Sinhala epistemic indefinites with a certain *je ne sais quoi*

Benjamin Slade*

13th June 2014

Abstract

This paper focusses on the examination of the properties of epistemic indefinites in Sinhala, with comparison to other languages, including English and Hindi. In particular, attention is placed on the determination of the felicity conditions for the two morphologically- and pragmatically-distinct Sinhala epistemic indefinites: WH+*də* and WH+*hari*, with a concentration on the (un)availability of identification methods associated with particular epistemic indefinites.

1 Introduction

Crosslinguistically, we find both interlanguage and intralanguage variation in the felicity conditions of epistemic indefinites. The primary purpose of this paper is to examine the felicity conditions for epistemic indefinites in Sinhala, particularly with respect to their signalling of the availability/unavailability of particular identification methods, with comparison to epistemic indefinites in various European languages, as described in previous literature, as well as using novel data concerning English and Hindi epistemic indefinites.

Sinhala is especially interesting in its possession of two morphologically-distinct epistemic indefinites with distinct sets of felicity conditions: WH+*də* and WH+*hari*. I note further that Sinhala appears to not be alone in the possession of multiple lexically- and pragmatically-distinct epistemic indefinites, as too in English the felicity conditions of *some NP* differ from those of *some NP or other*, with a similar distinction being found in Hindi between reduplicated and non-reduplicated forms of indefinites.

In the next section, Section 2, I examine the properties of Sinhala’s two epistemic indefinites, with comparison to epistemic indefinites in other languages (particularly English), focussing on the felicity conditions of the two epistemic indefinites in Sinhala as they relate to identification method. Section 3 examines the specificity requirements of Sinhala and English epistemic indefinites. Finally, Section 4 provides concluding remarks and brief discussion of epistemic indefinites in other South Asian languages, specifically Hindi and Malayalam. This section also examines the relation between epistemic indefinites and their component morphology, given that in many languages, including Sinhala, the component pieces of epistemic indefinites appear in other environments, such as in the construction of interrogatives. Given this recurrence of the morphological components of epistemic indefinites in languages like Sinhala in a wide variety of environments, an analysis which treats epistemic indefinites in isolation fails to provide any deep insights into the internal semantics of epistemic indefinites. Thus this section also includes a short examination of the broader synchronic facts about—as well as brief remarks on the historical development of—the particles *də* and *hari* in Sinhala, and their relation to the analysis of epistemic indefinites more generally.

*Many thanks to Luis Alonso-Ovalle and Paula Menéndez-Benito for numerous helpful comments, corrections, and discussion. Thanks also to audiences at the 86th meeting of the Linguistics Society of America [LSA 86] (Portland, Oregon), the 2012 meeting of the Dallas/Fort Worth Metroplex Linguistics Conference (Arlington, Texas), the 3rd meeting of the conference on Formal Approaches to South Asian Languages [FASAL-3] (Los Angeles), and the University of Utah Department of Linguistics. All remaining errors and infelicities are, however, mine alone. All Sinhala and Hindi data reported here was collected by the author unless otherwise indicated.
2 Sinhala’s two epistemic indefinites

Sinhala employs two morphologically- and pragmatically-distinct epistemic indefinites: the first having the form of wh-word + the particle hari, as in (1-b); the second having the form wh-word + the particle də, as in (1-c); contrast these again the “plain” indefinite in (1-a).1

(1) a. sanat deyak gatta.
   Sanath thing.INDEF buy.PAST.A
b. sanat monəva hari gatta.
   Sanath what hari buy.PAST.A
c. sanat monəva də gatta.
   Sanath what də buy.PAST.A
   “Sanath bought something.”

The NP-type of (1-a) is pragmatically-unspecified/”neutral”, providing no information about whether or not the identity of the person in question is known; morphologically it involves a NP with an indefinite suffix (-ek). The two wh-based indefinites, however, involve different felicity conditions.

Consider the following context: the speaker observes a person dancing on a table—that is, the person is fully visible to the speaker—who is somehow otherwise unidentifiable by the speaker. In such a context the WH+də indefinite may be felicitously employed, while the WH+hari indefinite may not, as show in (2).

(2) a. Kau do mese uda natanava.
   Who do table on dance.PRES.
b. #Kauru hari mese uda natanava.
   Who hari table on dance.PRES.
   “Someone is dancing on the table.”

On the other hand, in the context where the speaker knows only the name of the individual in question (say, in the case of reporting second-hand information), WH+də is infelicitous, while WH+hari may be employed, as in (3).

(3) a. #Ranjit namin kau do Chitra.ta hambavunā.
   Ranjit named who do Chitra.DAT meet.PAST.
b. Ranjit namin kauru hari Chitra.ta hambavunā.
   Ranjit named who hari Chitra.DAT meet.PAST.
   “Chitra met someone named Ranjit.”

