú and yàoshì. Compared to the written data, the frequencies of using them are not noticeably contrasting; however, yàoshì are almost five times more common than jiārú in the spoken data.
Table 2 The occurrence of conditional in the four corpora

<table>
<thead>
<tr>
<th>markers occurrence</th>
<th>rúguǒ Freq %</th>
<th>jiārú Freq %</th>
<th>yàoshi Freq %</th>
<th>yàobushi Freq %</th>
<th>Total Freq %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCMC2</td>
<td>601 88.5</td>
<td>38 5.6</td>
<td>38 5.6</td>
<td>2 0.3</td>
<td>679 100</td>
</tr>
<tr>
<td>ToRCH</td>
<td>2172 88.8</td>
<td>82 3.4</td>
<td>167 6.8</td>
<td>24 1.0</td>
<td>2445 100</td>
</tr>
<tr>
<td>UCLA2</td>
<td>867 92.5</td>
<td>39 4.2</td>
<td>28 3.0</td>
<td>3 0.3</td>
<td>937 100</td>
</tr>
<tr>
<td>Spoken</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISCUSS</td>
<td>1334 79.0</td>
<td>53 3.1</td>
<td>294 17.4</td>
<td>8 0.5</td>
<td>1689 100</td>
</tr>
<tr>
<td>Total</td>
<td>4974 212</td>
<td>527 37</td>
<td></td>
<td></td>
<td>5750</td>
</tr>
</tbody>
</table>

The following figures show the use of the four conditional markers. Figure 2 roughly shows the distribution of each conditional marker, and Figure 3 demonstrates the number of occurrences for each conditional marker and their proportions in each corpus.

Figure 2 The Distribution of conditional markers

Figure 3 Frequency percentages of conditional markers in different corpora
From Figures 2 and 3 above, it is obvious that *rúguǒ* appears the most often in the total number, and it also has the largest number of occurrences in each corpus, which is principally taken as a prototypical conditional marker, occupying 86.5% in all. *yàoshì* is the second most frequently-used conditional marker, even though its proportion is far less than that of *rúguǒ*, accounting for 9.2%, followed by *jiārú* and *yàobushi*. However, the use of the latter can almost be ignored, only occupying 0.6% in total. Additionally, according to the data of the DISCUSS, it seems that people tend to use *yàoshì* more often in spoken utterances than in written contexts.

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Markers</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
<th>Type 6</th>
<th>Type 7</th>
<th>Type 8</th>
<th>Type 9</th>
<th>Null</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCMC2</td>
<td><em>rúguǒ</em></td>
<td>86</td>
<td>48</td>
<td>13</td>
<td>32</td>
<td>283</td>
<td>39</td>
<td>29</td>
<td>41</td>
<td>29</td>
<td>1</td>
<td>601</td>
</tr>
<tr>
<td></td>
<td><em>jiārú</em></td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td><em>yàoshì</em></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td><em>yàobushi</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ToRCH</td>
<td><em>rúguǒ</em></td>
<td>332</td>
<td>84</td>
<td>62</td>
<td>65</td>
<td>930</td>
<td>224</td>
<td>120</td>
<td>182</td>
<td>154</td>
<td>19</td>
<td>2172</td>
</tr>
<tr>
<td></td>
<td><em>jiārú</em></td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>28</td>
<td>10</td>
<td>7</td>
<td>19</td>
<td>3</td>
<td>5</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td><em>yàoshì</em></td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>74</td>
<td>19</td>
<td>11</td>
<td>28</td>
<td>11</td>
<td>8</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td><em>yàobushi</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>UCLA2</td>
<td><em>rúguǒ</em></td>
<td>121</td>
<td>45</td>
<td>13</td>
<td>17</td>
<td>387</td>
<td>84</td>
<td>79</td>
<td>69</td>
<td>46</td>
<td>6</td>
<td>867</td>
</tr>
<tr>
<td></td>
<td><em>jiārú</em></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td>5</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td><em>yàoshì</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td><em>yàobushi</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>DISCUSS</td>
<td><em>rúguǒ</em></td>
<td>27</td>
<td>9</td>
<td>20</td>
<td>29</td>
<td>908</td>
<td>113</td>
<td>39</td>
<td>118</td>
<td>60</td>
<td>11</td>
<td>1334</td>
</tr>
<tr>
<td></td>
<td><em>jiārú</em></td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>19</td>
<td>10</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td><em>yàoshì</em></td>
<td>8</td>
<td>13</td>
<td>1</td>
<td>15</td>
<td>152</td>
<td>52</td>
<td>7</td>
<td>25</td>
<td>19</td>
<td>2</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td><em>yàobushi</em></td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>586</td>
<td>207</td>
<td>120</td>
<td>180</td>
<td>2835</td>
<td>568</td>
<td>310</td>
<td>521</td>
<td>354</td>
<td>69</td>
<td>5750</td>
</tr>
</tbody>
</table>

Table 3 The occurrence of conditional markers in the four corpora

From Table 3, it is distinct that, in general, most conditional sentences, almost half of them in total, are related to Type 5, namely, the strong condition and result, and the number of sentences in Type 1, 6, and 8 are quite similar. Except for the markers contained in conditional sentences, there are irrelevant cases in corpora, such as repetition, polysemy, and wrong combination, which are categorized into the Type Null. It is excluded in Figure 4.

