

Handling *ki* in Hindi for Hindi-English MT

R. Mahesh K. Sinha

Indian Institute of Technology Kanpur
rmk@iitk.ac.in

Anil Thakur

Indian Institute of Technology Kanpur
anilt@iitk.ac.in

Abstract

ki is an indeclinable element (particle) in Hindi which is used in multiple roles that have multiple mapping patterns in English. In one of its uses, *ki* functions as a clause complementizer and is mapped usually by *that* in declarative clauses and by various wh-words (such as *what*, *why*, *where*, *how*, etc.) in interrogative clauses. The contexts of these mappings are dependent on syntactic-semantic types of the clause. In its non-complementizer use, *ki* is used to denote various other functions such as coordinate conjunction, purpose and reason clause conjunction, yes-no question particle, etc. It is a difficult task to identify the different uses of *ki* and determine its multiple mapping patterns in the context of Hindi-English machine translation. A detailed linguistic analysis is needed to disambiguate the different contexts of *ki* in Hindi. In this paper, we examine the multiple uses and patterns of *ki* in Hindi and propose strategies for their identification and disambiguation for Hindi-English MT.

1 Introduction

ki is a particle in Hindi with multiple functions and multiple mapping patterns in English. In certain types of declarative sentences, *ki* functions as a clause complementizer and is mapped by *that* in English (1)¹.

- (1) a. *raama-ne kah-aa ki siitaa bhii aa-egii.*
{Ram-ERG tell-PST COMP Sita also come-FUT}
'Ram said that Sita would also come.'
b. *yah kaha-naa ki bhaarat meN sabhii amiir*

¹ The Hindi examples have been given in italics. Abbreviations/acronyms:

ACC: Accusative Case, CMT: Completive Aspect, COMP: Complementizer Particle, DAT: Dative Case, ERG: Ergative Case, FocP: Focus Particle, FUT: Future Tense, GER: Gerund Suffix, INF: Infinitive, Q: Question Particle, PST: Past Tense, PR: Present Tense, SU: Subjunctive Form

hEN galat hE.

{this say-GER COMP India in all rich be.PR
wrong be.PR}

'To say that all are rich in India is wrong.'

- c. *yah samaacaara ki ve aa cuke hEN sahii hE.*
{this news COMP they come CPT be.PR
correct
be.PR}

'The news that they have arrived is correct.'

- d. *use maalum hE ki ve aa rahe hEN.*
{him known be.PR COMP they come PROG
be.PR}

'S/he knows that they are coming.'

In certain types of sentences with interrogative subordinate (*ki*-) clause, *ki* functions as a clause complementizer and in this case its mapping patterns in English depend on the nature of the interrogative clause (2).

- (2) a. *us-ne puch-aa ki kyaa tum aa-oge.*
{he-ERG ask-PST COMP what you come-FUT}
'He asked whether you/I would come.'
b. *us-ne soc-aa ki kyaa baata hE.*
{He-ERG thought-PST COMP what matter
be.PR}
'He thought what the matter was.'
c. *raam jaanataa hE ki kOn aayaa hE.*
{Ram knows be.PR COMP who come
be.PR}
'Ram knows who has come.'

In (1-2), the verb in the main clause selects a sentential complement as its argument and *ki* introduces that sentential complement. Thus *ki* in (1-2) is a complementizer that introduces a clause selected by the verb in the main clause. Further, we notice that the mapping of *ki* in these cases is dependent on the nature of the subordinate clause. For instance, if the subordinate clause is an affirmative sentence, *ki* is mapped by *that* in English (1). However, if the subordinate clause is a question clause, *ki* cannot be mapped by *that*. In this case, it is grouped with the interrogative word and only the interrogative word is mapped in English. Thus *ki* is mapped by null in this case.

Furthermore, *ki* also occurs as various conjunction markers. In (3a), it functions as a coordinate conjunction word, typically represented by *yaa* ‘or’ in Hindi. In (3b), it functions as a purpose clause conjunction, typically represented by *taaki* ‘so that’ in Hindi. In (3c), it occurs as a marker for contrastive negation (implicit assertion).

- (3) a. *raam aayegaa (yaa) ki nahiiN (aa-yegaa)?*
 {Ram come-FUT or not (come-FUT)}
 ‘Will Ram come or not?’
 b. *raam yahaan aayaa hE (taa) ki aapase baat kar sake.*
 {Ram here come be.PR so that you-to talk do can}
 ‘Ram has come here so that he can talk to you.’
 c. *raam dillii jaa-yegaa na ki kolkataa.*
 {Ram Delhi go-FU not Kolkata}
 ‘Ram will go to Delhi, not to Kolkata.’