A reasonable first rough approximation of the felicity conditions for these two indefinites is that WH+hari is felicitous where the individual in question cannot be visually-identified (but may be identifiable in other ways) while WH+də is felicitous where no non-visual means of identifying an individual (such as naming) exists.

The conceptual cover approach of Aloni & Port (2010, forthcoming), proposed in the context of handling variation in the properties of epistemic indefinites crosslinguistically (specifically accounting for differences between English, German, Spanish, Italian), treats felicity conditions in a similar fashion, in that it refers to potential identification methods. More specifically, Aloni & Port (2010, forthcoming) propose that epistemic indefinites involve a pragmatic component signalling that the speaker cannot identify the individual in question in a way that is appropriate to the context—although the speaker may be able to identify this individual in other ways. Aloni & Port (2010, forthcoming) propose an account of the crosslinguistic distribution of epistemic indefinites couched in a

---

1 There is actually another indefinite involving wh-word + the particle vat, which appears only in NPI contexts. Since vat appears to simply be the NPI-counterpart of hari I do not provide separate discussion.
2 The alternation between kauru in (1-b) and kau in (1-c) is purely allomorphic: “who” appears as kauru unless immediately followed by də, in which case it appears as kau (see, amongst others, Fairbanks et al. 1968, Lalith Ananda 2008).
3 Alonso-Ovalle & Menéndez-Benito (2003) introduced this scenario, where they discuss the felicity of English and Spanish examples in this context.
theory of conceptual covers (Aloni 2001), suggesting three relevant methods of identification: definite description, naming, and ostension. Different contexts may require different types of identification. To take an example involving pragmatically-appropriate answers to questions, if a speaker asks (pointing) "Who is that man over there?", they are inviting an answer couched in terms of naming or definite description (i.e. "Oh, that’s John Jones" or “He’s the president of the company”), not an answer in terms of ostension; while a speaker at a linguistics conference who asks “Which one is Chomsky?” likely desires an ostensive answer. This, informally, is Aloni and Port’s proposal—see Aloni & Port (2010, forthcoming) for the formal technical implementation of the conceptual cover analysis.

On Aloni & Port’s analysis 2010, the English epistemic indefinite some NP signals that the individual in question cannot be identified by the speaker by the contextually-relevant means of identification (though other means may be available to the speaker). Thus (4) could felicitously be uttered by someone observing an unknown professor dancing on a table (that is, where the speaker could not identify the dancer by name), and (4) would also be felicitous in the context of hearing noises of feet dancing upon a table coming from an adjoining faculty meeting room (that is, where the speaker could not identify the dancer by ostension), while the Spanish sentence in (5) would be felicitous in the latter context but not in the former.

(4) Look! Some professor is dancing the lambada on the table. (Alonso-Ovalle & Menéndez-Benito 2003)

(5) ¡Mira! algún professor está bailando la lambada encima de la mesa
Look! algún professor is dancing the lambada on the table
(Alonso-Ovalle & Menéndez-Benito 2003)

Aloni & Port (2010) are concerned with correctly predicting crosslinguistic differences in the distribution of epistemic indefinites, focussing on differences between Germanic languages (exemplified in their discussion by German) and Romance languages (exemplified in their discussion by Italian). In order to capture pragmatic differences between the English epistemic indefinite some NP and the Italian epistemic indefinite un qualche NP, Aloni & Port (2010) suggest that there exists a universal ranking of identification methods—shown in (6)—and that in Italian (and other Romance languages) but not in English (or other Germanic languages) the contextually-required identification method must be higher in rank than the identification method by which the speaker can identify the individual in question in order for an epistemic indefinite to be felicitous.

(6) ostension > naming > description

Thus Aloni & Port (2010) can explain the differences in felicity in examples like the following pairs of English and Italian parallel sentences:

(7) Speaker can identify=ostension; target=naming
   a. Look! Some footballer just got injured. Do you know who he is?
   b. ?Guarda! Un qualche giocatore si è fatto male. Sai chi è?

(8) Speaker can identify=ostension/naming; target=description
   a. There is some pharmaceutical rep here to meet you. His name is Schulz. Should I show him in?
   b. ?C’è qui un qualche rappresentante farmaceutico per te. Si chiama Schulz. Posso farlo entrare?

(9) Speaker can identify=description; target=naming
   a. I have to meet some professor. He’s the Department Chair, but I don’t know his name.
   b. Devo incontrare un qualche professore. È il capo del dipartimento, ma non so come si chiama.

(10) Speaker can identify=naming; target=ostension
   a. I have to meet some professor. His name is John Smith, but I don’t know what he looks like.
   b. Devo incontrare un qualche professore. Si chiama John Smith, ma non so che aspetto abbia.
That is, the English examples are all felicitous because the speaker cannot identify the individual in question by a contextually-appropriate method. While the same is true in the Italian contexts, only (9) and (10) are felicitous since only in those two examples does the required relationship hold between the known-to-speaker and contextually-demanded identification methods.