<table>
<thead>
<tr>
<th>Markers</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
<th>Type 6</th>
<th>Type 7</th>
<th>Type 8</th>
<th>Type 9</th>
<th>Null</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>rúguǒ</em></td>
<td>566</td>
<td>186</td>
<td>108</td>
<td>143</td>
<td>2508</td>
<td>460</td>
<td>267</td>
<td>410</td>
<td>289</td>
<td>37</td>
<td>4974</td>
</tr>
<tr>
<td><em>jiārú</em></td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>65</td>
<td>27</td>
<td>20</td>
<td>43</td>
<td>14</td>
<td>17</td>
<td>212</td>
</tr>
<tr>
<td><em>yàoshì</em></td>
<td>10</td>
<td>17</td>
<td>7</td>
<td>23</td>
<td>257</td>
<td>79</td>
<td>23</td>
<td>64</td>
<td>34</td>
<td>13</td>
<td>527</td>
</tr>
<tr>
<td><em>yàobushi</em></td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>17</td>
<td>2</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>586</td>
<td>207</td>
<td>120</td>
<td>180</td>
<td>2835</td>
<td>568</td>
<td>310</td>
<td>521</td>
<td>354</td>
<td>69</td>
<td>5750</td>
</tr>
</tbody>
</table>

Table 4 The summary of conditional markers of different types in the corpora
Figure 4 is the diagramization of Table 4, showing the general distribution of each conditional marker in the four corpora. It is super obvious that all four markers appear the most when conveying the meaning (5), namely, the strong condition or result. _rúguǒ_, as the most popular conditional marker, is most frequently-used in all types of sentences. For example,

(19) 目前遗址还没有遭到大规模的破坏，如果
múqián yízhǐ hái méiyǒu zāodào dàguīmóde pōhuài rúguǒ
so far site still NEG get extensively damage if
发掘，肯定会有重大的发现。
fājué kěndìng huìyǒu zhòngdàde fāxiàn
excavate for sure yield significant finding
“So far, the site has not been extensively damaged; if it is excavated, it will yield significant findings for sure.”

The sentence (19) is a typical representation of Type 5 because it points out the affirmative prediction (significant findings will be yielded) under certain circumstances (if the site could be excavated). _rúguǒ_ also greatly works in the similar but weakened condition of Type 5, namely 6 and Type 1, showing general truth as in sentences (20) and (21).

(20) 可如果形成了许多人的争购的态势，
kě rúguǒ xíngchéng le xǔduō rén zhēnggòu de tài shì
but if form PRT a lot of people buy PRT situation
就可能出现价格高于价值的情况。
jiù kěnéng chūxiàn jiàgé gāoyú jiàzhí de qíngkuàng
PRT may appear price high value PRT situation
“But if a lot of people buy (something), there may be a situation where the price is higher than the value.”

Sentence (20) is an example of Type 6, stating a possible prediction with a not that strong attitude, while sentence (21) illustrates Type 1, a scientific fact that everyone knows is true and will change.

(21) 如果生长激素降低，胰岛素作用相对占
rúguǒ shēngzhǎng jīsù jiàngdī yídǎosù zuòyòng xiàngduì zhàn
if growth hormone reduce insulin effect relatively BE
优势，可以使脂肪合成增多造成肥胖。
yōushì kě shǐ zhīfáng héchéng zēngduō zàochéng féipáng
priority so make fat synthesis increase result in obesity
“If the growth hormone is reduced, the effect of insulin is relatively dominant, making fat synthesis increase and resulting in obesity.”

Besides ṛúɡuò, the marker yàoshi is often used in Type 5, followed by Type 8 and 7. Sentences (22) and (23) exemplify a strong plan (going to the aquarium) when a particular condition (it rains) is applied.

(22) 要是 下雨 咱 就 去 那个 海洋馆。
yàoshi xiàyǔ zán jiù qù nà gè hǎiyángguǎn
if rain we PRT go that aquarium
“If it rains, we will go to that aquarium.”

For other markers, jiǎrú is also used in Type 8, showing impossible CF circumstance, which is opposite to the current fact as in sentence (23). It is a completely impossible CF conditional sentence because nobody can turn back the clock. yàobushì, as the least frequently-used, appears relatively more when it is related to Type 9, aiming to convey a CF situation in the past and its probable result in the past, as in sentence (24).

(23) 假如 时光 能 倒流!
jiǎrú shíguāng néng dǎoliú
if time can turn back
“If I could turn back the clock!”