In (3), *ki* is not a complementizer. It occurs as a coordinate conjunction which is translated (mapped) in English by *or*, as in (3a). We notice that in (3a), *ki* can occur either with *yaa* ‘or’ or without it. In (3b), *ki* functions as a marker of a purpose clause and can alternate with *taaki* ‘so that’. In (3c), *ki* occurs with *na* ‘not’ in the role of a contrastive negation marker.

In yet another set of its uses, *ki* is used in different senses where its meaning (and mapping pattern) is dependent not only on factors such as the nature of the sentence but also on nature of the neighboring word, etc. (4).

- (4) a. *us-ne vEsaa hii kiyaa jEsaa ki apane kahaa thaa*
 {he-ERG like-that EMP did REL PAR you-ERG told be.PT}
 ‘He did exactly as you had told.’
 b. *Esaa na ho ki ve aaja hii jaa rahe hoN.*
 {this-like not be that they today FocP go PROG be-SU}
 ‘Maybe they are leaving today itself.’
 c. *ho na ho ki raam kahiiN yahii chupaa ho.*
 {may not be that Ram somewhere here-itself hiding be.PR}
 ‘Maybe Ram is hiding somewhere here itself.’

In (4a), *ki* occurs (optionally) with relative pronoun *jEsaa* ‘that-like’. In (4b-c), *ki* introduces a “hypothetical/possibility” clause. These represent the closest semantic interpretation rather than exact translations that are hard to achieve at the sentence level. The word ‘itself’ is used for Hindi focus

particle ‘hii’. In 4(c), this focus particle appears as a suffix.

The set of examples mentioned above clearly show the ambiguous nature of *ki* with its multiple functions. This also clearly shows the difficulty it poses for the Hindi-English MT. The existing literature; the traditional grammars (Guru, 1925, Vajpai, 1958, Sharma, 1958) as well as the modern linguistics literature (Kachru, 1980, Mohanan, 1994, Subbarao, 1984) do not discuss this aspect of *ki* in Hindi in any detail. There is also no comparative grammar of Hindi and English that deal with the issue of mapping patterns of such elements from Hindi to English. The issue certainly needs a detailed linguistic analysis. To correctly identify the exact category of *ki* in a particular clause, we need to identify the type of the verb in the main clause as well as the nature of the subordinate clause. In many other cases, the disambiguation of *ki* is dependent on various other factors. In this paper, we examine those linguistic markers that determine the different functions of particle *ki*. On the basis of these linguistic markers, we formulate strategies for the proper categorization of *ki* and its exact mapping in English. In section 2, we discuss the various mapping patterns of the different uses of particle *ki* in Hindi and propose rules for their mapping into English. The implementation details are briefly outlined in section 3 followed by the concluding remarks.

2 Structural Patterns and Disambiguation Rules of *ki*

For the purpose of the present study, we categorize *ki* on the basis of its different uses. For instance, *ki*, in its complementizer use, can be divided into two broad categories: “that-comp” and “wh-comp”. In its non-complementizer use, *ki* is used in various roles such as “coord-conj” (coordinate conjunction), “purpose-adv” (purpose adverbial clause conjunction for *taaki* (‘so that’), “reason-adv” (reason adverbial clause conjunction for *kyoNki* (‘because’), “relative” (occurs with relative pronouns, e.g. *jo ki* (‘who/which’), “subjunctive-mood” (subjunctive mood marker), “subjunctive-conditional” (subjunctive clause with conditional subordinate clause, for *yadi* (‘if’), “temporal-adv”, among others. Hence we categorize *ki* according to its different functions and discuss examples of each category through subsections 2.1-2.16. We also discuss the disambiguation and mapping rules for them in each section. The rules are based on the linguistic analysis of the different patterns of *ki*. For instance, we can notice that *ki_1* and *ki_2* are directly dependent on the type of the verb in the main

clause. The different mapping patterns of *ki* in *ki_2* are further dependent on the nature of the interrogation in the subordinate clause.

