



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2009

Deixis and information structure in Mon: the multifunctional particle kəh

Jenny, Mathias

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-21060>
Journal Article

Originally published at:

Jenny, Mathias (2009). Deixis and information structure in Mon: the multifunctional particle kəh. *Journal of the Southeast Asian Linguistics Society*, 2:53-72.

DEIXIS AND INFORMATION STRUCTURE IN MON – THE MULTIFUNCTIONAL PARTICLE *kəh*

Mathias Jenny
University of Zurich
<jenny@spw.uzh.ch>

1. Abstract

The topic of this study is a multifunctional particle in Mon, viz. *kəh*. This morpheme is found in all stages of Mon since the 11th century, where it appears as a deictic noun phrase marker indicating distal spatial and temporal deixis. The use and development of *kəh* exhibits a number of semantic and pragmatic features that make it difficult to define its actual function(s). The aim of this paper is to present data from Old, Middle and Modern Mon and try to find explanations of the development paths and uses of this very frequent particle.

In the next section, examples from all recorded stages of Mon will be presented, followed by previous studies and definitions of the morpheme under discussion. Section 3 is concerned with demonstratives in general and the position of *kəh* in the system of demonstratives in Mon. Section 4 considers the possibility of explaining *kəh* as a marker of definiteness in Modern Mon. Section 5 takes a look at the pragmatic functions of *kəh*, i.e. as a marker of topic or identifiable information.

2. History and previous studies

2.1 The development of deixis in Mon

Old Mon seems to have had a two way distinction in its deictic system. Proximal deixis was expressed by *woʔ*, distal by *goʔ*. The exophoric use of *woʔ* is clearly present in expressions like (1), describing the scene in an accompanying picture.

- (1) Old Mon (Ananda plaques)
woʔ brumha ku ʔin.
this Brahma COM Indra
These are Brahma and Indra.

The originally probably distal deictic *goʔ* is used mainly anaphorically, either as noun phrase marker or in adverbial expressions such as *gun goʔ* ‘therefore, by virtue of that’ and *blah goʔ* ‘afterwards, having finished that’. The different uses of *woʔ* and *goʔ* are seen clearly in the frequent expression *row woʔ* ‘thus, like this’ to cataphorically introduce direct speech, while *row goʔ* anaphorically follows the quotation. This use is illustrated in sentence (2), where the first instance of *woʔ* is cataphoric, while it is exophoric in the reported speech. This example of exophoric *woʔ* also shows that it is not primarily speaker-centred, but rather spatial proximal.

(2) Old Mon (Shweizigon inscription)

smāñ row **wo?** ma tirla pa k<ir>?im **wo?** ci
 ask manner this ATTR master do <NML>smile this EMPH

mu het yo? row **goḥ** tarley ?ānan smāñ da.¹
 what reason Q manner that master:1s Ananda ask FOC

*He asked like this: What is the reason for this smile you had on your face?
 Like that our lord Ananda asked.*

The deictic system of Old Mon is summarised in table 1.

Table 1: Demonstratives of Old Mon

Form	Exophoric uses	Discourse functions
<i>wo?</i>	proximal	cataphoric
<i>goḥ</i>	?	anaphoric

By Middle Mon, the deictic system was expanded by two new elements, the distal *gah*, i.e. the MM spelling of OM *goḥ*, now competing with *te?* and the proximal *wo?*, variously spelt *wo?*, *wwo?*, *wwa?*, *wa?*, *wwá?* in MM, being replaced by the adverbial *?ano?* ‘here’ in some contexts. The exact use of these new deictics is not known and their occurrence is not very frequent in Middle Mon texts.

It is not certain whether the forms *te?* and *?ano?* were in fact absent from Old Mon or whether their absence from the documents is due to the incompleteness of the data. They may have belonged to a less formal register of the language and were therefore not used in literary documents. It should be noted that in closely related Nyah Kur, widely believed to be a direct offshoot of Old Mon in Thailand, there are the forms *te?*, *te?* and *tee?*, marking three degrees of distal deixis (L.Thongkham 1984:173f). The proximal *?ano?* has good cognates in other Austroasiatic languages (cf. Shorto 2006:90f) and appears in Nyah Kur in the expression *pətam nɔ?* ‘tonight’ (Diffloth 1984:147).

In MM *wo?* is still used exophorically as proximal deictic and cataphorically to introduce direct speech, and *gah* anaphorically closes quotations. In addition, *gah* is used increasingly to mark NPs in prominent position, e.g. fronted objects. As Burmese influence becomes stronger during the Middle Mon period around the 14th century, the frequent fronting of objects, originally probably for pragmatic reasons, can be seen as influence of Burmese OV structures. This is illustrated in (3).

¹ In transliteration of Old and Middle Mon, I basically follow the common spelling conventions as used in Shorto 1971, with the following two exceptions: For the glottal stop ʔ I use the IPA symbol ʔ in all positions and for the Mon-Burmese digraphic vowel symbol ʔ̃, I use the transliteration <iu> as suggested by Yanson (2002), which corresponds to the ordering of the symbols in indigenous spelling in both Mon and Burmese (in spite of the tradition among Western (and consequently also Burmese) scholars to use <ui>). A still better solution would be to use a separate symbol in transliteration altogether (e.g. <ə>, which is probably close to the intended sound in Old Burmese and Old Mon), similar to the use of <o>, which is made up of <e> and <ā> in Indic scripts.

- (3) Middle Mon (Shwedagon inscription)
 na ʔidhi **gah** dhāt swok **gah** ket ket tuy.
 INSTR power that relic hair that take TAKE FINISH
With that power he took the hair relic by himself.

That *gah* was not grammaticalised as a marker of objects is shown in sentence (4), where it is the subject that receives the marker.

- (4) Middle Mon (Shwedagon inscription)
 tapussa **gah** goʔ sotāpatiphiuw ra.
 Tapussa that get conversion FOC
Tapussa achieved religious conversion.

Another frequent use of *gah* in Middle Mon is at the end of adverbial expressions, similar to Old Mon usage, functioning as a kind of phrase boundary marker, as in sentences (5) and (6), both formally relative clauses (both from the Shwedagon inscription).

- (5) het ñah ma ha goʔ liñwor pūjau **gah** kium ṇa.
 reason person ATTR NEG get pay.homage worship that ADD FOC
It was because they could not worship and pay homage [to the relics].