At first blush this approach would seem extendable to the Sinhala examples, given that “visually-identifiable” seems to be something like identification by ostension. However, there is a crucial distinction between the behaviour of the Sinhala epistemic indefinites and the Italian epistemic indefinites: Italian un qualche, according to Aloni & Port (2010), does not specify any particular method of identification. Rather un qualche NP simply signals that the speaker cannot identify the individual in question by the contextually-relevant means of identification and that additionally any method by which the speaker can identify the individual in question is lower in the ranking than the contextually-relevant method of identification. Sinhala epistemic indefinites on the other hand appear to be tied to the method of identification (and make no reference to hierarchical differences between known-to-speaker and contextually-targeted identification methods).

In order to get a more fine-grained sense of the felicity conditions for WH+do and WH+hari, (12) provides instances of variety of contexts in where the two epistemic indefinites may or may not be employed. Consider again the “dancing on the table” examples in Sinhala, shown with the two epistemic indefinites from (2), repeated below as (11), and the possible contexts in which they may used in (12).

(11) a. Kau do mese uda natanava.  
   Who do table on dance.pres.
   "Someone is dancing on the table."

b. Kauru hari mese uda natanava.  
   Who hari table on dance.pres.
   "Someone is dancing on the table."

(12) a. Speaker is sitting in a boring meeting with his friend Chitra. Suddenly some other person at the meeting gets up on the table and starts to dance. The person in question is not known to the speaker. **Only (11-a) is possible**

b. Speaker is sitting in a boring meeting with his friend Chitra. Suddenly Gunapala, who is friends with the speaker and Chitra, gets up on the table and starts to dance. **Neither (11-a) nor (11-b) is possible.**

c. Speaker is sitting in a boring meeting with his friend Chitra. Suddenly the chair of the meeting—who is unknown to the speaker, but who the speaker knows is the Dean of the School of Liberal Arts & Sciences—gets up on the table and starts to dance. **Neither (11-a) nor (11-b) is possible.**

d. Speaker is sitting his office with his friend Chitra. They hear loud noises coming from the next room that sound like rapid footsteps on a raised surface. The speaker knows that the people in the neighbouring office like to get up on the table and dance (one at a time) now and then. Chitra asks what the noise is. **(11-b) is possible. Some speakers allow for (11-a), but (11-b) is generally preferred.**

Where the individual in question may be identified by the speaker by ostension, as in (12-a)–(12-c), the WH+hari indefinite is infelicitous. Where the person in question may be identified by the speaker either by name or description, the WH+do indefinite is infelicitous, as shown by (12-b) and (12-c). This conforms more closely to the first approximation, namely that WH+hari requires that speaker cannot identify the individual in question by naming or description, and WH+do requires that the speaker cannot identify by ostension.

Where there is no possibility of visual identification or identification by other means, as in (12-d), then the WH+hari indefinite is preferred over the WH+do indefinite. That is, WH+hari seems to be strongly associated with lack of visual identifiability.

Again, this makes the Sinhala epistemic indefinites descriptively unlike both Germanic and Romance epistemic indefinites. The Sinhala epistemic indefinites are not, like the Germanic, simply signalling that the contextually-relevant identification method is unavailable to the speaker; nor do they, like the Romance, make reference to any specific ordering relationships between identification methods. Rather, the Sinhala epistemic
indefinites appear to be tied directly to availability of particular identification methods. Further, it is not entirely clear whether “ostension” is even the appropriate label for the relevant identification method, given speakers’ judgements regarding contexts like (13).

(13) The speaker is walking down a long hallway with his friend Chitra. At the far end of the hallway there is an open door. Through the open door the speaker can make out the shape of a humanoid figure dancing on a table, but can not see the figure clearly—the speaker cannot, for instance, even determine whether the person is a male or female, or make out any distinguishing features.

In contexts like (13), speakers report that either (11-a) or (11-b) may be employed. This suggests that “ostension” is not entirely an ideal term for describing the constraints on WH+hari. WH+hari thus seems to be infelicitous where the referent can be visually-identified, but not in all contexts where the referent could be identified by ostension (since a far-away dancer could still be identified by ostension).

English too provides additional data which do not conform to the “Germanic” or “Romance” patterns discussed by Aloni & Port (2010). That is, while English some NP fits the pattern, English some NP or other behaves rather differently: some NP or other appears to be infelicitous in contexts where the speaker can identify the referent by name or ostension, as shown by the contrast between (14) and (15).

(14) a. Based on the noises coming from next-door, some man is dancing on the table again.  
   b. Look there! Some man is dancing on the table!  
   c. I heard that some man named Joe Bloggs is dancing on the table.

(15) a. Based on the noises coming from next-door, some man or other is dancing on the table again.  
   b. #Look there! Some man or other is dancing on the table!  
   c. #I heard that some man or other named Joe Bloggs is dancing on the table.