Therefore we propose that the strategies for disambiguation of *ki* should take into account the types of the verb, the types of the main and/or the subordinate clause and also the nature of the neighboring elements. For instance, we notice that in certain patterns (*yaanii ki, yaa ki, na ki, jEse ki, maano ki, ho na ho ki, Esaa hE ki, kyoN ki*, etc.) *ki* is always dependent on the nature of the neighboring element and it is always grouped together with the relevant neighboring words. In other patterns, *ki* is an independent word and the rules for its identification/categorization are dependent on context within the sentence.

2.1 *ki_1* (that-comp)

The use of *ki* as a clause complementizer is its most common use in Hindi. The context is defined by the type of the verbs and a certain type of nouns that select a *ki*-clause as a complement. Some of the representative patterns are discussed in (i-ii) below.

i. *ki*-clause selecting verbs: The communicative verbs (e.g. *kahanaa* ‘tell/say’, *bolanaa* ‘speak’, *samajhaanaa* ‘convince’, *bataanaa* ‘tell’, etc.), sense_related verbs (*samajhanaa* ‘understand’, *dekhanaa* ‘see’, *sunanaa* ‘hear’, etc.), wish verbs (*caahanaa* ‘want/wish’, *praathanaa karanaa* ‘pray’, *icchaa karanaa* ‘wish/desire’, etc.) select a *ki*-clause as the complement of the verb.

(5) communicative verbs

raam kahataa hE ki vah nirdosha hE
{Ram says be.PR *ki_1* he innocent be.PR}
‘Ram says that he is innocent.’

(6) sense_related verbs

raam-ne dekhaa ki siitaa dukhii hE.
{Ram-ERG saw *ki_1* Sita sad be.PR}
‘Ram saw that Sita was sad.’

(7) wish verbs

raam caahataa hE ki siitaa yahaaN rahe.
{Ram wants be.PR *ki_1* Sita here-only live}
‘Ram wants that Sita should live here.’

ii. *ki*-clause selecting noun: The gerunds which are derived from a *ki*-clause selecting verbs also selects a *ki*-clause as their complement. Besides, there are a few nouns such as *kathan* (‘saying’), *baat* (‘matter’), *tathy* (‘fact’), *samaacaar* (‘news’), *khabar* (‘news’) that select a *ki*-clause as a complement. These nouns are mostly preceded by a demonstrative element (*yah* (‘this’), *vah* (‘that’),

etc), which in this case functions as a definite determiner and is mapped by *the* rather than demonstrative *that/this* in English.

(8) Gerunds

yah kahanaa ki raam nirdosh hE galat hE.
{this saying *ki_1* Ram innocent be.PR wrong be.PR}
‘To say that Ram is innocent is incorrect.’

(9) Simple Nouns

yah khabar ki raastrapati bhii aa rahe hEN
sahii nahiiN hE.
{this news *ki_1* president also come PROG be.PR correct not be.PR}
‘The news that the President is also coming is not correct.’

Rule1: *ki* = *ki_1* in the following cases. *ki_1* is mapped by *that* in English.

- If the category of the verb preceding the *ki*-clause is either communicative or wish or sense-related. E.g. (5-7).
- If the *ki*-clause is preceded by a gerund that is derived from the verb of the category communicative or wish or sense-related. E.g. (8).
- The *ki*-clause is preceded by a noun of the class “factual” (e.g. *baata* ‘matter’, *samaacaara* ‘news’, etc). E.g. (9).

2.2 *ki_2* (wh-comp)

A class of words such as communicative verbs (including gerund and some nouns as mentioned in section 2.1 above), select an interrogative subordinate clause as the complement clause. However, all the verbs which select a complement clause, as in the case of *ki_1*, can also select an interrogative complement clause. In interrogative clauses, *ki* is mapped by null marker. In this case, only the interrogative (*k-*) words are mapped by matching wh-words in English. That is, [[*ki* + *k*-word] = [wh-word]]. The *k*-words are either the argument NP or an adjunct NP. We include all the *k*-words in one category for the reason that they behave in the same way with respect to the neutralization of *ki* in all the cases. Some examples illustrating the different types of interrogative subordinate clauses are listed in (10-16).

(10) *ki_2a*: who

raama-ne puuchaa ki kOn aayaa hE.
{Ram-ERG asked *ki_2a* who come be.PR}
‘Ram asked who had come.’

(11) *ki_2b*: what

raam-ne puuchaa ki kyaa baat hE.
{Ram-ERG asked *ki_2b* what matter be.PR}
‘Ram asked what the matter was.’

