

## Discursive Use of *ma* (吗) in Chinese News Commentaries

**Narisong**

*School of International Studies, Hangzhou Normal University, Hangzhou, China*

Email: narisujin@163.com

**Ming Yue**

*School of International Studies, Zhejiang University, Hangzhou, China*

Email: dr.yueming@gmail.com

**[Abstract]** The discursive usage of *ma* (吗) as an important interrogative particle in Mandarin Chinese has been neglected. This paper does a quantitative investigation on the *ma*-sentences in a Chinese news commentary corpus, and finds strong correlations between their semantic/pragmatic meanings and text/paragraph locations. As *ma* is nearly always followed by a question mark in written Chinese, its strong questioning connotation always elicits an expectation for answers/solutions. Thus, the Question-Answer (or Problem-Solution) structure is frequently employed as a key device for the intended argumentative purposes of the commentary writer.

**[Keywords]** mandarin Chinese, interrogative particle, *ma*, discourse corpus, rhetorical structure theory

### Introduction

*Ma* (吗) as an important Chinese interrogative particle has been discussed in the literature on its origin, phonological changes, syntactic functions, and semantic meanings. However, most discussions focus on isolated sentences, many of which base on the linguist's intuitive examples. The few empirical studies on the discursive use of *ma* have superficially investigated its frequency distributions in daily conversations, literary works, popular science articles (Lan, 2013), without paying adequate attention to the role *ma*-sentences play in natural written monologue.

This paper therefore intends to make use of a discourse corpus to investigate the discursive use of *ma*-sentences in news commentaries. The following three research questions are asked:

- (a) How are *ma*-sentences distributed in Chinese news commentaries?
- (b) What is the typical usage of *ma*-sentences?
- (c) Is there any statistical correlation between the locus of *ma*-sentences and their discursive uses?

Through detailed semantic/pragmatic annotation and statistical analysis, we find that the *ma*-sentences often appear in titles as well as the beginnings and endings of paragraphs in Chinese news commentaries. They usually ask interrogative questions in the titles and paragraph openings to prompt answers in the following text, but ask a series of rhetorical questions to bring up opposition or call up readers' retrospections at paragraph endings. There are statistical correlations between discursive locations of both *ma*-sentences and their immediate upper constructions to their rhetorical usages. The following three sections report the methodology, results, and discussion of this study. A short conclusion is offered at the end of the paper.

## Literature Review

### *Previous Study on Ma-Sentences*

Previous studies on *ma*-sentences differentiated question types in Chinese with three major criteria: formal, i.e., whether the question has characteristic words or structures; semantic, i.e., whether the question has certain transformational correspondence with a declarative sentence, or whether the question is true or fake; and pragmatic, i.e., how certain the questioner is about the issue questioned. Such criteria help us understand the multiple characteristics of *ma*-questions, but are not truly concerned about their discursive usage, i.e. the roles they play for the communicative purposes of the speaker/writer.

### *Discourse relations and RST*

Rhetorical Structure Theory (Mann & Thompson, 1988) views natural texts as roughly trees and captures textual structures in terms of how one span of text supports another for the writer's communicative purposes in a set of rhetorical relations (RRs). Using schematic representations, RST is flexible with regard to the size of the text spans being related to one another: RR's can be applied between spans within a clause, between clauses, or between lengthy spans of text made up of multiple sentences and even paragraphs up to the whole text. Over the years, organized efforts have also been made to collect consistent RST analyses for statistical investigation. There are English, German, Spanish, Portuguese and Chinese discourse corpora annotated fairly homogeneously with rhetorical relations (Carlson et al, 2003; Stede, 2004; Taboada & Mann, 2006; Yue, 2008). These annotated discourse corpora have assisted the study of discourse markers and typical communicative schemas (Ford, 1994, 2000; Spenader & Lobanova, 2009; etc.) in theoretical linguistic studies as well language engineering applications.

## Methodology

### *Corpus Data*

We use the 400 texts in the Caijingpinglun Corpus (Yue, 2008), which will hereafter be referred to as CJPL400. The CJPL400 corpus has over 750,000 word tokens, and has rich linguistic information useful in this study on *ma*: the texts are representative of Chinese news commentaries; they have been cleaned, tokenized and POS-tagged for further processing; and they have been annotated with rhetorical structure trees in 50 Chinese rhetorical relations of 10 clusters (namely BACKGROUND, CAUSATION, CONDITION, CONJUNCTION, CONTRAST, DISJUNCTION, ELABORATION, JUSTIFICATION, MEANS, TOPIC-TRANSITION, and OTHERS). In the SOLUTIONHOOD cluster in particular, there are Question-Answer and Problem-Solution relations.