- (6) gamī truh gamī brau sāmaṇī truh sāmaṇī brau khā
 monk male monk female novice male novice female time

 ma nwom tuy **gah**, khā **gah** ...
 ATTR exist FINISH that time that
When there are male and female monks, male and female novices, at that time...

The deictics of Middle Mon are summarised in table 2.

Table 2: *Demonstratives of Middle Mon*

Form	Exophoric use	Discourse function
<i>ʔanoʔ</i>	‘here’	?
<i>woʔ</i>	proximal	cataphoric
<i>gah</i>	?	anaphoric
<i>teʔ</i>	distal	?

In modern spoken Mon, a three way distinction has emerged with *kəh* functioning as medial deictic. The form *wəʔ* from Middle Mon *woʔ* is used only in literary and formal style where it is apparently freely interchangeable with *noʔ*. It is replaced by *noʔ* in the spoken language. The exophoric uses of *noʔ*, *kəh* and *teʔ* are rather straight forward, the choice being based on relative distance from the point of reference. The cataphoric-anaphoric distinction made in Old and Middle Mon is no longer followed in the spoken language, though there are traces of it in literary Mon. Sentences (7a-c) illustrate the exophoric deictic use, with the object placed at an increasing distance from the point of

reference. Sentence (8), uttered after the speaker finished telling a story, shows the anaphoric use of *nɔʔ*.

(7) a. ket ʃɛʔ **nɔʔ**.²
 take REF PROX
Take this one.

b. ket ʃɛʔ **kɔh**.
 take REF MEDL
Take that one.

c. ket ʃɛʔ **trʔ**.
 take REF DIST
Take that one over there.

(8) ʔəkhak **nɔʔ** raʔ ləə kɔʔ raʔ.
 manner PROX FOC tell GET FOC
That's how I can tell [stories].

While *nɔʔ* and *trʔ* behave like typical demonstratives (see section 3 below), *kɔh* has a number of features and functions that differentiate it from its proximal and distal counterparts. In any given text in Mon, *kɔh* is easily the most frequent word. It occurs in a large number of sentences, often more than once in a single sentence. Among the functions of *kɔh* are the marking of noun phrases as in (9), including complex NPs as in (10) with an unmarked relative clause, and the marking of clauses as complements (11) or adverbial clauses (12).

(9) pɔ mɔŋ ŋəə **kɔh** ɗəə pələŋ **kɔh**.
 watch STAY frog MEDL LOC bottle MEDL
They were watching the frog in the bottle.

(10) kɔʔ tɛh ʃɛʔ kon ŋəə həkaoʔ klày **kɔh**.
 get HIT REF child frog body search MEDL
He found the young frog he was looking for.

(11) jəh hùʔ kɔ məkəh hùʔ kɔʔ ɕiəʔ **kɔh**, ʔuə
 person NEG give if NEG GET eat MEDL 1s
 tɛm mɔŋ raʔ.
 know STAY FOC
I know that I cannot eat anything if they don't give me [food].

² Spoken Mon data are transcribed according to the phonological system developed in Jenny 2005 (pp. 23-42). Where quoting other sources, the original spelling is retained.

- (12) kla dɛh hù? tɛh ʔɔp **kəh,** dɛh priəŋ
 before 3 NEG HIT surrender MEDL 3 prepare
- lò senat.həlòk kəh.
 DEPOSIT cannon MEDL
- Before they had to surrender, they prepared that cannon.*

In some cases it is not clear what the scope of *kəh* is in an expression, as shown in (13).

- (13) laʔ həmèə pən həcɔt thɔʔ rəə kwan poy **kəh**
 when Burmese shoot kill DISCARD group village 1p MEDL
- when the Burmese shot the people in our village*

In this sentence, *kəh* can have scope over the pronoun *poy* ‘we’, over the NPs *kwan poy* ‘our village’ or *rəə kwan poy* ‘the people in our village’, or over the whole adverbial clause.

The deictic system of Spoken Mon is summarised in table 3.

Table 3: *The demonstratives of Spoken Mon*

Form	Exophoric use	Discourse function
<i>nɔʔ</i>	proximal	anaphoric, cataphoric
<i>kəh</i>	medial	anaphoric
<i>trʔ</i>	distal	anaphoric

The frequency of use of *kəh* suggests that it is more strongly grammaticalised than the proximal and distal demonstratives. The functions of grammaticalised *kəh* will be discussed in sections 4 and 5. I will now first turn to an overview of earlier studies and explanations of this particle.

2.2 Previous studies and definitions of *kəh*

The early western grammars and dictionaries of Mon describe *kəh* variously as a (deictic) pronoun or adjective ‘that’ (Halliday 1955 [1922]:115), or as

that; also a sign of the accusative case before the verb; it is often emphatic, laying stress on the word or phrase which precedes it. [...] As an emphatic particle = the Burmese <kāh> kà,³ very often it is simply indicative of the subject, and is equivalent to the nominative case. (Duroiselle 1962 [1921]:171f)

Shorto (1971) gives the following definition for *goh* in Old Mon and *gah* in Middle Mon:

³ “kà: 1 (Phr~) as for Phr; however, but; highlighting the subject or topic of a sentence, or contrasting it to a different topic.” (Okell and Allott 2001:5)

goh, goh, gohh /gɔh/ *ns. & n (ns.)⁴ deictic, that, and otherwise with nn. and n.phr. with mooted referent, esp. at head of sentence, the, then often serving to mark end of complex n.phr. [...] gah, rarely gâh ns., n phr. marker, rarely deictic except in such fixed phr. as khā gah then, het gah ra therefore [...].* (pp.82f)

In his dictionary of Spoken Mon (1962), he explains *kəh* as

Particle marking nn. and nominal phr., esp. in initial prominence position, and occasionally complete sents. [...] kəh is occasionally deictic, esp. in phr. following doa ['in, at'], **nù** ['from'], *where it stands as n.* (pp.78f).