In contexts where the individual is not identifiable by the speaker by either name or ostension either some NP or some NP or other are felicitous, as shown by (14-a) and (15-a). Where the individual in question is identifiable by ostension, some NP is possible, as in (14-b) by some NP or other is not, as shown by (15-b). Where the speaker can identify the individual in question by name again some NP is felicitous, as (14-c), but some NP or other is at best highly degraded, as shown by (15-c).

Here too, as for the Sinhala examples, it is not clear that visually-identifiable is the same thing as identifiable-by-ostension, given that while (15-b) is infelicitous where the speaker is sitting in the same room as the (unknown) dancing man, in the case that the dancing man is far enough way that the speaker cannot see enough of his features to identify him at close range, (15-b) becomes felicitous. Alonso-Ovalle & Menéndez-Benito (2012) point to similar data in Spanish also suggesting that identification by ostension may not be the relevant factor.

Again, as found for the Sinhala epistemic indefinites, English some NP or other appears to differ from Romance epistemic indefinites in being constrained not by a hierarchy of ranked identification methods, but rather by explicitly ruling out specific identification methods as being accessible to the speaker. Preliminary empirical research based on Google web-searches confirms the judgements given in (14) and (15) [searches performed on 13 April 2012].

(16) “some X named”
   a. "some guy named": about 12,000,000 results
   b. "some man named": about 1,500,000 results
   c. "some bloke named": about 84,600 results

(17) "some man or other": 325 results

4Though one speaker indicated a slight preference for (11-a) in this context, another speaker explicitly explained that in the case of the individual being too far away to see clearly either WH+hari or WH+da could be used.
5See Slade(this volume) for further discussion of English some NP or other.
6At least if interpreted as signalling ignorance rather than indifference. Some NP or other appears to often function to signal "feigned" indifference, as discussed below.
Here the sole result for the search "some man or other called" is an example in Becker (1999), where, in fact, the example is meant to be synonymous with "some man or other phoned", and so is irrelevant.

The only results for the collocation some guy or other called/named appear to be (feigned) indifference readings (however they are to be interpreted, it is clear that they are not sincere speaker-ignorance readings), as in the representative examples reproduced here. (In this and all following examples the expression of interest is **bolded**; this emphasis is not present in any of the original texts.)

The collocation some bloke or other, unsurprisingly, produces results similar to those for some guy or other, as shown by the representative example below.

These results confirm that the some NP or other construction is infelicitous where the speaker can identify the referent by name.

Likewise, while a search for "look there some guy" turns up results like:

Searches for "look there some guy or other", "look there some bloke or other", and "look there some man or other" all turn up 0 results, confirming the intuition that the some NP or other construction is infelicitous where the referent is visually-identifiable by the speaker.

It is interesting to note, however, that someone or other has somewhat different properties than some NP or other, as a search for "someone or other named" turned up 45 results. Some of these results are of the feigned indifference type observed for some guy/bloke or other named/called:
(25) “It sometimes seems as if Joe Henry has spent most of his nearly two-decade career standing just out
of the frame in other people’s glamour shots. He is, of course, the brother-in-law of someone or other
named Madonna...” [http://archives.nodepression.com/2006/01/new-soul]

But at least one result appears to require a “sincere ignorance” interpretation:

(26) “Anyway, speaking of paintings, there were paintings hanging in the Philly airport painted by
someone or other named Miho Chung, a name that kind of flipped me out (“what? she’s Japanese and

In summary, Aloni & Port (2010, forthcoming) characterise epistemic indefinites as signalling that the speaker
cannot identify the individual in question in the contextually-relevant way, and develop an analysis aimed at
accounting for differences in the felicity conditions of epistemic indefinites crosslinguistically, particularly between
Germanic and Romance epistemic indefinites. They characterise the stricter conditions on Romance epistemic
indefinites in terms of a posited universal ordering of identification methods:— ostension > naming > description,
suggesting that in Romance the contextually-relevant identification method must be higher in order than any
identification method available to the speaker. Though thinking about epistemic indefinites in Sinhala and in
English in terms of identification methods is useful, it appears that ostension does not provide a good
characterisation of the felicity conditions. Further, the two Sinhala epistemic indefinites together with English some
NP or other also exemplify a third type of epistemic indefinite alongside of what Aloni & Port (2010) characterise as
“Germanic” and “Romance”, namely epistemic which are associated with particular identification methods.

The next section examines another property of epistemic indefinites: their relation to specificity.

3 Epistemic indefinites and specificity

Epistemic indefinites show crosslinguistic differences (and intralanguage differences, in the case of languages like
English and Sinhala which use multiple epistemic indefinites) in their relation to specificity.

In modern English, for instance, some NP is compatible with contexts in which the indefinite receives a
non-specific interpretation, as shown by (27).

(27) John wants to kiss some girl—he doesn’t care who.

Contrast (27) with the infelicitous (28)

(28) #John wants to kiss some girl or other—he doesn’t care who.