(24) 要不是 邻居 小孩 发现 告诉 他 妈, 我 真是 死 了。
yàobushì línjū xiǎohái fāxiàn gàosù tā má wǒ zhēnshì sǐ le
if+NEG neighbor kid find out tell his mother I really die PRT
“I would have died if the neighbor kid hadn’t found out and told his mother.”

Sentence (24) describes a CF condition different from the past fact. The subject “I” in the main clause did not die because the neighbor kid noticed something had happened and saved “my” life. Therefore, I could be alive to say this CF conditional sentence to explain the past situation.

In terms of genre distinctions, each conditional marker has a different performance, which has been demonstrated in Table 5 and its corresponding Figure 5.

<table>
<thead>
<tr>
<th>Categories</th>
<th>ṛúɡuò</th>
<th>jiǎrú</th>
<th>yàoshi</th>
<th>yàobushì</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: News reportage</td>
<td>195</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>B: News editorial</td>
<td>237</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>C: News review</td>
<td>141</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>D: Religious</td>
<td>101</td>
<td>15</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>E: Skill/trade/hobby</td>
<td>530</td>
<td>12</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>F: Popular lore</td>
<td>458</td>
<td>21</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>G: Biography and essay</td>
<td>429</td>
<td>27</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>H: Miscellaneous (official document, report, etc.)</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>J: Academic prose</td>
<td>587</td>
<td>13</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>K: General fiction</td>
<td>181</td>
<td>17</td>
<td>41</td>
<td>3</td>
</tr>
<tr>
<td>L: Mystery and detective fiction</td>
<td>235</td>
<td>14</td>
<td>33</td>
<td>5</td>
</tr>
<tr>
<td>M: Science fiction</td>
<td>98</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>N: Adventure and martial arts fiction</td>
<td>179</td>
<td>8</td>
<td>21</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 5 Distribution of markers in different genres

<table>
<thead>
<tr>
<th>Types</th>
<th>rúguǒ</th>
<th>jiārú</th>
<th>yàoshi</th>
<th>yàobushi</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialogue</td>
<td>826</td>
<td>44</td>
<td>254</td>
<td>8</td>
<td>1132</td>
</tr>
<tr>
<td>Monologue</td>
<td>508</td>
<td>9</td>
<td>40</td>
<td>0</td>
<td>557</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1334</td>
<td>53</td>
<td>294</td>
<td>8</td>
<td>1689</td>
</tr>
</tbody>
</table>

Table 6 Distribution of markers in dialogue and monologue spoken data

As for spoken data, rúguǒ is undoubtedly employed the most, occupying more than four times the number of the second one, yàoshi. Compared to the monologue, the interlocutors who have dialogues are inclined to use the conditional markers more often. However, no matter the dialogue or the monologue, the occurrence frequency of rúguǒ, yàoshi, jiārú, and yàobushi decreases in turn.

5. Conclusion
In summary, the present study concentrates on conditional sentences in Mandarin from a semantic perspective. Among the four Mandarin conditional markers (rúguǒ, jiārú, yàoshi, and yàobushi), rúguǒ is dominant and occurs the most, followed by yàoshi, jiārú, and yàobushi. Of the 15 genres of written texts under consideration, rúguǒ takes full priority in each genre compared to the other three markers, and it is the most commonly-used in the type of ‘academic prose’. Authors of general fiction tend to use yàoshi more than those of other genres, while it is likely for people to choose jiārú in the type of ‘biography and essay’. yàobushi, as the least popular marker, is rarely used in written and spoken texts. When it comes to the types in which conditional sentences...
occur the most in written and spoken texts, the answer is the condition expressing strong conditions and results such as predictions, plans, and contingencies. Besides, the meanings consisting of scientific facts, less certain preconditions or plans, and counterfactual connotations opposed to the current fact are not uncommon. In terms of spoken Mandarin, conditional markers tend to be used in conversations instead of monologues, and the use of 如果, 如果, if, and 会 reduces in turn.

The current research, focusing mainly on four Mandarin conditional markers, may not entirely capture the diversity of conditional constructions in Mandarin. Future studies should consider additional markers and regional variants, including dialects and habitual uses, to broaden the understanding of Mandarin conditionals across different dialects and sociolects. The methodological approach, primarily corpus-based, might not reflect the latest language use trends, as it depends on existing written and spoken data. Future research could benefit from incorporating experimental and ethnographic methods, offering a more dynamic view of language use, especially in informal and evolving linguistic settings.

Further research avenues include comparative studies between Mandarin conditionals and those in other languages, which could illuminate unique aspects, similarities, and differences influenced by cultural and linguistic contexts. Additionally, exploring practical applications in areas like language teaching and translation could enhance Mandarin teaching methodologies and improve machine translation systems for better handling of Mandarin conditional sentences.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

**ORCID:** 0000-0002-1233-3166

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

**References**


