Specific unknowns: a case study of epistemic indefinites in Cantonese

Tommy Tsz-Ming Lee

Abstract. This paper concerns how languages bundle an existential claim and an ignorance inference in a nominal expression. I present a case study on epistemic indefinites (EIs) in Cantonese and show that Cantonese EIs have a different morphological makeup ($m + zì + \text{WH} \; \text{‘not + know + WH’}$), when compared to other more discussed EIs. I suggest that the ignorance component associated with $mzì$-WH is a conventional implicature and that $m$-$zì$ obtains an adnominal usage via grammaticalization. It denotes a choice function that comes with an ignorance component that is inherited from the predicative meaning of $m$-$zì$.

Keywords. indefinites; ignorance; specificity; choice function; grammaticalization

1. Introduction. This paper concerns how languages bundle an existential claim and an ignorance inference (over the witness of an indefinite) in a nominal expression. For example, the indefinite marker algún in Spanish convey both meanings at the same time, illustrated with its incompatibility with the “namely”-phrase that explicitly conveys the speaker’s knowledge.

(1)#María se casó con algún estudiante del departamento de lingüística:
María se casó con ALGÚN estudiante del departamento de lingüística:

en concreto con Pedro
namely with Pedro
‘María married a linguistics student, namely Pedro.’
(Alonso-Ovalle and Menéndez-Benito 2010, p.2)

It should be noted that not all indefinite markers conventionally convey the ignorance component. Take English a/some as an example. Both are compatible with the “namely”-phrase.

(2) Mary married a/some linguistics student, namely, Peter.

The former type of indefinites thus represents a subtype of indefinites, which is also known as epistemic indefinites (EIs, Alonso-Ovalle and Menéndez-Benito (2015)). The primary goal of this paper is to present a case study on EIs in Cantonese, which take the form of $m + zì + \text{WH}$, literally, ‘not + know + WH’.¹ It is incompatible with the “namely”-phrase.

(3) Aaming tai-zo [mzì bin-bun syu], (# zîkhai Hunglaumung)

Aaming read-PERF MZI which-CL book namely Dream.of.the.red.chamber
‘Aaming read some book, namely, *Dream of the Red Chamber.* ’

Despite the presence of a predicate, the whole string is used as an indefinite nominal expression (i.e. the object of ‘read’). Since $mzì$ in (3) occupies a position unavailable to other

¹ EIs that involve a $w$-expression are not uncommon, e.g. Japanese WH-$k$î (Alonso-Ovalle and Shimoyama 2014), Sinhala WH-$d$à (Slade 2015), Tiwa WH-$k$î (Dawson 2018).
predicates, I thus treat it syntagmatically as an adnominal marker and gloss it as MZI, in order to differentiate it from the ordinary use of m-zi as a negated predicate.

(4) ngo m-zi [Aaming tai-zo bin-bun syu]  
 I not-know Aaming read-PERF which-CL book  
‘I don’t know which book Aaming read.’

Focusing on the usage of the mzi-WH string in (3) (henceforth the mzi-indefinite), I suggest in section 2 that it represents a novel type of EIs in terms of the properties of the ignorance component, earning mzi-indefinites a unique empirical profile in contrast to EIs in other languages.

Based on these observations, I argue in section 3 that the ignorance component in Cantonese EIs is best characterized as a conventional complicature (Grice 1975; Potts 2005; Horn 2007), in contrast to a conversational implicature. I propose that mzi is semantically a choice function that select an alternative set as its argument and return a member of it. Importantly, mzi is associated with a conventional implicature, where the ignorance component represents a non-at-issue content.

In section 4, I touch on a less discussed but important link between the nature of the ignorance component and the source/origin of the markers of EIs. I suggest that the ordinary predicate m-zi ‘not-know’ obtains a non-predicative usage and develops into an adnominal marker. Crucially, the lexical meaning of m-zi, which depict at-issue content in its predicative use, becomes a non-at-issue content in its adnominal use. I conclude in section 5.