Example (28) carries a pragmatic signal that the speaker does not know the identity of the girl who John wants to
kiss, but is infelicitous since it is used in a context where there is not in fact a PARTICULAR girl that John wants to
kiss. This indicates that in fact the English construction some NP or other—in contrast to the some NP
construction—obligatorily assumes that there is a specific referent, which, however, the speaker has not means of
uniquely identifying.\footnote{Some NP or other can be non-specific in the scope of non-downward entailing quantificational elements, e.g. in examples like (i),}

This contrast also seems to be borne out by the results of Google searches. A search for "don’t care which|who"
"some boy|girl|man|woman|bird|chick|chap|fellow|guy|bloke" turns up about 500 hits, like that shown in (29).

\footnote{i Every boy wants to kiss some girl or other.

the girls can vary with respect to boys.}
"I'd love to have her get chased by some guy — I don't care who — and for her to turn him down."

http://www.tumblr.com/tagged/glee?before=1301456283

Though some of the hits are spurious in the sense that "don't care which/who" does not qualify "some boy/girl/man/woman/bird/chick/chap/fellow/guy/bloke", a similar search for a some NP or other type phrase correlating with I don't care who/which produces vastly different results: a search for "don't care which|who" some boy|girl|man|woman|bird|chick|chap|fellow|guy|bloke or other turns up only results. Four of these results are irrelevant as "don't care which|who" does not qualify the some NP or other phrase.

The two Sinhala epistemic indefinites discussed here also differ from each other in terms of their specificity requirements. Consider example (30) below.


Every boy saw every girl kiss someone.

Here the only available interpretation of (30-b), which employs a WH+do indefinite, is that there is a specific (though unknown) person x such that every boy saw every girl kiss x. The other logically possible interpretations—i.e. that each boy saw all of the girls kissing some (particular) person x, where x may vary for each boy; or that each boy saw every girl kiss some person x, where x may vary for each girl—are not available.

These two latter interpretations are available for (30-a), which uses a WH+hari indefinite rather than a WH+do indefinite. However, the only reading of (30-b)—where someone out-scopes the other quantifiers—appears, for some reason, to be unavailable for (30-a).

That WH+hari indefinites can be non-specific is also apparent in the following example:

| maṭə kāva hari iṁbinə onə. |
| I DAT who ACC hari kiss INF want PRES |
| "I want to kiss someone(. anyone)."

That is, both Sinhala WH+do and English some NP or other appear to be obligatorily specific indefinites which can only be used in contexts where the speaker believes that there is some unique (but unknown) individual fulfilling the existential requirements of the utterance.

4 Conclusions and avenues for future inquiry

Sinhala epistemic indefinites thus differ from the European epistemic indefinites examined in Aloni & Port (2010) in terms of their felicity conditions, summarised in Table 1.

In terms of their internal morphological structure, Sinhala epistemic indefinites are also markedly different from their closest European counterparts. And, I suggest, it is important to pay attention to morphosyntactic differences between epistemic indefinites crosslinguistically—especially in case of epistemic indefinites of the Sinhala-type,

6 The only non-spurious hit is actually from Frazier & Bader (2007) in the phrase "some book or other by Chomsky, I don't care which" which is used (I would argue, perhaps inappropriately) to gloss the German irgendein Buch von Chomsky.

7 It is not entirely clear to me why this should be the case, given that WH+hari indefinites do not seem to generally be obligatorily non-specific, as shown by examples like (3-b), repeated below as (i).

(i) Ranjit namin kauru hari Chitra.ṭa hambavunā. Ranjit named who hari Chitra DAT meet PAST. "Chitra met someone named Ranjit."

8 The English construction some NP or other is necessarily specific in particular environments, but can be non-specific in other environments see fn. 7.
where the morphological components appear in a variety of other syntactic environments—when considering potential formalisations of epistemic indefinites. Thus, while I do not provide a formal analysis for epistemic in this study (see Slade 2011, Slade in progress for explicit formal proposals), I argue that any attempt at a proper formal analysis of this type of epistemic indefinite must include a broader view of the network of contexts in which particles like Sinhala də or Japanese ka appear (see also Slade 2011, Szabolcsi 2013). This section examines the morphosyntax of epistemic indefinites in selected South Asian languages, with brief remarks on the relationship between the morphological make-up of epistemic indefinites and the formal semantic analysis of epistemic indefinites.

In Sinhala, as remarked above, epistemic indefinites are formed by the addition of a particle (də or hari) to a wh-word, with close parallels being observed in Japanese. Sinhala epistemic indefinites are similar, in fact, in their morphosyntactic structure not only to those of Japanese, but also to epistemic indefinites in the geographically-proximate Dravidian languages, like Malayalam. In Malayalam epistemic indefinites are formed from a wh-word combined with a particle (-o), as shown in (32). Similar to the situation in Sinhala, a “plain”, pragmatically-unmarked indefinite, formed from a NP preceded by “one”, is available, as in example (33).