### *Annotation*

Systemic annotation is necessary so that linguistic analysis could be made explicit, openly examined and carefully replicated. A lot of work remains to be done before we could actually carry out a quantitative study on the discursive usage of *ma*. In this project, data are retrieved and further processed in the following steps:

- First, retrieve all the sentences with a *ma* from CJPL400, and assign a unique identity for each and every token of *ma*. Sentences are defined as strings between selected punctuation marks (the usual period, question mark, exclamation mark, and paragraph-endings, as well as colon and semicolon taken in CJPL as Elementary Discourse Unit delimiters), and will thereafter be used as equivalents to EDUs.
- Retrieve and transform the positional information of all the *ma*-EDUs: paragraph id (title, first-p, mid-p, last-p), sentence/EDU id (sole, first-s, mid-s, last-s).
- Retrieve all the EDUs with a *ma* from CRST, together with their rhetorical relation tags (name, nuclearity and spans) from CJPL corpus, and transform the data to an excel file, under the variables RR, RRnuc, and RRspan.
- Retrieve all the Immediate Upper Discourse Units with *ma* (thereafter IUDUs of *ma*), together with their rhetorical relation tags (name, nuclearity, and span range) from the CRST corpus, and transform the data to the excel file, under the variables RA, RAnuc, and RAspan; The values of the six retrieved variables in a discourse (sub)tree is illustrated in Figure 1 for Example 1:
- Group all the RRs and RAs into larger semantic clusters. Since most RRs and RAs belong to the ELABORATION group, they are further divided into two sub groups: ELABORATION with Elaboration, Restatement, and Evaluation relations; and SOLUTIONHOOD with Question-Answer and Problem-Solution relations.
- Annotate all the *ma*-sentences for the following variables: sentence type (omitted, simple, complex); question type (not-a-Q, suspicion/doubt, interrogation, strong negation), answer type (no-answer, implied, delayed, immediate).
- For a few *ma*-s used within an EDU (i.e. between commas, or within parentheses), add their corresponding RRs in the excel file.

Example1 <sup>3</sup> <duan ID=2> 去年夏天，一位越南女记者曾在柏林向记者抱怨说：<sup>4</sup>“中国货太便宜啦，把越南的市场几乎都占了，到处都是‘中国制造’！”<sup>5</sup> 笔者当时回敬她：<sup>6</sup>“这又什么不好吗？”<sup>7</sup> 世界在享受‘中国制造’带来的便宜时，应想到中国如何辛苦才对啊！”</duan>

Translation: <sup>3</sup>*Last summer, a Vietnamese female journalist complained to me: <sup>4</sup>“Chinese goods are so cheap that they have almost occupied the Vietnamese market. There is ‘made in China’ everywhere!” <sup>5</sup>I argued back then: <sup>6</sup>“What’s wrong with that? <sup>7</sup>The world should appreciate Chinese for their hard work when enjoying the inexpensiveness of ‘made in China’.”*)

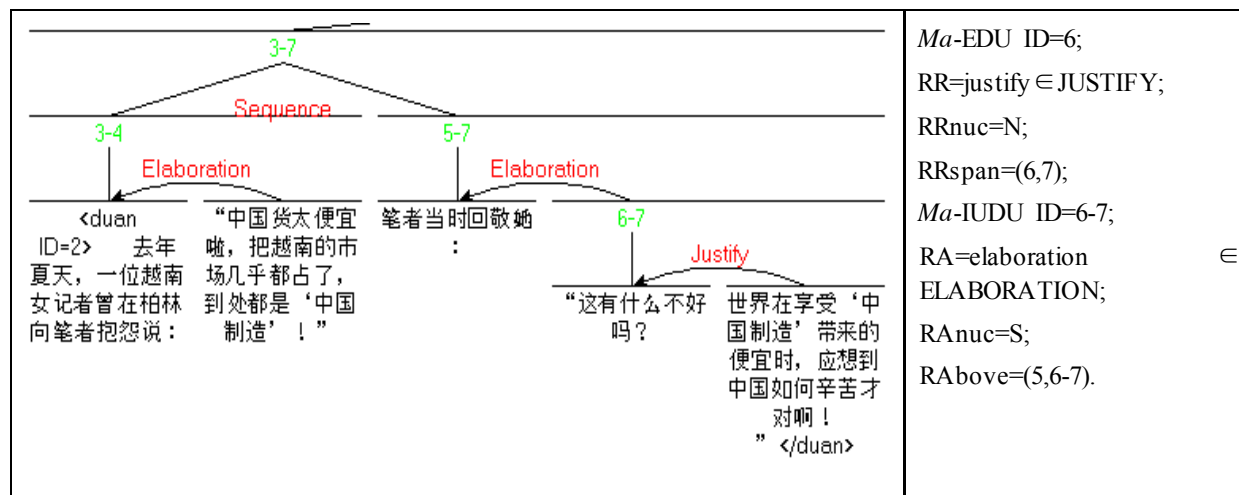


Figure 1. Sample rhetorical sub-tree from CRST (CJPL1995119.2)