(12) *ki_2c*: which
raama-ne puuchaa ki kOn-sii kitaab khariidii jaaye.
 {Ram-ERG asked *ki_2c* which book bought go.SU}
 ‘Ram asked which book to buy.’

(13) *ki_2d*: how
raama-ne puuchaa ki kEse vahaan jaayeN.
 {Ram-ERG asked *ki_2d* how there go.SU}
 ‘Ram asked how to go there.’

(14) *ki_2e*: why
raama-ne puuchaa ki kyoN vahaan jaayeN.
 {Ram-ERG asked *ki_2e* why there go.SU}
 ‘Ram asked why to go there.’

(15) *ki_2f*: where
raam-ne puuchaa ki kahaan jaayeN.
 {Ram asked be.PR *ki_2f* where go.SU}
 ‘Ram asked where to go.’

(16) *ki_2g*: when
raam-ne puuchaa ki kab jaayeN.
 {Ram asked be.PR *ki_2g* when go.SU}
 ‘Ram asked when to go.’

Rule2: *ki* = *ki_2* when followed by an interrogative clause.

2.3 *ki_3* (Coord-Conj)

ki often occurs either along with the coordinate conjunction marker *yaa* (‘or’) or in place of *yaa* (‘or’). In this use *ki* either denotes an affirmative clause (17a-b) or (implicit negation) question clauses (18).

(17) *ki_3a*: whether
 a. *unheN koi matalab nahiiN ki aap yahaan rahate hE (yaa) ki kahiiN aur.*
 {he any interest not *ki_2* you here live be.PR *ki_3a* somewhere else}
 ‘They do not care whether you live here or somewhere else.’
 b. *raam socataa hE ki vah yah baat kisii se kahe ki naa kahe.*
 {Ram ponder be.PR *ki_3a* he this matter anybody to tell *ki_3b* tell}
 ‘Ram ponders whether he tell this matter to anybody or not.’

(18) *ki_3b*: or + Q (implicit question)
 a. *raam jaa-yegaa (yaa) ki siitaa (jaa-yegii)?*
 {Ram go-FUT *ki_3b* Sita (go-FUT)}
 ‘Will Ram go or Sita (will go)?’
 b. *raam aa-yegaa (yaa) ki nahiiN?*

{Ram come-FUT or *ki_3b* not}
 ‘Will Ram come or not?’
 c. *ki_3c*: 0
raam aayegaa yaa ki nahiiN?
 {Ram come-FU or *ki_3c* not}
 ‘Will Ram come or not?’

Rule3: *ki* = *ki_3* under the following conditions.

- ki* = *ki_3a* when *ki* joins two clauses in which the subordinate clause contains *ki* of *ki_3b* type and the sentence is not interrogative.
- ki* = *ki_3b* when *ki* occurs in an interrogative (yes-no question) and joins two verb phrases or is followed by the negation particle *naa/nahiiN* (‘not’) or by a noun/adjective/adverb.
- ki* = *ki_3c* when *ki* is preceded by *yaa* (‘or’).

2.4 *ki_4* (purpose-adv)

In this pattern, *ki* introduces a purpose adverbial clause. The usual marker for the purpose adverbial clause in Hindi is *taaki* (‘so that’). However, *ki* also frequently occurs in place of *taaki*. Therefore this *ki* is another marker for the purpose clause in Hindi. *ki* in the sense of *taaki* (‘so that’) also occurs in another context where it occurs after *jisase* (‘by which’), as in (19d). In this case, *ki* is mapped by *so that* in English.

(19) *ki* = *taaki* (‘so that’)
 a. *raam aayaa hE ki aap-se baateN kara sake.*
 {Ram come be.PR *ki_4* you-to talk do can.SU}
 ‘Ram has come so that he can talk to you.’
 b. *raam jyaadaa khaataa hE ki moTaa ho sake.*
 {Ram more eat be.PR *ki_4* fat become can.SU}
 ‘Ram eats more so that he can become fat.’
 c. *(kyaa) maalaa tEyaar hE ki mEN le jaaun?*
 {(Q) garland ready be.PR *ki_4* I take go.SU}
 ‘Is the garland ready (so) that I can take it?’
 d. *us-ne bahut mehanat kii ki vah paasa ho sake.*
 {he-ERG lots hardwork did *ki_4* he pass be can.SU}
 ‘He worked hard so that he could pass.’

Rule4: *ki* = *ki_4*:

The subcategorizational frame of the verb is saturated without a *ki*-clause, i.e., the categories of verb are other than communicative, wish, or sense-related.