<table>
<thead>
<tr>
<th>Felicitous Ident. Methods</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>WH+hari</td>
<td>visually-identifiable</td>
</tr>
<tr>
<td>WH+də</td>
<td>identifiable by non-visual means</td>
</tr>
</tbody>
</table>

Table 1: Properties of Sinhala epistemic indefinites

11) However, Japanese is unusual in that both “plain” and epistemic indefinites are formed in this fashion, where the distinguishing property appears to be the relative position of the wh-word and the particle. Japanese wh-words like dare “who” and nani “what” are similar to wh-words in Sinhala in that they are actually “indeterminate pronouns”, sometimes functioning as interrogative pronouns (i-a), sometimes as indefinite pronouns (ii-b). Which function the indeterminate pronoun takes on depends on the positioning of the “question-particle” [Q] ka: when ka appears adjacent to the wh-word, it functions as an indefinite, when the particle appears clause-finally, the wh-word functions as an interrogative pronoun. (See Kuroda (1965); Nishigauchi (1991); Hagstrom (1998); Slade (2011), amongst others, for further discussion. Cf. Ramchand (1997), Kratzer & Shimoyama (2002)).

(i) a. John-ga nani-o kaimasita-ka?
   John-NOM what-ACC bought.POL-Q
   “What did John buy?” [Hagstrom 1998: 36]

   John-NOM what-Q-ACC bought
   “John bought something”

Where ka appears immediately adjacent to the wh-word, it functions as plain, non-epistemic indefinite. However, ka may, in certain structures, appear close to, but not immediately adjacent to, the wh-word, in which case it behaves as an epistemic indefinite; compare (ii-a), where ka is immediately adjacent to dare “who”, with (ii-b), where ka appears following the postposition kara “from”.

(ii) a. [Dare-ka-kara] henna tegami-ga todoi-ta.
   [who-ka-from] strange letter-NOM arrived.
   “A strange letter arrived from somebody.” (Nishigauchi 1990: 221)

   b. [Dare-kara-ka] henna tegami-ga todoi-ta.
   “A strange letter arrived from god knows who”(ibid.)

The same pattern emerges where ka appears after the postpositional genitive no:

(iii) a. [Dare-ka-no hahasoya]-ga paatii-ni kita.
   [who-ka-gen mother]-NOM party-loc came
   “Someone’s mother came to the party.” (Yatsushiro 2001: 12)

   b. [Dare-no-ka hahasoya]-ga paatii-ni kita.
   [who-gen-ka mother]-NOM party-loc came
   “Someone (or other)’s mother came to the party.” (ibid.)

For remarks on other complexities of Japanese epistemic indefinites, see further Alonso-Ovalle & Shimoyama (2013).
‘I touched somebody in the dark.’ (Jayaseelan 2001: 66)

(33) ṇāṉ innale oru āḷ-e paricayappetṭu
I yesterday one person-ACC met
‘I met someone yesterday’ (Lit. ‘I met a person yesterday.’) (Jayaseelan 2001: 66)

An investigation of the felicity conditions of Malayalam epistemic indefinites in more fine-grained detail remains to be done.

Yet a different morphosyntactic means of producing epistemic indefinites is through reduplication, as is found in Hindi, another South Asian language. In Hindi every indefinite pronoun and determiner X has an epistemic counterpart X-na-X; e.g. koi “someone” and koī-na-koī “someone or other”, kuch NP “some NP” and kuch-na-kuch NP “some NP or other”. The X-na-X epistemic indefinite are like English some NP or other in that they appear to only be felicitous where the referent cannot be uniquely identified by any means: compare the infelicitous (34) against (36); whereas “plain” indefinites like koi “someone” have no such restriction, see (35).

(34) #Dekho! Koi-na-koī mez par nāc rahā hai.
    Look! Someone-not-someone table on dance CONT is
    “Look! Someone-or-other is dancing on the table.”

(35) Dekho! Koi mez par nāc rahā hai.
    Look! Someone table on dance CONT is
    “Look! Someone is dancing on the table.”

(36) Suno! Koī-na-koī mez par nāc rahā hogā.
    Listen! Someone-not-someone table on dance CONT will-be
    “Listen! Someone-or-other must be dancing on the table.”

Unlike English some NP or other, which it resembles in terms of its epistemic felicity conditions, the Hindi X-na-X indefinites do not appear to be obligatorily specific, though they may bear this interpretation:

(37) Rām kisi-na-kisi-ko cummā cahtā hai.
    Ram someone-not-someone-DAT kiss.INF want AUX
    “Ram wants to kiss someone or other. (but I don’t know who)” OR
    “Ram wants to kiss someone, anyone.”