The most extensive study of Mon is Bauer's *Morphology and Syntax of Spoken Mon* (1982). He dedicates three pages to *kəh*, which he calls a clitic or particle marking "determination". Other nominal categories listed by Bauer are possession, plurality, and deixis. The markers of all these categories follow the head noun in this order, with determination occupying the final slot. Bauer does not elaborate on his use of the term "determination", but from the discussion and examples it looks like he is referring to definiteness: "The most versatile clitic and, on some contexts, particle of the nominal phrase is kəh; in most environments, it translates into English 'the'." (Bauer 1982:322f) Another function of *kəh* mentioned by Bauer is "marking a boundary of any nominal phrase" (p. 323). Unlike Halliday, Bauer does not see *kəh* as a marker of grammatical relations or case marker:

In no case can kəh function as overt marker of grammatical subject or object. [...] in a sequence of subject verb object where subject and object may be simple or compound nouns with any number of noun-clitics either noun (subject or object) may be followed by kəh. (p. 324)

In other cases *kəh* is explained as marking a noun phrase as emphatic or in "prominence position", while Bauer's translations ('as for ...') suggest that *kəh* functions to mark a nominal expression as topic. The use as emphatic marker "is particularly obvious in cases where the noun position, preceding kəh, is occupied by a personal name or a personal pronoun" (p. 325).

Later in his study, Bauer states that *kəh* is the "only native and 'natural' device to nominalize verbs or incorporating verbal phrases into complex NPs" (p. 331). He concludes that, while *kəh* can be used to mark subordinate (relative) clauses,

it is inappropriate to call these sentence types 'relative clauses' [...] but rather to interpret them as nominalized VPs and to retain kəh simply as a determining and nominalizing (or noun) particle. (p. 332)

We have seen that for earlier authors *kəh* was part of the deictic system with some special functions, Bauer explains *kəh* as belonging to a distinct category, i.e. "determination". The main functions of *kəh* as given by Bauer are

⁴ n. = noun, n.phr. = noun phrase, ns. = noun suffix

1. Determination (i.e. definiteness?)
2. Emphasis
3. Marking of boundary of nominal phrase
4. Nominalisation of verbal phrases.

In the following sections more data from Modern Mon will be given and checked against the theoretical properties of the categories to which *kəh* may belong or has been analysed as belonging to.

3. Demonstratives

3.1 Theoretical overview

The most comprehensive survey of demonstratives is probably Diessel (1999), which covers morphological, syntactic, semantic and pragmatic aspects of demonstratives in a crosslinguistic perspective and also explores grammaticalisation paths involving demonstratives. Diessel gives three criteria that are relevant for the notion of demonstratives:

First, demonstratives are deictic expressions serving specific syntactic functions. [...] The notion that I will use [...] subsumes not only demonstratives being used as pronouns or noun modifiers but also locational adverbs such as English *here* and *there*.⁵

Second, demonstratives generally serve specific pragmatic functions. They are primarily used to focus the hearer's attention on objects or locations in the speech situation [...], but they may also function to organize the information flow in the ongoing discourse. [...] Demonstratives are often used to keep track of prior discourse participants and to activate specific shared knowledge. [...]

Finally, demonstratives are characterized by specific semantic features. All languages have at least two demonstratives that are deictically contrastive: a proximal demonstrative [...] and a distal demonstrative. (Diessel 1999:2)

According to Diessel, demonstratives have two main functions, viz. exophoric and endophoric (Diessel 1999:6). The exophoric use is seen as more basic and historically predating the endophoric uses. Languages vary in the distinctions they make in exophoric demonstratives. Most common are languages with a two-way or three-way distinction, based either on the relative distance from the point of reference or on the closeness to speaker/hearer/other person. Typically exophoric demonstratives are accompanied by a pointing gesture towards the object of reference in the discourse situation.

The endophoric use is divided into anaphoric, discourse deictic and recognitional. Diessel (1999:93) gives the following definitions of these terms:

Anaphoric and discourse deictic demonstratives refer to elements in the ongoing discourse. [...] Anaphoric demonstratives are coreferential with a prior NP; they keep track of discourse participants. Discourse deictic demonstratives refer to propositions; they link the clause in which they are embedded to the proposition to which they refer.

⁵ It is noteworthy that Diessel does not include other adverbial demonstratives (e.g. temporal, manner) in his survey.

Recognitional demonstratives do not refer to elements in the surrounding discourse; rather, they are used to indicate that the hearer is able to identify the referent based on specific shared knowledge.

This classification of demonstratives is largely identical to one proposed by Himmelmann (1996), with some differences in terms of labels rather than content. Himmelmann uses “situational use” for what Diessel calls exophoric, and “tracking use” for Diessel’s anaphoric. Diessel and Himmelmann do not mention the cataphoric use of demonstratives as seen in example (2) above. If we add this notion, we get the following possibilities for demonstratives:

- | | | |
|----|-------------|---|
| 1. | Exophoric: | Proximal, medial, distal (language specific scales) |
| 2. | Endophoric: | Anaphoric, cataphoric, discourse deictic, recognitional |

Demonstratives can occur either as pronouns or as attributes to nouns (adnominal demonstratives). In some languages there are different forms for adnominal demonstratives and pronominal demonstratives, while others use some kind of derivation to derive pronouns from adnominal demonstratives. The combination of an adnominal demonstrative with a relator noun or a demonstrative pronoun with an adposition can result in an adverbial expression, as in English ‘this way, that way’ and ‘like this, like that’. For lack of adequate data, Diessel mentions this adverbial use of demonstratives but does not give any details or further discussion (Diessel 1999:74). It is especially in this use, though, that the cataphoric – anaphoric distinction is relevant in many languages, including older stages of Mon.

Demonstratives tend to grammaticalise along different but consistent paths across languages. The most common endpoints of these grammaticalisation paths include

1. Pronouns (third person and relative)
2. Complementisers
3. Sentence connectives
4. Possessives

from pronominal demonstratives and

1. Definite articles or noun class markers
2. Boundary markers of postnominal relative clauses or attributes
3. “Determinatives”

from adnominal demonstratives (Diessel 1999:119ff).

“Determinatives” according to Diessel (1999:135) are demonstratives that mark the nominal head of a relative clause or function as head of a relative clause. This use is clearly different from Bauer’s label “determination” for the Mon particle *kəh* seen above.

I will now turn to more data from Modern Mon to see how the particle *kəh* fits in the category of demonstratives.