2.5 *ki_5* (reason-adv)

ki also introduces a reason adverbial clause. In this case, the function of *ki* is identical to the function of *kyoNki* ‘because’. In this case, *ki* is mapped by *because* (of) in English.

(20) *ki* = (*kyoNki*) (‘because’)

- a. *raam dukhii hE ki aap usase baat nahiiN karate hEN.*
 {Ram sad be.PR ki_5 you him-to talk not do be.PR}
 ‘Ram is sad because you do not talk to him.’
- b. *siitaa khush hE ki aap usase milane aaye.*
 {Sita happy be.PR ki_5 you her meet came}
 ‘Sita is happy because you came to see her.’

Rule5: *ki* = ki_5 when the predicate of the main clause contains a psych_adjective (e.g. *naaraaja* (‘angry’)) and the adjective is not preceded by a degree marker element such as *itanaa/itanii etc.*²

2.6 ki_6 (relative)

ki optionally occurs along with almost all the relative pronouns in a relative clause construction. *ki* occurs jointly with the relative pronoun but it can also occur separately. It is the latter case that has been discussed here. In both the cases, where the relative pronoun occurs with *ki*, it denotes non-restrictive relative clauses. It is always grouped with the relative pronoun it occurs with and it itself is mapped by a null marker in English.

- (21) *ki* = 0
- a. *vah laRakii jo ki aapake paas bETii hE raama kii bahana hE.*
 {that girl who ki_6 you near sitting be.PR Ram of sister be.PR}
 ‘The girl who is sitting near you is Ram’s sister.’
- b. *vah keval unase milaa jinase ki usako kaam thaa.*
 {he only them met whom ki_6 he-DAT work be.PST}
 ‘He only met those with whom he had a work.’
- c. *vah utanaa nahiN jaanataa hE jitanaa ki Ram jaanataa hE.*
 {he that-much not know as-much ki_6 Ram know be.PR}
 ‘He does not know as much as Ram knows.’

Rule6: *ki* = ki_6 when it follows a relative pronoun (e.g. *jo, jisane*, etc.).

2.7 ki_7 (subjunctive-clause)

ki in this category occurs in different types of sentences usually categorized as “hypothetical” or

² An example is given below:

siitaa itanii khush hE ki vaha bola nahiiN sakatii.
 {Sita so happy be.PR ‘*k_defaulti*’ she speak not can}
 ‘Sita is so happy that she can not speak’.

“wish” clause. In this case, *ki* is mapped in English by a null marker. *ki* is always grouped with the preceding clause.

- (22) *ki* = 0
- a. *ho naa ho ki raam kahiiN yahiiN chupaa ho.*
 {may be ki_7 Ram somewhere here-only hiding be.SU}
 ‘Maybe, Ram is hiding somewhere here (only).’
- b. *kahiiN yEsaa na ho ki aap baarish meN phaNs jaayeN.*
 {may be ki_7 you rain in trapped go-SU}
 ‘Maybe, you get trapped in rain.’
- c. *sambhav hE ki ve paTanaa pahuNc gaye hoN.*
 {maybe be.PR ki_7 they Patna reached be.SU}
 ‘Maybe, they have arrived in Patna.’
- d. *iishwar kare ki usa-ne hameN naa dekhaa ho.*
 {God do ki_7 he-ERG us not seen be.SU}
 ‘Maybe (hopefully), they have not seen us.’

Rule7: *ki* = ki_7 when *ki* follows a hypothetical or “possibility” clause (e.g. *ho naa ho* (‘maybe’), *yEsaa naa ho* (‘maybe’)), which is in subjunctive mood.

2.8 ki_8 (hypothetical-conditional)

This type of *ki* occurs in place of *yadi* (‘if’) in a hypothetical sentence and denotes hypothetical-conditional function. In this case, *ki* is mapped by ‘if’ in English.

- (23) *ki* = if
- a. *kitanaa acchaa hotaa ki ham pahale mile hote.*
 {how-much good be.SU ki_8 we earlier met be.SU}
 ‘It would have been so nice if we had met earlier.’
- b. *bahut kuch ho gayaa hotaa yahiiN (/bas) ki kucha aur koshisha kiye hote.*
 {a lots be went be.SU only (only) ki_8 some more effort did be.SU}
 ‘A lot could have been achieved only if we had made some extra effort.’