While epistemic indefinites like French je ne sais qu- NP, as in (38) below have fairly transparent origins (though the precise processes leading these being reanalysed as epistemic indefinites are still of some interest), the use of reduplication as in Hindi or the collocation of a wh-word with a particle as in Sinhala in the production of epistemic indefinites is somewhat more opaque, and the relationship between these morphological processes/structures and epistemicity thus would be productively examined when formulating a formal semantic analysis.

(38) Il nous reste encore je ne sais quel désir vague, je ne sais quelle inquiétude.
    It us remains still 1 NEG know which desire vague, 1 NEG know which restlessness.
    “We are left with some kind of vague desire, some kind of restlessness.” (Voltaire; cited in Haspelmath 1997: 133)

The relationship of particles like Sinhala də to epistemicity is the more perplexing given that the particles which appear in all of these WH-based indefinites have much wider distributions, appearing in a number of non-indefinite context. This is true not only of Sinhala də and hari, but also of Malayalam -oo and Japanese -ka, and appears to be a more general feature of languages employing this type of particle (see further Slade 2011, Szabolcsi 2013).

In fact, the use of də and hari in the formation of epistemic indefinites is a relatively recent development in the history of Sinhala. Də originates in the formation of alternative questions, and hari in the formation of
(non-interrogative) disjunctions. Examples of do and hari in the other environments in which they appear in modern colloquial Sinhala are given below in (40)–(43).

The particle do (< earlier Sinhala da) is found (obligatorily) in a wide variety of syntactic contexts in modern colloquial Sinhala. Thus, not only does do appear in the formation of epistemic indefinites, but wh-questions in Sinhala also employ this Q-particle do, as do alternative questions: any serious account of epistemic indefinites like Sinhala WH+do must provide an analysis of do which is compatible with all of the environments in which it appears. Compare the declarative in (39) with the corresponding interrogative in (40).12

\begin{align*}
\text{(39) } & \text{Chitra potə gatta} \\
& \text{Chitra book bought.}_A \\
& \text{“Chitra bought the book.”} \quad \text{[Colloquial Sinhala]} \\
\text{(40) } & \text{Chitra monəwa do gatte} \\
& \text{Chitra what } do \text{ bought.}_E \\
& \text{“What did Chitra buy?”} \quad \text{[Colloquial Sinhala]} \\
\end{align*}

In yes/no-questions, do also appears obligatorily, normally in clause-final position, as in example (41).13

\begin{align*}
\text{(41) } & \text{Chitra ee potə kieuwa do?} \\
& \text{Chitra that book read-}_A \text{ do } \\
& \text{“Did Chitra read that book?” (Kishimoto 2005: 11)} \quad \text{[Colloquial Sinhala]} \\
\end{align*}

The particle do may also appear after a constituent smaller than IP—in which case it marks that constituent as focussed (42).14

\begin{align*}
\text{(42) } & \text{Chitra ee potə do kieuwe?} \\
& \text{Chitra that book do read-}_E \\
& \text{“Was it that book which Chitra read?” (Ibid.)} \quad \text{[Colloquial Sinhala]} \\
\end{align*}

Finally, do is also involved crucially in the formation of interrogative disjunctions (i.e. alternative questions), appearing obligatory after each of the disjuncts, as in example (43).

\begin{align*}
\text{(43) } & \text{Gunəpālə do Chitra do Ranjit do gaməṭə go past.}_E \\
& \text{Gunapala do Chitra do Ranjit do village.dat go.past.}_E \\
& \text{“Was it Gunapala or Chitra or Ranjit who went to the village?”} \quad \text{[Colloquial Sinhala]} \\
\end{align*}

In addition to appearing in the formation of epistemic indefinites, hari also appears in the formation of declarative disjunctions:

\begin{align*}
\text{(44) } & \text{Gunapala hari Chitra hari gaməṭə go past.}_A \\
& \text{Gunapala hari Chitra hari village.dat go.past.}_A \\
& \text{‘Gunapala or Chitra went to the village.’} \\
\end{align*}

These particles in Sinhala show significant changes in their distribution diachronically. Table 2 provides an overview of the distributional spread of the particles do and hari in Sinhala in the different syntactic environments in which they appear over time. Given the wide range of contexts in which do and hari appear, and given that they

---

12The verb over which the wh-word takes scope appears with the special “focussing” -e ending (following Kishimoto 2005, I refer to this as the -e ending, glossed as -E), distinguished from the neutral ending (the -a ending, glossed as -A). For further discussion of “focussing” and “neutral” forms of the verb in Sinhala, see Gair (1983[1998]), Gair (1986)[1998][b], Slade (2011).

13Note that when do appears in the default clause-final position, the matrix verb appears with the “neutral” -a inflection.

