Mandarin Chinese *ma*: Q morpheme, SA intensifier, or PQP?

Using a novel paradigm for eliciting semantic judgments, this work establishes the facts for the embedding of Mandarin yes-no question marker *ma* and evaluates three competing proposals for its distribution.

Background One way to form a yes-no question in Mandarin is by using the particle *ma* (1), which is traditionally considered unembeddable (e.g., Pan 2019), but recent observations show that *ma* may be embedded under restricted conditions (2). Theoretical proposals for *ma* make different predictions about the range of contexts that allows its embedding. Among others: Cheng (1991) proposed that *ma* realizes a C^0 and is a clause-typing question morpheme (3); Kim (2019) proposed that *ma* adjoins to a speech act phrase and is a speech act–intensifying operator (4); more recently, Dayal (2023) proposed that *ma* resides in a perspective phrase between SAP and CP (5a) and triggers perspectival centering, introducing a semantic/pragmatic requirement that the CP question is p(otentially)-active for the perspectival center (5b).

(1) a. Mingtian hui xia yu.
      tomorrow will fall rain
      'It will rain tomorrow.'

    b. Mingtian hui xia yu <u>ma</u>?
      tomorrow will fall rain MA
      'Will it rain tomorrow?'

(2) John xiang zhidao [xia yu le <u>ma</u> zuotian].
    John want know fall rain LE MA yesterday
    'John wants to know whether it started raining yesterday.'   (Bhatt & Dayal 2020: 1122)

(3) $[_{CP}[_{TP}\ldots][_{C^0}\ ma]]$ *CP proposal*   (4) $[_{SAP}[_{CP}[_{TP}\ldots]]\ ma]$ *SAP proposal*

(5) a. $[_{SAP}[_{PerspP}[_{CP}[_{TP}\ldots]]][_{Persp^0}\ ma]]$ b. $[\![\text{Persp}^0]\!] = [\lambda Q.\lambda x: Q$ is p-active for $x.\,Q]$

The PerspP proposal is motivated by the study of embedded inverted questions (EIQs; McCloskey 2006) in, among other English varieties (Woods 2016), Irish English (6). McCloskey convincingly showed that EIQs are not quotations, as personal pronouns (e.g., 3SG) would behave differently if they were (7).

(6) [Every male physicist]$_i$ wonders [will he$_i$ be awarded a Nobel Prize].   (McCloskey 2006: 89)

(7) [Every male physicist]$_i$ wonders, ["will he$_{j/*i}$ be awarded a Nobel Prize?"]

The embeddability of EIQs is affected by three factors (McCloskey 2006, Dayal 2023): the matrix predicate type, subject person, and illocutionary force. This can be captured by Dayal's semantics of Persp^0 (5b), as these factors determine who the perspectival center is (which interacts with the felicity conditions of the speech acts of asking or asserting) and whether the embedded question is p-active for the perspectival center. For example: the unacceptability of (8b) can be explained by a semantic/pragmatic mismatch, viz., if someone knows the answer to a question, it cannot be p-active for them; and the acceptability of (9b) can be ascribed to the observation that a positive answer leads to a resolution of the embedded question.

(8) *Effect of the matrix predicate type: Matrix declarative and 3rd-person matrix subject*
    a. Mary$_i$ wonders [will she$_i$ be awarded a Nobel Prize].   ✓wonder-*type*
    b. #Mary$_i$ knows [will she$_i$ be awarded a Nobel Prize].   #know-*type*

(9) *Effect of the matrix illocutionary force: 3rd-person matrix subject and* know-*type matrix predicate*
    a. #Mary$_i$ knows [will she$_i$ be awarded a Nobel Prize].   #declarative
    b. Does Mary$_i$ know [will she$_i$ be awarded a Nobel Prize]?   ✓interrogative

Methods To determine the most suitable proposal for *ma* among these three, I survey native Beijing Mandarin speakers and test how the matrix predicate type, subject person, and illocutionary force affect the embeddability of *ma*. These factors amount to 2 types × 3 persons × 2 forces = 12 conditions; each condition has four items in the survey, using different matrix predicates and embedded questions. An example item in the *wonder*-3rd-declarative condition looks like (10). To ensure that the items cannot be read as quotations, all embedded questions in the survey contain a 3SG referring to the matrix subject (if 3SG) and a 1SG referring to the speaker. To discourage a reading in which *ma* takes scope over the matrix clause, a comma is added after the matrix predicate, and the embedded questions are ended with a question mark.