3.2 Demonstratives in Modern Mon

As seen above, Modern Mon has a deictic system with three members based on relative or absolute distance from the point of reference, usually i.e. the speaker in the spatial dimension and the present time in the temporal dimension. The basic forms in Mon, i.e.

nəʔ, *kəh* and *trəʔ*, are mainly used adnominally though they may occur as quasi-pronominals after prepositions like *dəə* ‘LOCATIVE’ and *nù* ‘ABLATIVE’. All three can be used to refer to objects in the real world located at different distances from the place of the speaker. They therefore count as demonstratives according to Himmelmann’s defining characteristic, i.e. “the element must be in a paradigmatic relation to elements which – when used exophorically – locate the entity referred to on a distance scale: as proximal, distal, etc.” (Himmelmann 1996:210f). as will be seen below, the Mon demonstratives also satisfy Diessel’s characterising criteria given above. The medial form can be used to indicate closeness to the hearer, but this does not necessarily have to be the case. If the speaker is located between the hearer and the object referred to, he still uses *kəh* if the object is some distance away from him. In the temporal dimension all three deictics can refer to a point in the past or future, but only *kəh* seems to be used with relative time reference. As relative time reference always implies anaphoric use, the use of *kəh* as the most frequent anaphoric particle is not surprising in this context. In some contexts, exophoric *trəʔ* seems to be used merely to express a great distance in time or space without pointing to a specific referent, as seen in (14). This use is not possible with *kəh* and *nəʔ*, which always point to a specific referent when used exophorically as in (15) and (16).

(14) *kla trəʔ ...*
 before DIST
Long time ago ...

(15) *pətəm nù nəʔ*
 start ABL PROX
from now on

(16) *pətəm nù kəh*
 start ABL MEDL
from then on, from that time

All three demonstratives have derived nominal forms functioning as demonstratives pronouns or locative adverbs. In most cases the prefix *ʔiʔ-* forms pronouns and *ʔə-* forms locative adverbs, but there is some inconsistency in the use of the forms, i.e. the forms with *ʔiʔ-* prefix are sometimes used as locative demonstratives.⁶ The general adverbial and temporal forms originate in collocations of a nominal head with the adnominal demonstratives, *sac* ‘manner’ for the former and *laʔ* (from Mon/Pali *kāla* ‘time’) for the latter.⁷ While the presence of a separate lexeme *həmùh* ‘now’ explains the gap in the proximal-temporal slot, no explanation can be given for the gaps in the distal-adverbial and distal-temporal slots. I will return to the gap in the medial-plural slot below. Table 4 summarises the demonstratives and their derivate forms in Modern Mon.

⁶ The prefix *ʔiʔ-* has other grammatical functions besides nominalisation. It is realised as -y- infix after some initials, including the ones occurring in the demonstratives, causing palatalisation (cf. Jenny 2003).

⁷ The proximal-adverbial has another, less common form, viz. *kənəʔ*, obviously contracted from *kiəŋ nəʔ* ‘this habit, this custom’.

Table 4: *Basic demonstratives and derivate forms in Mon*

	Basic form	Nominalised	Locative	Adverbial	Temporal	Plural
PROX	<i>nɔʔ</i>	<i>ʔiʔnɔʔ / ɲɔʔ</i>	<i>ʔənɔʔ</i>	<i>hənɔʔ</i>	<i>(həmùh)</i>	<i>tənɔʔ</i>
MEDL	<i>kəh</i>	<i>ʔiʔkəh / cəh</i>	<i>ʔəkəh</i>	<i>həkəh</i>	<i>ləkəh</i>	<i>(tɔʔ kəh)</i>
DIST	<i>ʔɻʔ</i>	<i>ʔiʔʔɻʔ / cɻʔ</i>	<i>ʔəʔɻʔ</i>	<i>(sac ʔɻʔ)</i>	<i>(laʔ ʔɻʔ)</i>	<i>təʔɻʔ</i>

As seen above (table 3), the use of all three demonstratives has been extended to anaphoric use, but only *kəh* has also acquired discourse deictic and recognitional functions. In cataphoric use, only the proximal *nɔʔ* seems to be used. In anaphoric function, *kəh* is by far the most frequent. The proximal and distal particles are chiefly used to explicitly indicate closeness or distance either in the discourse or in the real world, i.e. there is a close relationship between exophoric and anaphoric use.

Although *nɔʔ*, *kəh* and *ʔɻʔ* share a number of common features, there are some characteristics that distinguish *kəh* from the other two demonstratives. The first important difference is the singularising effect that is common to *nɔʔ* and *ʔɻʔ*, but not *kəh*. Mon has two nominal plural markers, viz. *tɔʔ* for definite or inclusive plural and *həlàŋ* for indefinite or open plural. If the number of referents of a nominal expression is either known from the context or irrelevant to the discourse, plurality is not overtly marked, except for personal pronouns.⁸ The pure nominal *klɔ* can therefore mean ‘(the/a) dog’ or ‘(the/some) dogs’. If the noun is modified by a proximal or distal demonstrative, it becomes singular. Plural referents must in this case obligatorily be marked by either the definite or the indefinite plural marker. This is not the case with the medial demonstrative, which does not necessarily imply singularity of the referent, irrespective of the use as exophoric or anaphoric deictic. The expression *klɔ nɔʔ* can only be interpreted as ‘this dog’; ‘these dogs’ is always *klɔ tənɔʔ* or *klɔ həlàŋ nɔʔ*. The same is true for *klɔ ʔɻʔ* ‘that dog’ and its plural forms *klɔ təʔɻʔ* and *klɔ həlàŋ ʔɻʔ*. Compare with these the use of *kəh* in sentences (17) and (18), both from a recorded conversation. While the plural is overtly marked in (17), it is implied in (18) by the use of the quantifier *həʔɔt* ‘all’.

- (17) ʔeŋkəlɔc **tɔʔ** **kəh** lɛ ʔa ləkəh.
 English PL MEDL ADD go then
Then the Englishmen went away.

- (18) la **kəh** thɔʔ thɔʔ həʔɔt.
 donkey MEDL discard DISCARD all
They left all the donkeys there.

The fact that the plural marker is obligatory with *nɔʔ* and *ʔɻʔ* but not with *kəh* explains why only the former two occur in the contracted form with the definite plural marker as prefix *tə-*, while *tɔʔ* is never shortened before *kəh*. That the collocation of a noun with a demonstrative entails singularity of the referent is a common feature in many languages of

⁸ There is a special pronoun for the first person plural, viz. *poɣ*. Second and third person pronouns always receive the definite plural marker with plural referents.