Rule8: *ki* = ki_8 when *ki* is followed by a “conditional-hypothetical” clause with verb ending with *-taa* and the subordinate clause is in subjunctive mood with the verb ending in *-taa/-te*.

2.9 ki_9 (temporal-adv)

This type of *ki* is used to denote temporal function. In this case, *ki* can be mapped in English either by a null marker (23) or by the phrase *the moment* (24-25)

(24) ki_9a: 0

raam jEse hii ghar pahuNcaa ki baarisha banda ho gayii.

{Ram as soon as home reached ki_9a rain stopped be went}

‘As soon as Ram reached home, it stopped raining.’

(25) ki_9b: the moment

a. *gaaRii rukii hii thii ki ve dikh paRe.*

{train stopped only be.PST ki_9b they see dropped}

‘The moment the train had stopped, we saw them.’

b. *dinesh aayaa kyaa ki sabhii us par barasa pare.*

{Dinesh come what ki_10 all him on shouted start}

‘The moment Dinesh came, everybody started shouting at him.’

Rule9: *ki* = ki_9 under the following conditions:

a. *ki* = ki_9a when the main clause contains temporal adverbial such as *jEse hii* (‘as soon as’).

b. *ki* = ki_9b when *ki* follows a clause in which the main verb is focused with particle *hii* or *ki* is preceded by *kyaa* (‘what’) and the subordinate clause contains an inchoative verb (e.g. *paRanaa* (‘start’), *laganaa* (‘start’)).

2.10 ki_10 (resultative clause)

ki-clause denotes the result of the action of the verb in the main clause. In this case, *ki* is mapped by *that* in English.

(26). ki_10: that

a. *us-ne itanaa khaayaa ki usase chala nahiiN jaa rahaa tha.*

{he-ERG so much ate ki_10 him by walk not go PROG be.PST}

‘He ate so much that he was unable to walk.’

b. *use itanaa piitaa ki vah behosh ho gayaa.*

{him so much beat ki_10 he unconscious become went}

‘He was beaten so much that he became unconscious.’

c. *unako itanaa kaam hE ki unahen samay hii nahiiN milataa.*

{them so much work be.PR ki_10 they time even not get}

‘They have so much work that they do not get time.’

Rule10 (Default Rule): *ki* = ki_10 when no context/condition is defined.

2.11 ki_11 (imperative clause)

This type of *ki*-clause introduces an imperative clause. *ki* is mapped by a null marker and the *ki*-clause as a whole is mapped by an infinitive clause in English.

(27) ki_11: 0

usane tumheN kaha ki vahaan mat jaao.

{he you told ki_11 there not go-IMP}

‘He told you not to go there.’

Rule11: *ki* = ki_11 when *ki* introduces an imperative clause and the main verb belongs to the category as given under rule ki_1.

It is interesting to note here that the Hindi imperative clause is transformed to a complementary gerund clause in English.

2.12 ki_12 (alternate-negation clause)

This type of *ki*-clause introduces an alternate-negation clause. In this case, *ki* is always grouped with *naa* and it itself is mapped in English by a null marker.

(28) ki_12: 0

ham-ne aap-ko bulaayaa thaa naa ki unako.

{we-ERG you called be.PST not ki_12 them}

‘We had called you, not them.’

Rule12: *ki* = ki_12 when preceded by *naa* (‘not’).

2.13 ki_13 (simile clause)

This type of *ki*-clause introduces a clause that functions as a simile to compare the action in the main clause. *ki* occurs with words such as *maano* (‘suppose’) or *jEse* (‘like’). In this case, *ki* is always grouped with *maano* or *jEse* and the group is mapped in English by ‘as if’.

(29) ki_13: as if

a. *vah yEse bolataa hE maano ki meraa dusmana ho.*

{he as speak be.PR suppose ki_13 my enemy be.SU}

‘He speaks as if he is my enemy.’

b. *vah Ese calataa hE jEse ki sharaaba pii rakhaa ho.*

{he as walk be.PR suppose ki_13 bear drink kept be.SU}

‘He walks as if he is drunk.’

Rule13: *ki* = ki_13 when *ki* follows *maano* (‘suppose’) or *jEse* (‘like’) and the subordinate clause is in the subjunctive mood. The main clause contains a correlative pronoun *Ese*.

2.14 ki_14 (hypothetical clause)

This pattern of *ki*-clause introduces a “hypothetical” clause. The main clause contains phrases like (*kahiiN*) *Esaa to nahiiN* (‘may be’) and *ki* is grouped with this phrase. *ki* does not have separate map in English.