2. Properties of mzi-indefinites. In this section, I report three observations on the ignorance component on mzi-indefinites: (i) it is not cancellable or reinforceable; (ii) it survives intensional operators (i.e. it denotes “specific unknowns”) and (iii) it can scope below quantifiers and be distributed.

2.1. Cancellability and reinforcement. The ignorance component of mzi-indefinites do not show the signature properties of a conversational implicature, as the sentence with a mzi-indefinite cannot be cancelled by a follow-up ‘I know’-clause as in (5), suggesting that the ignorance component is not calculable.

(5) Non-cancellability  
#Aaming tai-zo mzi bin-bun syu, ji ngo zidou hai bin-bun  
Aaming read-PERF MZI which-CL book, and I know be which-CL  
‘Aaming read some book, and I know which (book it is).’

Also, it cannot be reinforced by a follow-up ‘I don’t know’-clause without giving rise to a sense of redundancy.

(6) Redundancy of Conjunction (Horn 1972); Non-reinforceability (Sadock 1978)  
#Aaming tai-zo mzi bin-bun syu, ji ngo m-zi hai bin-bun  
Aaming read-PERF MZI which-CL book, and I not-know be which-CL  
‘Aaming read some book, and I don’t know which (book it is).’

Note that the opposite is observed with Spanish and Japanese EIs (see examples in Alonso-Ovalle and Menéndez-Benito 2010; Alonso-Ovalle and Shimoyama 2014, respectively).

Here, it is instructive to consider a domain requirement on EIs, namely, the anti-singleton
constraint (Alonso-Ovalle and Menéndez-Benito 2010). Arguably, it is crucial to ignorance component that is derived via a quantity implicature. It requires the domain of quantification of the EI to be non-singleton, such that the use of EIs would implicate the speaker’s ignorance over the witness.

Adopting a similar test in Alonso-Ovalle and Menéndez-Benito (2010), consider a scenario where the speaker entered a room and saw a professor dancing on the table. (7) can be felicitously uttered, with the speaker pointing at this professor. This suggests that mzi-indefinites are compatible with a singleton domain.

(7) Absence of the anti-singleton requirement

taihaa! mzi bin-go gausau hai toi soengmin tiumou
look mzi which-CL professor at table top dance
‘(Pointing at the professor) Look! Some professor is dancing on the table!’

These diagnostic tests speak against a conversational implicature approach to ignorance component on mzi-indefinites, which is suggested for EIs in Spanish and Japanese (Alonso-Ovalle and Menéndez-Benito 2010; Alonso-Ovalle and Shimoyama 2014).\(^2\)

2.2. OBLIGATORY WIDE INTENSIONAL SCOPE. Another property of a mzi-indefinite concerns its interpretation with regard to intensional operators. The ignorance component is retained when embedded under attitude verbs and deontic modals. In both sentences below, the speaker still conveys the his/her ignorance over the witness of the indefinite. In other words, the ignorance component survives intensional contexts (i.e. it is projective).\(^3\)

(8) Wide scope over attitude verbs

Aafan soeng tong mzi bin-go jisang jitfan
Aafan want with mzi which-CL doctor marry
‘Aafan wants to marry to some doctor ... ’
(i) ✓ ... they know each other for two years. scopally specific
(ii) × ... but she does not know any doctor. scopally non-specific

(9) Wide scope over deontic modals

Aafan jatdingjiu tong mzi bin-go naamjan gitfan
Aafan must with mzi which-CL man marry
a. ✓ ‘There is some man that Aafan must marry to.’ scopally specific
b. × ‘Aafan must marry to a man (whoever he is).’ free choice

Notably, the mzi-indefinites are unambiguously scopally specific. They refer to specific referent unknowns from the perspective of the speaker. This is reminiscent of certain types of indefinites as in Stát’ímct (Matthewson 1999) and Tiwa (-khi, Dawson 2018), where they

\(^2\) There are also proposals suggesting that the ignorance component may be a manner implicature, which is attributable to lexical competition with another expression (cf. lexical blocking, McCawley 1978), as in Russian (Geist 2008) and Tiwa (Dawson 2018). In view of the absence of obvious competitors to mzi, I do not consider this possibility.