Southeast Asia, including Burmese, which has two grammatical morphs marking plurality (optional in most contexts except for pronouns) and Thai, which lacks fully grammaticalised plural markers. This fact sets *kə̀h* apart from the other demonstratives, not only in Mon, but also in an areal perspective.

Secondly, *kə̀h* can occur after constituents other than nominal expressions, including verbs, adverbs and whole clauses, whereas *nɔʔ* and *tɔʔ* are restricted to nominals. Relevant examples were given above in (11) and (12). There are some exceptions to this rule with *nɔʔ* and *tɔʔ* occurring after prepositions, as seen in sentences (14) and (15) above. In these cases, the adnominal forms function like pronominal forms.

The third difference is that *kə̀h* can occur after the other demonstratives, including *kə̀h* itself.⁹ Bauer explains the function of *kə̀h* in this position as merely emphatic. Relevant examples are given in sentences (19) - (21). In (19) the distal and proximal demonstratives must be interpreted as exophoric, reinforced by the deictic directional *ʔa* ‘go’ in the second part. Neither Chiangmai nor Three Pagoda Pass are mentioned in the preceding discourse, so anaphoric reading is excluded here. In (20) an anaphoric reading is more natural (the hill has been mentioned in the preceding sentence). In (21), the proximal is anaphoric, while the medial has another discourse function, most likely to indicate the element which the subsequent discourse is about.

- (19) cək ceh nù kəpac cə̀ŋmə̀y tɔʔ **kə̀h** mùə ləpac,
 march go.down ABL part Chiangmai DIST MEDL one side
 cək ʔa nù klə̀ŋ kyac.pə̀əʔ **nɔʔ** **kə̀h**
 march go ABL way Three.Pagoda.Pass PROX MEDL
 one part
 mùə kəpac.

One part [of the army] marched down from Chiangmai, one part marched from here at the Three Pagoda Pass.

- (20) dɔə kɔ ʔə̀ʔ tɔ hmoŋ.plə̀y **kə̀h** **kə̀h**
 LOC OBL REF hill prince MEDL MEDL
at the Prince Hill

- (21) dɔə ʔə̀rə̀ə nɔʔ **kə̀h**
 LOC matter PROX MEDL
in this matter

The function of *kə̀h* in these sentences is clearly no longer demonstrative (or deictic) in these expressions, i.e. the distance from the point of reference is no longer relevant. An analysis as anaphoric (as opposed to exophoric for the first of the two in each expression) is excluded by the fact that the localities in (19) are both not mentioned before in the text.

A last feature distinguishing *kə̀h* from the other demonstratives is stress. While *nɔʔ* and *tɔʔ* are always stressed, *kə̀h* can be either stressed or unstressed. There seems to be

⁹ In newer publications such as Newspapers, one can find other combinations of demonstratives like N *wùʔ* *nɔʔ*. This development seems to be very recent and may or may not be an extension of the secondary use of *kə̀h* to other demonstratives.

some regularity in the distribution and function of stressed and unstressed *kəh*, but it is not clear at the present stage of research whether stress can be assigned to certain functions of the particle.

Obviously *kəh* has some grammatical or pragmatic functions not shared by the other demonstratives. That *kəh* is more grammaticalised, i.e. has extended its area of applicability to more contexts, is further supported by the fact that it is far more frequent than the other two. In a short narrative text of one page, *kəh* occurs 86 times, while there are only two instances each of *nəʔ* and *tʁʔ* in the same text. As seen above, Bauer lists *kəh* as a marker of “determination”, presumably meaning definiteness. The next section takes a closer look at definiteness and tries to answer the question whether *kəh* is a kind of definite marker.

4. Definiteness

4.1 Defining definiteness

Definiteness is a grammatical category present in some but not all languages. Definiteness is typically a nominal category, closely associated with the noun phrase. According to Lyons (1999:1), “in many languages a noun phrase may contain an element which seems to have as its sole or principal role to indicate the definiteness or indefiniteness of the noun phrase.” Definite markers regularly arise from grammaticalised demonstratives, while the unstressed numeral ‘one’ becomes an indefinite marker in many languages. In Mon, there is a tendency to add unstressed *mùə* ‘one’ to noun phrases when they are first mentioned in a discourse, i.e. *mùə* can be taken as a kind of indefinite marker. Though the semantics of definite and indefinite markers seems to be straight forward, it is not easy to reach a crosslinguistically valid definition. Three features appear to be involved in definiteness, viz. uniqueness or inclusiveness, identifiability or familiarity and specificity or referentiality. In his study of definiteness, Lyons comes to the conclusion that the features ‘identifiable’ and ‘unique/inclusive’ are expressed in many languages in a single grammaticalised morpheme. This grammaticalisation of ‘identifiability’ and ‘uniqueness/inclusiveness’ is definiteness. There seems to be no reason to see these two concepts as related, but “the evidence for identifiability and inclusiveness being distinct features is lacking” (Lyons 1999:158). If a language expresses only ‘identifiability’ with a grammatical morpheme, this is taken to be definiteness in that language (Lyons 1999:278). Some languages use definite expressions for specific referents only, while others can use definite noun phrases also in generic contexts. That means that specificity is not a prototypical feature of definiteness.

Lyons only briefly mentions the use of definiteness “beyond the noun phrase” (Lyons 1999:45f). He does not pursue this topic, but in a later chapter (pp. 60ff) he returns to nominalising and other functions of definite articles: “An important aspect of the behaviour of definite articles is their use other than with nouns.” (p. 60). In many languages the definite article can be used to nominalise adjectives and verbs.

More strikingly, a definite article can sometimes serve to introduce an entire finite clause, thus functioning somewhat as a complementizer. [...] This applies particularly to subordinate clauses with an argument function. (ibid.).

According to Lyons definitions and extensions of use of definite articles, it seems plausible that *kəh* in Mon is a grammaticalised definite article which originates in and coexists with medial demonstrative.

4.2 Is *kəh* a definite article?

In this section I will give more examples of the different uses of *kəh*, trying to determine whether they can be united in the single category of definiteness. In addition to the data already presented in sections 2 and 3, the data presented in this section will concentrate on the features typically associated with definiteness, i.e. identifiability/familiarity, uniqueness/inclusiveness and specificity/referentiality. It should be noted from the beginning that, unlike definite articles in many languages that have them, *kəh* in Mon is never syntactically obligatory.