**Statistical Tool**

SPSS 20.0 is employed for correlation tests. Since many of the linguistic variables are nominals, Lambda coefficient is used. And since it is not yet clear whether Variable A influences Variable B or vice versa, symmetric correlation is assumed between the variables.

**Results**

There are altogether 129 tokens of *ma* retrieved from CJPL400, scattered in 70 articles. Among them, 121 tokens appear in questions marked with (/or shared with) a question mark. 2 appear in quotations (or Chinese bookmarks), and 6 others at the end of titles/subtitles. Table 1 presents the frequency distribution of *ma* in the corpus.

Table 1

Distribution of *ma* in CJPL400 Texts

ParaS	OnlyS	1 <sup>st</sup> -S	Mid-S	Last-S	Total	Only-S	1 <sup>st</sup> -S	Mid-S	Last-S	Total
Title/1 <sup>st</sup> -p	10	2	3	3	18	7.75%	1.55%	2.33%	2.33%	13.95%
Mid-p	23	15	42	26	105	17.83%	11.63%	32.56%	20.16%	81.40%
Last-p	0	1	3	2	6	0.00%	0.78%	2.33%	1.55%	4.65%
Total	33	18	48	30	129	25.58%	13.95%	37.21%	23.26%	100.00%

**Correlation between *ma*-EDU location and RR**

The RR distribution in descending order of all the *ma*-EDUs is shown in Table 2: the top 2 RRs are SOLUTIONHOOD (41.1%) and CONJUNCTION (13.2%).

Table 2

*Rhetorical relations ma-EDUs play in CJPL400 texts*

RR	Title/1 <sup>st</sup> P	Mid-P	Last-P	Total	Only-S	1 <sup>st</sup> -S	Mid-S	Last-S	Total
<b>SOLUTIONHOOD</b>	9	<b>44</b>	0	<b>53</b>	30	8	11	4	53
	7.0%	<b>34.1%</b>	0.0%	<b>41.1%</b>	23.3%	6.2%	8.5%	3.1%	41.1%
<b>CONJUNCTION</b>	1	<b>12</b>	4	17	0	1	<b>8</b>	<b>8</b>	17
	.8%	<b>9.3%</b>	3.1%	13.2%	0.0%	.8%	<b>6.2%</b>	<b>6.2%</b>	13.2%
<b>ELABORATION</b>	0	15	0	15	0	2	9	4	15
	0.0%	11.6%	0.0%	11.6%	0.0%	1.6%	7.0%	3.1%	11.6%
<b>JUSTIFICATION</b>	3	11	1	15	2	2	7	4	15
	2.3%	8.5%	.8%	11.6%	1.6%	1.6%	5.4%	3.1%	11.6%
<b>CONTRAST</b>	2	<b>11</b>	0	13	0	3	5	5	13
	1.6%	<b>8.5%</b>	0.0%	10.1%	0.0%	2.3%	3.9%	3.9%	10.1%
<b>CAUSATION</b>	1	6	0	7	0	0	3	4	7
	.8%	4.7%	0.0%	5.4%	0.0%	0.0%	2.3%	3.1%	5.4%
<b>BACKGROUND</b>	2	4	0	6	0	1	4	1	6
	1.6%	3.1%	0.0%	4.7%	0.0%	.8%	3.1%	.8%	4.7%
<b>CONDITION</b>	0	1	1	2	0	1	1	0	2
	0.0%	.8%	.8%	1.6%	0.0%	.8%	.8%	0.0%	1.6%
<b>TOPIC-TRANSITION</b>	0	1	0	1	1	0	0	0	1
	0.0%	.8%	0.0%	.8%	.8%	0.0%	0.0%	0.0%	.8%
<b>Total</b>	18	105	6	129	33	18	48	30	129
	14.0%	81.4%	4.7%	100.0%	25.6%	14.0%	37.2%	23.3%	100.0%

A quick look by the naked eye finds that *ma*-EDUs in the first paragraphs are usually used in the SOLUTIONHOOD relation but never used so when they are used in the last paragraph of a commentary text. *Ma*-EDUs in the last paragraphs are never used in the ELABORATION or CONTRAST relations. In the mid-paragraphs, CONJUNCTION appears most often. The symmetric Lambda coefficient between the text location of *ma*-EDU and its rhetorical function (RR) reported by SPSS is 0.040, sig=0.042, indicating a correlation between the two variables.