14In case of do occurring inside of the c-command domain of the verb, the verb appears in the “focussing” -e form.
only acquire the ability to form epistemic indefinites relatively late, it would seem likely that their function of forming epistemic indefinites is not unrelated to their functions in other contexts.\textsuperscript{15}

\begin{tabular}{|l|c|c|c|c|c|c|}
\hline
 & Pre-Sinh 1 & Pre-Sinh 2 & Old Sinh & Class Sin & M Lit Sinh & M Collq Sinh \\
\hline\hline
\textit{Da/Do} & & & & & & \\
alt-questions & ✓ & ✓ & ✓ & ✓ & ✓ & ✓ \\
y/n-questions & (✓) & (✓) & (✓) & ✓ & ✓ \\
\textit{Ho/Hari} & & & & & & \\
indefinites & & & & ✓ & ✓ & \\
declarative disjunctions & ✓ & ✓ & ✓ & ✓ & ✓ & ✓ \\
\hline
\end{tabular}

Table 2: Spread of particles \textit{da/da, ho/hari} in Sinhala\textsuperscript{16}

The other devices used to form epistemic indefinites in languages like English, Spanish, Italian, German (determiners) and Hindi (reduplication) are formally rather different, and observe very different syntactic distributions. Further investigation into the historical development of epistemic indefinites is thus a promising avenue of investigation for understanding the nature of WH+particle based epistemic indefinites, especially in languages like Sinhala and Malayalam where the morphosyntax of epistemic indefinites intimately connects them to a variety of other constructions.

The variety of morphosyntactic forms of epistemic indefinites crosslinguistically—especially in light of the close structural similarities of the morphological components of epistemic indefinites in languages like Sinhala, Malayalam, and Japanese—thus raises the question of whether we should expect there to exist a unified theory of epistemic indefinites, or whether rather different formal mechanisms may be involved in different languages. Further investigation of epistemic indefinites in non-European languages like Sinhala, Hindi, Japanese, and Malayalam will be important in addressing questions of this sort.

The full range of semantic/pragmatic differences between epistemic indefinites also remains to be determined. Looking across epistemic indefinites in the small set of languages in which there has been any substantial investigation into their distribution, at this point we observe epistemic indefinites (1) which simply signal that the speaker cannot identify the individual in question by the contextually-appropriate identification method (leaving open the possibility that the speaker can identify the individual in question by some other identification method); (2) those which signal that the speaker cannot identify the individual by the contextually-appropriate identification method AND further constrain which identification methods may be accessible to the speaker (the Romance cases examined by Aloni \& Port \textit{2010}); and (3) those which signal that the speaker cannot identify the individual by some particular identification method (\textit{e.g.} description), as appears to be the case for Sinhala WH+\textit{da} and WH+\textit{hari}, as well as for English \textit{some NP or other}. As has been observed in this paper, there are also secondary differences between epistemic indefinites involving their relationship with specificity. An examination of a larger set of languages is necessary in order to determine the not only the range of morphosyntactic devices for forming epistemic indefinites, but also the semantic/pragmatic typology of epistemic indefinites in terms of properties like accessible identification methods and specificity. As well, as indicated in this section, formulating a reasonable formal semantic analysis of epistemic indefinites also requires paying attention to the morphological make-up of such indefinites—especially in languages like Sinhala, Malayalam, or Japanese where the morphological

\textsuperscript{15}Slade (2011) and Slade (in progress) formulate a choice-functional analysis of Sinhala \textit{da}, \textit{hari} (as well as Malayalam -oo and Japanese \textit{ka}) which attempts to provide a coherent, unified account of the wide-range of contexts in which these particles occur, and of the sets of properties they possess.

\textsuperscript{16}Pre-Sinh \{1, 2\} = stages of pre-Sinhala, reconstructed on the basis of Old Sinhala \& Pāli. Old Sinh = Old Sinhala, represented by the graffiti texts on the Mirror Wall at Sihigiri (ca. 8th–10th c. A.D., see Paranavitana \textit{1956}); Class Sinh = Classical Sinhala [CS], represented largely by translations and commentaries on Pāli Buddhist texts (ca. 12th–15th c. A.D., see Wijemanne \textit{1984}); M Lit Sinh = Modern Literary Sinhala [LS], which differs from Classical Sinhala, but retains a number of archaism such as overt subject-verb agreement morphology; and M Collq Sinh = Modern Colloquial Sinhala [MCS]. The latter two varieties co-exist in a diglossic relationship, with the literary variety being employed in written and formal situations, but the general archaic nature of the literary variety justifies its treatment as representing an earlier variety than does the colloquial—on Sinhala diglossia, see further Gair (1968[1998], 1986[1998]a) and Paolillo (1992).
components of epistemic indefinites participate systematically in the formation of a variety of other structures. Likewise, our overall understanding of epistemic indefinites can be furthered by an examination of the historical development of epistemic indefinites, and changes in the distribution of the morphological components of epistemic indefinites.

References


Aloni, Maria & Angelika Port. 2010. Epistemic indefinites crosslinguistically. NELS 41.


Slade, Benjamin. in progress. Indefinite, unknown. Ms., University of Texas, Arlington.

Szabolcsi, Anna. 2013. KA and MO as presupposition inducers. Ms, New York University.