Definite noun phrases are used to refer to entities that are either known from the discourse context or that have been mentioned in the previous discourse. They are not used in presentational expressions ('there was a/*the NP'). In most cases, *kəh* is attached to nominals that are identifiable, either linguistically or extralinguistically. Personal names, pronouns and place names are inherently semantically definite, so that in many languages they are not explicitly marked as such. In Mon, *kəh* can be freely suffixed to names of people and places as well as to pronouns of all persons. This also suggests that the medial demonstrative value of *kəh* has been lost in this function. Otherwise the collocation of *kəh* with *ʔuə* 'I' and *poɣ* 'we' would be contradictory.¹⁰ That the (extralinguistic) identifiability can be based on shared knowledge or on the notions associated with a given frame is illustrated in sentence (22). While the Karen are not mentioned previously they are associated with the place name Mesali, obviously a Karen village.

- (22) tɛh nɪʔmɔ̀n dɔ̀ə mèsə̀li tɔ̀ʔ, cao kɔ̀.pɔ̀n ʔɔ̀t tɛk
 HIT invite LOC Mesali DIST return eat about beat
- pɔ̀n nə̀dɪ kəriə̀ŋ **kəh** kok “ʔɔ̀ khə̀nɔ̀, ʔɔ̀ khə̀nɔ̀”
 four hour Karen MEDL call “eat noodle eat noodles”
- ɕiə̀ʔ hə̀nɔ̀m.
 eat noodles
- We were invited to Mesali over there, and we went back to eat about four o'clock, and the Karen called us “ʔɔ̀ khə̀nɔ̀, ʔɔ̀ khə̀n ”, [that means] 'eat noodles'.*

Sentence (23) occurs at the beginning of a story, introducing the main participants in the text. The use of unstressed *mùə* resembles an indefinite article here, but the function of *kəh* is again not clear. Its cooccurrence with the quasi indefinite article *mùə* clearly excludes an interpretation as a definite marker/article. The referent of the nominal expression is clearly not identifiable in this context, neither within the discourse nor extralinguistically.

¹⁰ One might try to explain expressions like *ʔuə kəh* as indicating a kind of emotional distance, but this does not seem to be the case in Mon.

- (23) kla tʃʔ siəŋ, nùm mət̚ŋ kon.ŋàc mùə **kəh** ...
 before DIST be.so exist STAY child one MEDL
Long time ago, right, there was a boy ...

As can be seen from this sentence, *kəh* is not restricted to identifiable or known referents but can be used with a completely new actant if it is relevant in the subsequent discourse.

We have seen above that, unlike the other demonstratives, *kəh* does not imply singularity of the referent. The same is true for inclusiveness. An expression like *kon.ŋàc ləŋʔ kəh* ‘some of the children’ is perfectly grammatical in Mon. Therefore uniqueness or inclusiveness is obviously not part of the semantics of *kəh*.

The following sentences show that *kəh* is not restricted to specific (as opposed to generic) referents either. The examples are taken from Ketumati’s translations of English proverbs into Mon (Ketumati 1965), they are therefore representative of Literary Mon rather than the spoken variety, but there are no obvious differences in use in this respect.

- (24) ɲəh pyʔ **kəh** tət̚ŋ ɲəh thiə.
 person hungry MEDL be person angry
A hungry man is an angry man. (p. 130)

In (24), the use of *kəh* could be motivated by the complex NP including an attributive verb, but this does not explain why it is not present in the second part of the sentence. Clearly the expression ‘a hungry man’ is not specific, but generic. The second NP has predicative function and is non-referential, i.e. there is a difference in referentiality involved here. Possibly *kəh* is only used with referential noun phrases.

In sentence (25), the first noun is generic and the second specific, but both have the marker *kəh*.

- (25) mənih **kəh** ʃon khyɔp, nəy tət̚wət̚ʔ **kəh** praɔʔ.priəŋ pəŋɔp.
 man MEDL though consider master god MEDL prepare command
Man proposes, God disposes. (p. 168)

Lyons states that definite articles can develop into markers of complement clauses. That this is true for Mon *kəh* is illustrated in the following sentence, again taken from Ketumati’s translation of proverbs. Here *kəh* is used to mark argument (i.e. non-predicative) function of a verbal expression.

- (26) hɔm **kəh** lət̚ə lət̚n, klon **kəh** wət̚ plən.
 peak MEDL easy exceed do MEDL difficult again
Easier said than done. (p. 144)

This use of *kəh* does not make it a full fledged nominaliser, as the resulting expression does not have all the syntactic possibilities of a typical nominal expression. It can, for example, not be modified by an attributive verb or demonstrative: **hɔm kəh khɔh* ‘good talk’, **hɔm kəh nəʔ* ‘this talk’. Rather, *kəh* is used to mark the boundary of a complex expression, as already seen above with noun phrases modified by relative clauses.

It appears that *kəh* does have some of the semantic properties of a definite article, but it can hardly be seen as a grammaticalised marker of definiteness, unless we are ready to accept a more vague definition of definiteness than the one given by Lyons. The fact that *kəh* is used to mark a verbal expression as an argument rather than the predicate indicates that there is a possible connection with the topic – comment distinction, i.e. *kəh* might be used to mark topics. This means that the function of *kəh* is one of information structure rather than syntax or semantics. To this topic we will turn in the next section.

5. Information structure – Topic and focus

5.1 Topicality

Different authors have suggested different definitions of the term ‘topic’. Haiman gives the following definition:

The topic represents an entity whose existence is agreed upon by the speaker and his audience. As such, it constitutes the framework which has been selected for the following discourse. (Haiman 1978:585)

Dik states that “a linguistic expression will [...] usually contain some *given* information and some *new* information” (1989:265ff). Given information is what the speaker assumes to be available to the addressee as pragmatic information, which consists of “*general*, *situational*, and *contextual* information” (Dik 1989:9, 265). He goes on that “partially corresponding to the “given”/“new” distinction, we may distinguish the dimension of *topicality* and *focality*” (p. 266). Topicality tends to coincide with given information, and focus with the most salient or important piece of new information that is given about the topic.