\(^3\) EIs in other languages display non-uniform scope interactions with different intensional operators: the EI may lose the ignorance component, resist embedding, or give rise to a free choice reading, etc. See, for example, discussions in Aloni and Port (2015) and Šimík (2014).
take obligatory wide scope. One difference, however, is that mzi-indefinites can in fact take narrow quantificational scope, which I discuss in the next subsection.

2.3. VARIABLE QUANTIFICATIONAL SCOPE. Let us first focus on the indefinite/existential meaning of a mzi-indefinite. (10) shows that it can scope above or below the universal quantifier, giving rise to two possible readings in (10b) and (10c). Thus mzi-indefinites do not take obligatory wide scope with respect to quantifiers.

(10) a. mui-go hoksaang dou hok-gwo [mzi bin-zung auzau jyujin]
   every-CL student all learn-EXP MZI which-CL European language
   ‘Every student has learned some European language.’
   b. Wide: \( \exists y [\text{an-unknown-European-language}(y) \land \forall x [\text{student}(x) \rightarrow \text{learned}(x,y)]] \)
   c. Narrow: \( \forall x [\text{student}(x) \rightarrow \exists y [\text{an-unknown-European-language}(y) \land \text{learned}(x,y)]] \)

Crucially, when the mzi-indefinite is interpreted narrowly, the ignorance component is distributed over the universal quantifier. Specifically, (10c) convey a meaning where for each student, s/he has learned some language unknown to the speaker (i.e. the speaker’s ignorance is spread over all student-language pairs).

To see one more example illustrating this property, (11) favors a narrow scope reading (for pragmatic reasons). It conveys that for each famous song in the 80’s, it is rearranged from some Japanese song unknown to the speaker.

(11) a. mui-sau batsap-nindoi coetman ge go dou hai goipin zi mzi bin-sau
   every-CL in.the.eighties famous GE song all be rearrange from MZI which-CL
   jatman-go
   Japanese-song
   ‘Every famous song in the eighties is rearranged from some Japanese song.’
   b. #Wide:
      \( \exists y [\text{an-unknown-Jap.-song}(y) \land \forall x [\text{a-famous-song}(x) \rightarrow \text{be.rearranged.from}(x,y)]] \)
   c. Narrow:
      \( \forall x [\text{a-famous-song}(x) \rightarrow \exists y [\text{an-unknown-Jap.-song}(y) \land \text{be.rearranged.from}(x,y)]] \)

Note that this observation on Cantonese EIs contrasts with Japanese ones. It is reported that the ignorance component disappears when interpreted narrowly.

(12) Japanese

| Dono kyooju-mo dare-ka gakusee-to odotteru. |
| which professor-MO who-KA student-with is.dancing |
| ‘Every professor is dancing with some student.’ (Alonso-Ovalle and Shimoyama 2014) |

Alonso-Ovalle and Shimoyama (2014) suggests that (12) can be felicitously continued by a follow-up question by the hearer: ‘Who is dancing with who?’, signaling the absence of the ignorance component. The same type of follow-up question sounds infelicitous to both (10a) and (11a) (e.g. as if the hearer is not listening to the speaker).

3. Analysis. Taking stock, the ignorance component of mzi-indefinites shows a unique empirical profile, in comparison to EIs in other languages.
The ignorance component of mzi-indefinites
a. it cannot be cancelled or reinforced;
b. it cannot be embedded under intensional operators;
c. it can take narrow quantificational scope and be distributed.

I suggest that the properties in (13a) and (13b) follow if the ignorance component is treated as a conventional implicature and that (13c) can be captured if mzi is a choice function that comes with this implicature (which is a type of non-at-issue content). I discuss some further motivations for this suggestion below.

3.1. Motivation for a Conventional Implicature Approach. I adopt a general definition of conventional implicature, taken from Potts (2015), which basically follows the suggestions in Grice (1975) and Horn (2007).

(14) Meaning p is a conventional implicature of phrase S if, and only if:
   a. p is a conventional (encoded) property of a lexical item or construction in S;
   b. p is entailed by S; and,
   c. p’s truth or falsity has no effect on the at-issue content of S.