- (30) ki_14: 0
(kahiiN) Esaa to nahiiN ki vah na aaye?
 {may be ki_14 he not come.SU}
 ‘May be he will not come?’ Or
 ‘Will he and will he not come?’

Rule14: *ki* = ki_14 when preceded by a specific clause containing phrase like (*kahiiN*) *Esaa to nahiiN*.

2.15 ki_15 (address clause)

ki can also be used to introduces an “address” clause. In this case, *ki* is always grouped with the main clause and *ki* itself is mapped in English by a null marker.

- (31) ki_15: well
Esaa hE ki tumhaaraa vahaaN jaanaa thiik nahin hE.
 {well ki_15 your there going good not be.PR}
 ‘Well, it is not good for you to go there.’

Rule15: *ki* = ki_15 when *ki* is preceded by a (main) clause containing *Esa hE*.

2.16 ki_16 (extent clause)

ki is also used to denote an extent clause. In this case, *ki* is always preceded by ‘*jahaaN taka*’ or ‘*yahaaN taka*’ with which it is grouped together and it does not have its own map in English.

- (32) ki_16a: as far as
jahaaN tak ki unakaa savaal hE, ve jaruur aayeNge.
 {as far as ki_16a his question be.PR he certainly come.FUT}
 ‘As far as he is concerned, he will certainly come.’

- (32) ki_16b: even
yaha savaala itanaa aasaana hE ki yahaaN tak ki baccaa bhii hal kara sakataa hE.
 {this question so easy be.PR ‘ki_default’ ‘ki_16b’ child also solve do can be.PR}
 ‘This question is so easy that even a child can solve (it).’

Rule16: a. *ki* = ki_16a when preceded by *jahaaN taka* (in non-correlative clause).

b. *ki* = ki_16b when *ki* follows a phrase *yahaaN tak*.

An example such as (34) below seems to have a similar pattern as in (32), but at close examination, we find that it is amenable to Rule6 (for relative-correlative constructions) mentioned above.

- (34) *vah vahaaN tak gayaa jahaaN tak ki aaja taka kooi nahiiN gayaa thaa.*
 {he there up to went as far as ki_16a today till nobody went be.PST}
 ‘He went that far as nobody had gone till today.’

3 Implementation Strategy

The ‘*ki*’ mapping rules as outlined above have been implemented in the Hindi-English machine translation system (Sinha, 2004) currently under development.

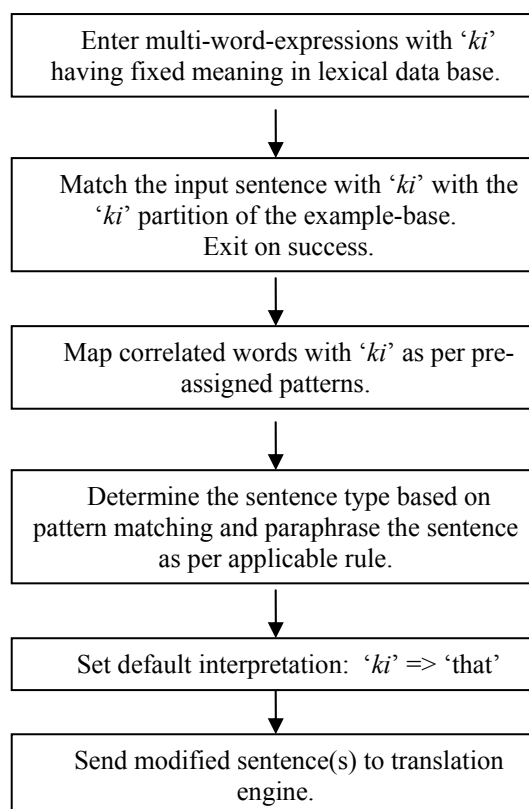


Figure 1: Implementation Strategy Flow-diagram

Our translation system uses a hybridized approach incorporating both the rule-based and example-based strategies. We use a lexical database which carries syntactic, semantic and ontological information of the lexicons along with their mappings to the target language forms. Rules are used primarily for extracting clauses/ chunks and for paraphrasing the input sentence. The core

of the translation engine uses a generalized Example-based approach.

Figure 1 depicts an outline of the basic strategy used.