*Ma*-EDUs in single-sentence paragraphs are mostly used in SOLUTIONHOOD. In multi-sentence paragraphs, *ma*-EDUs are mostly used in SOLUTIONHOOD at the opening of a paragraph, and in CONJUNCTION at the closing of a paragraph. The symmetric Lambda coefficient between the paragraph location of *ma*-EDU and its rhetorical function (RR) reported by SPSS is 0.229, sig=0.001, indicating a correlation between the two variables.

### ***Correlation between ma-IUDU Location and RA Function***

Table 3 reports the RA distribution of *ma*-EDUs (or the RR of *ma*-IUDU) in CJPL400. These IUDUs are mostly SOLUTIONHOOD (19.4%), ELABORATION (19.4%) and CONTRAST relations (14.0%). 7% of the *ma*-IUDUs serve as root of the discourse tree, or, in another word, cover the whole text.

The symmetric Lambda coefficient between text location of *ma*-IUDU and the RA function reported by SPSS is 0.141, sig=0.002, indicating a correlation between the two variables. The symmetric Lambda coefficient between the paragraph location of *ma*-IUDU and the RA function reported by SPSS is 0.108, sig=0.018, indicating a correlation between the two variables.

Table 3

*RA (Rhetorical relations of the Immediately Upper Discourse Unit) of ma-EDUs in CJPL400 texts*

RA	Title/1 <sup>st</sup> P	Mid-P	Last-P	Total	Only-S	1 <sup>st</sup> -S	Mid-S	Last-S	Total
SOLUTIONHOOD	8	16	1	<b>25</b>	4	4	11	6	25
	6.2%	12.4%	.8%	<b>19.4%</b>	3.1%	3.1%	8.5%	4.7%	19.4%
ELABORATION	0	23	2	25	6	1	11	<b>7</b>	25
	0.0%	17.8%	1.6%	19.4%	4.7%	.8%	8.5%	<b>5.4%</b>	19.4%
CONTRAST	1	15	2	18	2	1	8	7	18
	.8%	11.6%	1.6%	14.0%	1.6%	.8%	6.2%	5.4%	14.0%
JUSTIFICATION	0	14	1	15	2	2	7	4	15
	0.0%	10.9%	.8%	11.6%	1.6%	1.6%	5.4%	3.1%	11.6%
CAUSATION	0	13	0	13	3	2	4	4	13
	0.0%	10.1%	0.0%	10.1%	2.3%	1.6%	3.1%	3.1%	10.1%
BACKGROUND	0	9	0	9	4	0	3	2	9
	0.0%	7.0%	0.0%	7.0%	3.1%	0.0%	2.3%	1.6%	7.0%
ROOT	9	0	0	<b>9</b>	<b>9</b>	0	0	0	9
	7.0%	0.0%	0.0%	<b>7.0%</b>	<b>7.0%</b>	0.0%	0.0%	0.0%	7.0%
CONJUNCTION	0	8	0	8	0	4	4	0	8
	0.0%	6.2%	0.0%	6.2%	0.0%	3.1%	3.1%	0.0%	6.2%
TOPIC-TRANSITION	0	<b>7</b>	0	<b>7</b>	<b>3</b>	<b>4</b>	0	0	7
	0.0%	<b>5.4%</b>	0.0%	<b>5.4%</b>	<b>2.3%</b>	<b>3.1%</b>	0.0%	0.0%	5.4%
TOTAL	18	105	6	129	<b>33</b>	18	48	30	129
	14.0%	81.4%	4.7%	100.0%	<b>25.6%</b>	14.0%	37.2%	23.3%	100.0%

### Discussion

Corpus data report that the rhetorical functions of both *ma*-EDU and *ma*-IUDU (i.e. RR and RA) are statistically correlated with *ma*'s location in the commentary, and that a large portion of both RR and RA are in SOLUTIONHOOD relations (41.1% and 19.4% respectively). We speculate that such correlations exist mainly because *ma* has a strong questioning connotation, and that the question-answer scheme is a typical part of the argumentative structure (Liao, 1988). There are at least three supporting evidences:

First of all, 93.9% of *ma* tokens in CJPL400 are used with a question mark, taking up 13.7% of all the 845 question marks. The rest 6.1% of *ma* tokens are used in titles or parentheses where the question mark can be legitimately omitted. Since the question mark in written text indicates a

questioning tone, the high percentages demonstrate that *ma*-sentences in CJPL400 are nearly always used in questions.