The ignorance component of mzi-indefinites is obviously encoded by mzi and we have seen that it cannot be cancelled, satisfying both (14a) and (14b). Concerning (14c), I adopt the ‘yes, but...’-test to illustrate the claim (Karttunen and Peters 1979; Potts 2005). Observe that in response to (15a), the hearer can follow up by agreeing on the at-issue existential claim, while disputing the ignorance component.

(15) a. Aaming tai-zo [mzi bin-bun syu] = (3)
    Aaming read-PERF MZI which-CL book
    ‘Aaming read some book (I don’t know which).’

   b. hai aa3. batgwò nei jinggoi zidou hai bin-bun gaa3
      yes SFP but you probably know be which-CL SFP
      ‘Yes, but you probably know which book it is.’

To see a contrast, (15b) would be an infelicitous follow-up to (16), where the speaker of (15b) sounds to contradict himself/herself.

(16) ngo m-zi [Aaming tai-zo bin-bun syu] = (4)
    I not-know Aaming read-PERF which-CL book
    ‘I don’t know which book Aaming read.’

Note that different proposals have been suggested along the line of a non-Gricean approach to the ignorance component. For example, the ignorance component is argued to indicate intended referential vagueness where the EI marker encodes anti-specificity, e.g. French un quelconque (Jayez and Tovena 2006), Greek -dhipote (Giannakidou and Quer 2013). But we have seen that mzi-indefinites are specific indefinites. Alternatively, the EI marker is proposed to trigger an obligatory shift in identification method and the shift is regulated by a felicity condition (i.e. non-vacuous shift), as advocated by Aloni and Port (2015) for German irgendein and Italian un qualche and adopted by Šimík (2014) for Czech -si and Slade (2015).
for Sinhala harildo. As far as I can see, the properties of mzi are compatible with this proposal, with the difference being treating the ignorance component as a requirement by some felicity condition or a presupposition or a conventional implicature. I leave further comparison among these options to future research.

3.2. Motivation for a Choice-Functional Analysis. Mzi-indefinites display ‘exceptional’ wide scope behaviors, where they can take scope from within a syntactic island. In (17), the mzi-indefinite in a complex NP takes wide scope over the matrix subject.

(17) mui-go hoksaang dou tengdou [hokhaau kwaidingjiu hok [mzi bin-zung jyujin] every-CL student all heard school require learn MZI which-CL language ge siusik]
GE news
‘There is some language s.t. every student heard the news that the school requires (them) to learn it.’


3.3. Implementation. Assuming a multi-dimensional semantic framework (Karttunen and Peters 1979; Potts 2005), I suggest that the semantics of mzi can given as follows. While the existential meaning constitutes at-issue content, the ignorance component are taken to be non-at-issue (listed as a conventional implicature).

(18) A multi-dimensional semantics of mzi
a. At-issue content:
\[mzi\] = \lambda P_{<e,t>}. g(i)(P), where g(i) \in D_{\text{choice function}}<<e,t>,e>
b. Conventional implicature:
The speaker doesn’t know (i.e. fails to identity in a relevant way) the referent chosen by the choice function.

I illustrate how the suggestion derives the narrow scope reading of (10), where the ignorance component is distributed over the universal quantifier. The relevant example and the target reading are repeated below.

(19) a. mui-go hoksaang dou hok-gwo [mzi bin-zung auzau jyujin] = (10) every-CL student all learn-EXP MZI which-CL European language
‘Every student has learned some European language.’

b. Narrow:
\[\forall x[\text{student}(x) \rightarrow \exists y[\text{an-unknown-European-language}(y) \land \text{learned}(x,y)]]\]

Let us assume a modification structure of a mzi-indefinite as in (20a). Assume further that wh-expressions denote alternative sets (Kratzer and Shimoyama 2002; Beck 2006, i.a.) and they serve as the argument of the mzi, which is a choice function as proposed. The at-issue content is depicted in (20b).
(20) a. The internal structure of the mzi-indefinites:
   \[ \text{NP} \text{ mzi [NP which European.language ] } \]

b. At-issue-content:
   \[ mzi, \] \[ \text{g} (\text{[which European.language]}) \]
   via (18a)
   \[ = \lambda X. g(i)(X) (\{ x: \text{European.language}(x) \} ) \]
   by Functional Application
   \[ = g(i)[x: \text{European.language}(x) \] \]
   \[ = g(i)[\text{Spanish, German, ... } \] \]

Note that I adopt the suggestion in Winter (1997) that the choice function can be existentially bound at its base position. The meaning of (19a) can be stated as follows. Since the ignorance component is associated with the choice function, it is distributed altogether.