Dik distinguishes different kinds of topics: The discourse topic denotes the entity which the discourse is “about”. A discourse may have different discourse topics with different degrees of centrality to the discourse. While a topic usually refers to an entity that is known or given, i.e. mentioned in the previous discourse (“GivTop”, Dik 1989:267), new referents may be introduced to the discourse as “NewTop” (ibid.). A NewTop denotes an entity that is not mentioned before but that is relevant to the following discourse. A topic can be reactivated after a stretch of discourse. It is then called a “resumed topic” (ResTop; ibid.).

Other authors (e.g. Erteschik-Shir 2007:) take up the features of topic a “givenness” and “aboutness”. Givón (2001:254) states that “topicality involves two aspects of referential coherence, one anaphoric, the other cataphoric”. The anaphoric aspect is “referential accessibility” and the cataphoric aspect is “thematic importance”.

Lambrecht (1994) says that in his use of the term, “the topic of a sentence is the thing which the proposition expressed by the sentence is ABOUT” (p.118). Also, the topic must be relevant to the present discourse and the predicate must add some new information about it. “The definition of topic in terms of aboutness and contextual relevance entails that there is an inherent relationship between topic and pragmatic PRESUPPOSITION” (p. 150). That means according to Lambrecht’s definition, the topic is part of the presupposition of the utterance. This is related to the pragmatic (or information-structural) status of the NP referred to by the topic expression. Lambrecht (p. 109) distinguishes different degrees of identifiability:

Unidentifiable: anchored, unanchored

Identifiable: inactive, accessible (textually, situationally, inferentially), active

According to Lambrecht, topics (as part of the presupposition) must be identifiable and may be active or activated.

While the definitions of topic (and focus) employed by different authors vary to some degree, they largely agree in that topics must be accessible to the hearer in some way and must be relevant to the ongoing discourse. Dik's NewTop seems to contradict the prerequisite of accessibility, but it can be seen as activated by its introduction to the discourse.

5.2 Topicality and focality in Mon

Mon has a focal particle, viz. *raʔ*, originating in a weak form of the copula *das* in Old Mon (s. Jenny 2005, 2006). This focus marker contrasts in some contexts directly with *kəh*. While *kəh* marks an expression as argument or non-predicative, *raʔ* can be used to mark the same expression as a predicative clause, as seen in (27).

(27) a. mənìh klɿŋ ɕiəʔ pɿŋ **kəh**
 man come eat rice MEDL
the people coming to eat

b. mənìh klɿŋ ɕiəʔ pɿŋ **raʔ**.
 man come eat rice FOC
The people are coming to eat.

As *kəh* is mostly used anaphorically, it marks known or accessible referents (nominal, verbal or clausal). Many sentences have the form [X *kəh*] [Y *raʔ*], with X being the presupposed part of the sentence, i.e. the theme or topic, and Y the predicate or comment, as in (28) and (29).

(28) mənìh **kəh** klɿŋ ɕiəʔ pɿŋ **raʔ**.
 man MEDL come eat rice FOC
That man is coming to eat.

(29) pəəʔ-kləm-pəəʔ kəh dəh hùʔ cao nəm pùh,
 three-hundred-three MEDL 3 NEG return yet NEG
 ʃət dəh kyaʔ **kəh** kyaʔ ʔa **yaʔ** (< ʔiʔ-raʔ).
 just 3 lose MEDL lose GO NSIT (< PFV-FOC)
In 1303, they had not retreated yet, but they had already lost [the war].

In other cases, *kəh* is used to mark a constituent as antitopic or afterthought, which is usually nominal, but may be verbal or adverbial. In this function, *kəh* is always unstressed. This use is illustrated in example (30), where the negation particle *pùh* marks the end of the basic sentence.

- (30) ʔuə hùʔ ket raʔ pùh, lòc **kəh.**
 1 NEG take FOC NEG text MEDL
I don't want it anymore, that book.

This may easily lead to an analysis of *kəh* as a topic marker. If we take the definitions given in section 5.1 for topic, there is indeed a large degree of overlap in the use of *kəh* and topicality. As seen in the examples in earlier sections, *kəh* marks a constituent which is either mentioned in the previous discourse or is pointed at in the discourse situation, i.e. which is textually or situationally accessible according to Lambrecht's terminology. Where a new referent is introduced to a discourse and is marked by *kəh*, this referent is identified as being important or relevant to the discourse, i.e. it is activated rather than already active, as in sentence (23) above. This can be seen as pragmatic accommodation of a new element that is introduced as if it were identifiable (cf. Lambrecht 1994:65ff).

Haiman (1978) suggests that topic expressions and conditional clauses are comparable (or even identical) in many respects. A clause in Mon containing *kəh* as boundary marker can be interpreted as conditional or sequential, as illustrated in sentence (31). This use of clauses ending in *kəh* is quite frequent in the spoken language, especially in non-final clauses within complex sentences.

- (31) ɲəh kɔ kəpac toə **kəh** ʔuə ʔa ràn ɕiəʔ thɔʔ.
 person give part FINISH MEDL 1 go buy eat DISCARD
After he gave me half [of the money] I went to buy something to eat.

Obviously *kəh* has acquired a function in organising information structure, i.e. marking identifiable or accessible chunks of information, nominal, verbal and clausal. Once this function is established, *kəh* can also be used to activate pieces of information as relevant to the discourse by pragmatic accommodation. In this information structural function, *kəh* is directly opposed to the focus marker *raʔ*. It is not, however, incompatible with it. Cooccurrence of *kəh* and *raʔ* in the same clause is not infrequent, especially in adverbial expressions like *hɔt kəh raʔ* 'it is for this reason'. In this context *kəh* clearly has demonstrative value, i.e. it anaphorically refers to information given earlier in the discourse. Therefore it also marks the 'reason', i.e. circumstances leading to some ensuing situation, as known or accessible, and relevant to the following discourse, i.e. *kəh* here also expresses topicality. The focal particle *raʔ* indicates that it is new and important in the discourse that the circumstances mentioned earlier are the reason for what follows.

To conclude this section, we can state that *kəh* is used to mark topical pieces of information in an ongoing discourse. The topicality may be based on the previous discourse, i.e. anaphoric referential accessibility, or on the relevance to the following discourse, or both. The accessibility may be due to shared general, situational, or contextual information.