The interpretation of different roles of the particle 'ki' is handled at different levels of the system implementation. We have not adopted a rule-by-rule implementation approach. Although, we have outlined various linguistic parameters in the rules, we use only those linguistic parameters that are essential for disambiguation. We take advantage of the rule 'by elimination' approach. Firstly, all those word combinations involving the particle 'ki' that have a fixed meaning are entered in the lexical data-base. In the input sentence, these word-groups are identified and marked. Secondly, the default meaning of 'ki' is set to 'that' in English which is the most common interpretation representing approximately 32% of usage in a sample of corpus. Next, we take all interpretations that emerge due to words co-related with 'ki' such as 'itanaa .. (jitanaa) ki', 'vah .. jo ki' (rule ki_6) etc. A pattern matching is performed on the input sentence to invoke an applicable 'ki'-mapping rule. The input sentence is transformed (paraphrased) as per the interpretation given in the rule to yield a new sentence where ambiguity pertaining to 'ki' is resolved. It may even fragment the input sentence into more than one sentence and the final translation is obtained by recomposing the translations of fragmented components. The paraphrasing may add and/or delete words/phrases so that the input sentence becomes more easily translatable. For example, for rule 'ki_3b', a word 'kyaa' (what) is inserted in sentence initial position to denote 'yes/no' type interrogative sentence and the particle 'ki' is replaced with its interpretation 'yaa' (or). In some cases, where rules are complex, matching with example-base is invoked to obtain the target pattern. For example rules 'ki_3a', 'ki_11', 'ki_14' and 'k_16' are handled through example-base.

4 Concluding Remarks

In this paper we have discussed the multiple roles in which the particle *ki* in Hindi is used and have examined the factors that determine its different mapping patterns in English.

For this, we examined a parallel Hindi-English corpus of approximately a size of 150,000 sentences. The corpus used was of a mixed nature consisting of texts from different sources such as short stories, fairy tales, essays, and also grammar and linguistics books. The system identified a total of 12000 occurrences of *ki* in this selected corpus. The frequency (in terms of rounded-off percentage) of the different types of *ki* in the

selected corpus is: ki_1: 22%; ki_2: 3%; ki_3: 4%; ki_4: 5%; ki_5: 22%; ki_6: 3%; ki_7: 8%; ki_8: 2%; ki_9: 5%; ki_10: 10%; ki_11: 3%; ki_12: 2%; ki_13: 2%; ki_14: 1%; ki_15: 6%; ki_16: 2%.

The different types of *ki* have been examined for their multiple uses and translation patterns in English. The judgment on the multiple interpretations of these sentences is based on both written text corpus and the native speakers' intuition. On the basis of the different linguistic markers available in a sentence, we have formulated strategies to identify the different functions and mapping patterns of *ki*. These rules have been used in our Hindi to English MT system under development (Sinha 2004). Although, a final evaluation of the system's accuracy in 'ki' interpretation will be available only after the entire MT system gets built, our preliminary evaluation of the rules by comparing with the available parallel translation and those generated through these rules, exhibit an accuracy of more than 90%.

The patterns discussed in this paper are only representative and the rules proposed here present the initial attempt in this direction. There may still be some more patterns not encountered in the selected corpus. To our knowledge, no work of this nature exists in literature on the Hindi-English MT or related literature.

It is interesting to note that most of the languages belonging to the Indo-Aryan family, have a particle with similar behaviour as that of 'ki' in Hindi. It is hoped that the analysis and strategies presented here will be relevant to these languages as well.

References

- K.P. Guru. 1925. *Hindivyakaran* (in Hindi). Nagari Paracarini Sabha, Varanasi.
- Yamuna Kachru. 1980. *Aspects of Hindi syntax*. Manohar, Delhi.
- Tara Mohanan. 1994. *Argument structure in Hindi*. CSLI Publications, Stanford, California.
- A. Sharma. 1958. *A Basic Grammar of Modern Hindi*. Central Hindi Directorate, New Delhi.
- R.M.K. Sinha. 2004. An Engineering Perspective of Machine Translation: Angla Bharti – II and Anu Bharti – II Architecture. In *Proceedings of International Symposium on Machine Translation NLP and TSS*, The McGraw-Hill Companies, New Delhi: 10-17.
- K.V. Subbarao. 1984. *Complementation in Hindi Syntax*. Academic Publications, Delhi.
- K.D. Vajpai. 1958. *Hindi Shabdanushasan* (in Hindi). Nagari Paracarini Sabha, Varansi.