(21) The meaning of (19a)

a. At-issue-content: \( \forall x[\text{student}(x) \rightarrow \exists f \{ \text{learned}(x, f[\text{Spanish, German, ... }]) \} ] \] 

b. Conventional implicature: The speaker doesn’t know the referent chosen by f.

4. A note on grammaticalization. Returning to the origin/source of the ignorance component in m-zi, I suggest the (negated) attitude verb m-zi obtains a non-predicative usage and develops into an adnominal marker that denotes a choice function. The lexical meaning of m-zi is carried over to the choice function, constituting a sub-type of choice function. Because of this language specific development, the ignorance component associated display a different empirical profile when compared to other EIs.

4.1. THREE POSITIONS OF m + zi. To trace some development of m-zi, I suggest that it displays the following grammaticalization path:

(22) 1 an attitude verb \( \rightarrow \) 2 a “raising” verb \( \rightarrow \) 3 an adnominal modifier

The three usages correspond to the following three examples. Note that (24) is suggested to be “an attitudinal marker” (Yap and Chor 2014). But since it is substantially different from (23) in terms of the absence of the attitude holder (i.e. an embedded subject is occupying a matrix subject position) and the requirement on clause type (i.e. it is only compatible with interrogative clauses), I adopt the term “raising”.

(23) ngo m-zi [Aaming tai-zo bin-bun syu] an attitude verb; =\( 4 \)
     I not-know Aaming read-PERF which-CL book
     ‘I don’t know which book Aaming read.’

(24) Aaming m-zi tai-zo bin-bun syu a “raising” verb
     Aaming not-know read-PERF which-CL book
     ‘It is not known which book Aaming read.’

(25) Aaming tai-zo [mzi bin-bun syu] an adnominal modifier; =\( 3 \)
     Aaming read-PERF MZI which-CL book
     ‘Aaming read some book (I don’t know which).’

The difference of these three usages is summarized in Table 1. The usage in 2 appears to mark a transition stage of the other two usage, given the partial overlapping among these usages.
4.2. CORPUS DATA. This suggestion is also supported by corpus data. According to (i) Early Cantonese Colloquial Texts: A Database (data mainly in 19th century) and (ii) A Linguistic Corpus of Mid-20th Century Hong Kong Cantonese, the attested instances of the string \( m + zi \) are summarized as follows.\(^4\)

<table>
<thead>
<tr>
<th>m-zi as ...</th>
<th>Attitude holder</th>
<th>Complement</th>
<th>Ignorance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 an attitude verb</td>
<td>overt</td>
<td>clauses</td>
<td>at-issue</td>
</tr>
<tr>
<td>2 a raising verb</td>
<td>null</td>
<td>at-issue</td>
<td></td>
</tr>
<tr>
<td>3 an adnominal modifier</td>
<td>null</td>
<td>WH</td>
<td>non-at-issue</td>
</tr>
</tbody>
</table>

Table 1. Different usages of the string m-zi

<table>
<thead>
<tr>
<th>m-zi as ...</th>
<th>(i) Early Can.</th>
<th>(ii) Mid-20th HKC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 an attitude verb</td>
<td>4/60</td>
<td>44/110*</td>
</tr>
<tr>
<td>2 a raising verb</td>
<td>2/60</td>
<td>17/110*</td>
</tr>
<tr>
<td>3 an adnominal modifier</td>
<td>0/60</td>
<td>7/110*</td>
</tr>
</tbody>
</table>

Table 2. Frequency of \( m \text{-}zi \) (*total hit: 1098, counting the first 10%)

The data set is admittedly small, but it seems appropriate to suggest that compared to 2, 3 emerges relatively recently and that the usage of 2 is more frequent than 3. I leave a more comprehensive investigation into the development of \( m \text{-}zi \) to future work.