(10) Xiaohong$_i$ xiang zhidao, [ta$_i$ de fenshu bi wo$_j$ gao <u>ma</u>?]
     Xiaohong want know 3SG POSS score than 1SG high MA
     'Xiaohong$_i$ wants to know whether her$_i$ score was higher than mine$_j$.'

The brackets, subscripts, and underlining are only for the reader's convenience and are not in the survey.

The instruction in (11) was given at the beginning of the survey. As per the instruction, each target (in option (b)) was presented with a two-sentence equivalent (in option (a)). In the two-sentence equivalent, the first and second sentence conveyed the matrix and embedded content of the target, respectively as in (12).

(11) Below are some pairs of expressions, each with an accompanying context. In each pair of expressions, option (a) consists of two sentences, and option (b) attempts to convey the same meaning using only one sentence. Your task is to understand the accompanying context, read both expressions aloud, and decide whether the two sentences in (a) can be naturally conveyed using the single sentence in (b) in the accompanying context. If you believe it is possible, choose (b). If not, choose (a). If both (a) and (b) seem unacceptable for you, choose (a) as well.

(12) Xiaohong$_i$ xiang zhidao yi jian shi. Na jiu shi ta$_i$ de fenshu shi bu shi bi wo$_j$ gao.
      'Xiaohong$_i$ wants to know one thing. That thing is whether her$_i$ score is higher than mine$_j$.'

The survey forces the participants to choose between two options conveying the same meaning (cf. Ma-howald et al. 2016). One option is tedious, and the other succinct. If the participants prefer the tedious option, it suggests that the succinct option may be semantically unacceptable, tempting them to use more sentences; and vice versa. The predicted results by the three proposals are summarized in the middle columns of (13), where 0 represents the tedious option and 1 the succinct one. The CP proposal predicts that the acceptability of embedded *ma* should in general be unaffected, but due to the peculiarity to inquire one's own knowledge or desire, the 1st-interrogative conditions are excluded. The SAP proposal predicts that *ma* questions necessitate the speaker not knowing the answer and believing that the addressee knows, resulting in the exclusion of a few more conditions. The PerspP proposal predicts that results should mirror whether the embedded question is p-active to the perspectival center and resemble the distribution of English EIQs.

Results Judgments were collected from 12 native Beijing Mandarin speakers (ages 22–24, $\mu = 23.67$). During the survey, items were presented in a random order. Mean acceptability rates of the succinct option are presented in the last two columns of (13), where values higher than 50% are in bold. Mixed effects logistic regression models were fit and show that all distinctions between bolded and nonbolded conditions are statistically significant. We observe that none of the three proposals perfectly aligns with the the results, with the PerspP proposal having the fewest conditions significantly differing from the prediction.

(13)

			CP		SAP		PerspP		Results	
			Know	*Wonder*	*Know*	*Wonder*	*Know*	*Wonder*	*Know*	*Wonder*
	Declarative	1st	1	1	0	1	0	1	.02	**.83**
		2nd	1	1	1	0	0	0	.35	.21
		3rd	1	1	1	1	0	1	.15	**.73**
	Interrogative	1st	0	0	0	0	0	0	.23	.06
		2nd	1	1	0	0	1	0	**.77**	**.75**
		3rd	1	1	1	1	1	0	**.63**	**.67**

Discussion The difference between the PerspP proposal's prediction and the results lies in the *wonder*-2nd/3rd-interrogative conditions, suggesting that the contextual requirement of *ma* might be a weaker form of the proposed semantics of Persp^0. While they both require the perspectival center's possible ignorance to the embedded question (the **ignorance requirement**), *ma* does not require its investment in resolving the question but only the possibility that the addressee knows the answer (the **knowledge requirement**):

(14) $[\![\text{ma}]\!] = [\lambda Q.\lambda x\colon \diamond\neg\text{know}(x, Q) \wedge \diamond\text{know}(\text{addressee}_C, Q).Q]$

The *know*-declarative conditions are unacceptable because they violate the ignorance requirement (similar to Persp^0); the *wonder*-2nd-declarative condition is unacceptable because it violates the knowledge requirement (same result with Persp^0 but for a different reason); and the *wonder*-2nd/3rd-interrogative conditions are acceptable because, despite a positive answer not resolving the embedded question (which makes them unacceptable with Persp^0), they meet the knowledge requirement: when someone asks if the addressee or another person wants to know the answer to a question, it is possible that the addressee knows it.