6. Conclusion

We have seen that among the three demonstratives in Mon, *kəh* has a special position in many respects including semantics, syntax and phonology (stress pattern). Basically (and probably originally) a demonstrative particle indicating medial distance from the centre of interest, *kəh* has acquired a wide range of functions. The exact stages of the extension of function can not be traced from the documented material available in Mon. Already in Old

Mon, the particle *goḥ* had anaphoric and maybe boundary marking function, but the material is too scarce to draw a final conclusion. In Modern Mon, the polysemy (or multifunctionality) of *kəḥ* is rather far-reaching, but the different functions are interconnected and overlapping, so that in many cases it is not easy or even impossible to decide which function is prevalent in a given expression. The two main functions of *kəḥ* are as DEMONSTRATIVE with MEDIAL DEIXIS and marking IDENTIFIABLE INFORMATION RELEVANT to the ongoing discourse. This second function corresponds to a TOPIC MARKER, which can be seen as an extension of anaphoric uses of the demonstrative. Marking a constituent (phrase or clause) as identifiable or topical, *kəḥ* sets it apart from the new information given in the sentence, i.e. the PREDICATE. This leads to the frequent use of *kəḥ* as marker of a NON-PREDICATIVE expression.

The question of stress needs further investigation based on more extensive recorded texts. At the present stage of research it seems like *kəḥ* is fully stressed when it functions as exophoric demonstrative or as resumed topic marker (anaphoric), but unstressed when marking a given topic or antitopic.

Figure 1 summarises the possible development of the different functions of *kəḥ* in Mon.

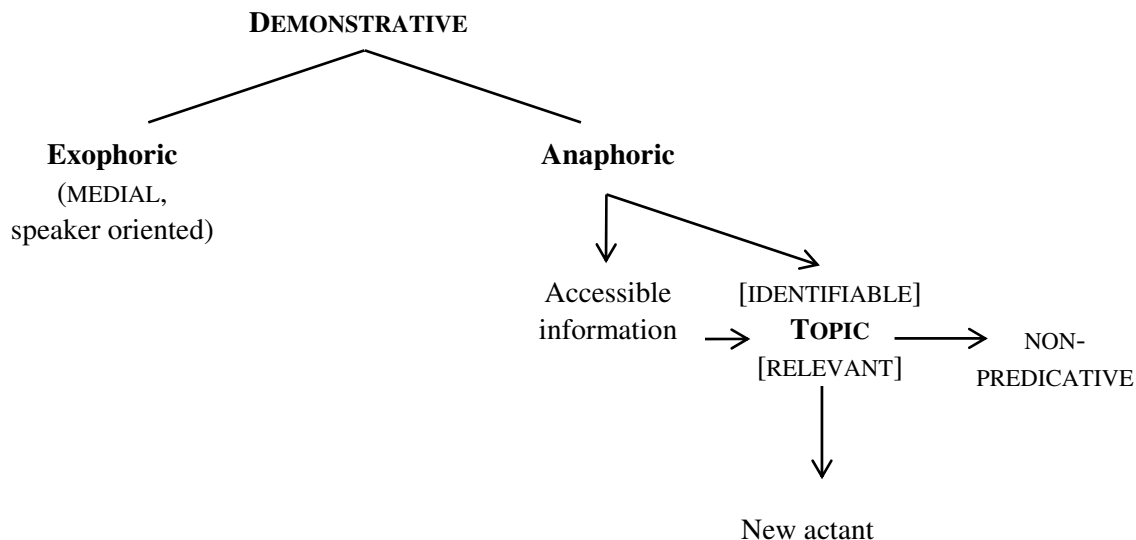


Figure 1: *Development of kəḥ*

Abbreviations

ABL	Ablative	MEDL	Medial
ADD	Additive	NEG	Negation
ATTR	Attributive	NSIT	New situation ('already, now')
COM	Comitative	OBL	Oblique
DIST	Distal	PFV	Perfective
EMPH	Emphatic	p/PL	Plural
FOC	Focus	PROX	Proximal
INSTR	Instrumental	Q	Question
LOC	Locative	REF	Referential

References

- Bauer, C. 1982. *Morphology and syntax of Spoken Mon*. Doctoral dissertation at SOAS, London.
- Diessel, H. 1999. *Demonstratives. Form, function, and grammaticalization*. Amsterdam/Philadelphia: John Benjamins.
- Dik, S. C. 1989. The theory of functional grammar. Part I: The structure of the clause. Dordrecht/Providence: Foris Publications.
- Duroiselle, C. (ed.) 1962 [1921]. *Epigraphia Birmanica. Vol. II. The Talaing plaques on the Ananda plates*. Rangoon : Government Printing and Stationery.
- Erteschik-Shir, N. 2007. Information structure. The syntax-discourse interface. Oxford: OUP.
- Givón, T. 2001. *Syntax*. Vol. 2. Amsterdam/Philadelphia: John Benjamins.
- Haiman, J. 1978. Conditionals are topics. *Language* vol. 54/3, 564-89.
- Halliday, R. 1955 [1922]. *A Mon-English dictionary*. Rangoon: Ministry of Union Culture.
- Himmelmann, N. P. 1996. Demonstratives in narrative discourse. In Fox, B. 1996. *Studies in anaphora*. Amsterdam/Philadelphia: John Benjamins; 201-254.
- Jenny, M. 2003. New infixes in Spoken Mon. In *Mon-Khmer Studies* 33, 183-194.
- Jenny, M. 2005. *The verb system of Mon*. Zurich: ASAS 19.
- Jenny, M. 2006. Mon *ra?* and *noj*: Assertive particles? *Mon-Khmer Studies* 36, 21-38.
- Ketumati, Bh.. 1965. *4 languages – English, Mon, Burmese and Pali*. Moulmein: Suvannavam sa Printing House.
- Lambrech, K. 1994. Information structure and sentence form. Topic, focus and the mental representations of discourse referents. Cambridge: CUP.
- Lyons, C. 1999. *Definiteness*. Cambridge: CUP.
- Okell, J. and A. Allott. 2001. *Burmese/Myanmar dictionary of grammatical forms*. London: Curzon.
- Shorto, H. L. 1962. *A dictionary of Spoken Mon*. London: Oxford UP.
- Shorto, H. L. 1971. *A dictionary of Mon inscriptions from the VIth to the XVIth centuries*. London: Oxford UP.
- Yanson, R. A. 2002. A list of Old Burmese words from 12th century inscriptions. In Beckwith C. I. (ed.) *Medieval Tibeto-Burman languages*. Leiden/Boston/Köln: Brill, 163-167.