4.3. FUSION OF PREDICATE AND \( \text{wh} \)-EXPRESSIONS. Cross-linguistic data reveal that it is not uncommon for \( \text{wh} \)-expressions to develop into indefinites by fusing with predicates (Haspelmath 1997, p.131). Here are some examples in European languages.

(26) a. Middle High German
   \( ne \text{ weil} \text{ wr} \) ‘(I) don’t know who’ \( \rightarrow \) \( neizwer \) ‘somebody’

b. Old English
   \( ne \text{ w"at} \text{ hw"a} \) ‘(I) don’t know who’ \( \rightarrow \) \( n"althw"a \) ‘somebody’

c. French
   \( Je ne sais \text{ pas} \text{ quel} \) ‘I don’t know which’ \( \rightarrow \) \( je ne sais quel \) ‘some kind of’

Interestingly, there is in fact independent evidence showing that the fusion of a predicate with a \( \text{wh} \)-expression is closely connected to the (non-)at-issue nature the ignorance component. As mentioned briefly in section 2.1, the ignorance component of EIs in (Tokyo) Japanese (i.e. \( \text{wh-ka} \)) is taken to be a conversational implicature as it is cancellable and reinforceable, as illustrated below.

\(^4\)The counting omits instances of \( m \text{-}zi \) in A-not-A form, in answer fragments, in idioms, or with null/nominal arguments, which do not necessarily indicate one of the three usages under discussion.
Tokyo dialect of Japanese

Ken-wa [dare-ka] gengogaku-no gakusei-to kekkonshita. jitsuwa dare-da-ka
Ken-TOP who-KA linguistics-GEN student-with married in.fact who-COP-Q
shitteru.
know

‘Ken married a linguistics student. In fact, (I) know who it is.’

(Alonso-Ovalle and Shimoyama 2014, p.14)

In Gifu dialect, an additional morpheme *syan* can be attached to a WH-ka expression. Morphology-wise, *syan* is presumably a phonologically reduced form of *shi + ran* “know + not”.5 While the sentence in (28) also conveys the ignorance component of the speaker in a similar way as (27), it cannot be cancelled.

Gifu dialect of Japanese

Ken-wa [dare ka-syan]-to kekkonshi-tot-ta kedo, (#boku-wa zituwa
dare-da-ka sit-to-ru).

who-COP-KA know

‘Ken married someone. In fact, I know who it is.’

(p.c. Teruyuki Mizuno)

The contrast between WH-ka and WH-ka-syan is by itself very intriguing but further comparisons must await a separation occasion. What is relevant to the current discussion is that the WH-ka-syan in Gifu dialect, together with Cantonese mzi-indefinites, lends support to a close link between the morphological makeup of EIs and the non-at-issue nature of the ignorance component.

5. Conclusions. In this paper, I have showed that Cantonese EIs have a different morphological makeup than other more discussed EIs. I suggested that the ignorance component associated with mzi-indefinites is a conventional implicature and that m-zi obtains an adnominal usage via grammaticalization. It serves as a choice function that comes with an ignorance component that is inherited from the predicative use of m-zi. For future work, a more comprehensive comparison on how languages bundle the existential claim and the ignorance inference is much desired. Specifically, it may be interesting to see to what extent the properties of the ignorance component reveal how EIs emerge (e.g. grammaticalization, lexical competition, conversational implicature, etc.) and why a language adopts a particular way of bundling, but not the other.

References.


---

5 I thank Ken Hiraiwa and Teruyuki Mizuno for discussions.


McCawley, James D. 1978. “What is conversationally implicated by an utterance depends not only on the utterance but on what other utterances the speaker could have produced but did not. For example, a declarative sentence A.” In Syntax and Semantics 9: Pragmatics, edited by Peter Cole, 245–259. Academic Press.