Second, of the 41.1% (53/129) of *ma*-EDUs in SOLUTIONHOOD relations, 7 are in the title or opening paragraphs with an elaborated answer covering the rest of the text as the author offers his viewpoint to the reader. 44 are in the middle of the text, usually followed by an immediate answer, and sometimes further elaborated with justifying evidences. When there is no immediate answer, the strong correlation between *ma* and the question intonation calls on reader's attention to expect a delayed answer. If such an answer is not supplied in the following text, the pending status would cause much anxiety for the reader. Consequently, he would retrospect and ask himself whether he is expected by the author to give his own answer. This is exactly what the author wants the reader to do, as one major purpose of an argumentative essay is to motivate the reader to take actions accordingly.

Thirdly, when *ma*-sentences in the middle of texts do not involve in SOLUTIONHOOD relations, they are often used as rhetorical negations for CONTRAST (11/76), indicating an opinion change in the text (cf. Ford, 2000). The *ma*-sentences found in CONTRAST relation serve as attention-getting devices, for counterarguments are always helpful in debate. Attention-getting may also be the reason why many *ma*-sentences at the end of paragraphs (8/30) are used in a CONJUNCTION series -- so as to make the questions more impressive and the argument more prominent. If an unanswered question is left at the end of a commentary, it is found used to express the author's doubt or suspicion, as a euphemistic criticism or indirect suggestion to the administration. For instance, Example 2 raises a doubt on the banks for taking sufficient mends. The lingering doubt may effectively evoke the reader's actual negation.

Example 2 <duan ID=15>只是不知道，我们的银行珍惜储户的激励吗？清楚自己的社会角色吗？（尚德琪）</duan> (end of text)

Translation: <duan ID=15>Except that we don't know the following: Will our banks value the depositors encouragement? Do they understand the social roles they are supposed to play? (By Shang Deqi) </duan> (end of text)

While news commentaries as argumentative essays are monologues in appearance, they are dialogic in nature -- apart from excerpts of narrative dialogues or imaginative talks going on in the texts, there are also dialogues between author and reader. Through the use of *ma*-questions, the author gets the reader actively involved in a dialogue across time and space, as if being asked on the spot. The reader will actively await a solution to the problem, or work out one by himself. Thus, the questioning connotation of *ma* and the argumentative purpose of the author have jointly determined its strong interactive role in the text.

### Conclusion

This paper studies the discursive role *ma*-sentences play in natural texts, and find that their locations in both text and paragraph have statistical collocations with their discursive roles. Such usage might be due to the strong interrogative connotation of *ma*, and/or to the frequently used SOLUTIONHOOD relations ('Question-Answer' and 'Problem-Solution') in Chinese argumentative essays. Hopefully this quantitative study would offer new insight to *ma* as an interrogative particle.

### References

- Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In J. van Kuppevelt & R. Smith (Eds.), *Current Directions in Discourse and Dialogue* (pp. 85-112): Kluwer Academic Publishers.
- Ford, C. E. (1994). Dialogic aspects of talk and writing: Because on the interactive-edited continuum. *Text*, 14(4), 531-554.
- Ford, C. E. (2000). The Treatment of Contrasts in Interaction. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause-Condition-Concession-Contrast: Cognitive and Discourse Perspectives* (pp. 283-312). Berlin: Mouton De Gruyter.
- Lan, J. (2013). *The Research on the Succession of Interrogative Sentences with Ma*. MA, Huazhong Normal University.
- Liao, Q. (1988). Argument Structure in Discourse. *Yuyan Jiaoxue yu Yanjiu (Language Teaching and Research)* (01), 86-101.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 3(8), 243-281.
- Spenader, J., & Lobanova, A. (2009). *Reliable Discourse Markers for Contrast Relations*. Paper presented at the Proceedings of the 8th International Conference on Computational Semantics, Tilburg. 210-221.
- Stede, M. (2004). *The Potsdam commentary corpus*. Paper presented at the Proceedings of the 2004 ACL Workshop on Discourse Annotation. 96-102.
- Taboada, M., & Mann, W. C. (2006). Applications of Rhetorical Structure Theory. *Discourse Studies*, 8, 567-588.
- Yue, M. (2008). Rhetorical structure annotation of Chinese news commentaries. *Journal of Chinese Information Processing*, 4(2), 19-23.