ON TRANSLATION OF INTERROGATIVE SENTENCES FROM HINDI TO ENGLISH

Dr. R.M.K. Sinha
Indian Institute of Technology, Kanpur 208016 India
E-mail: rmk@iitk.ac.in

Dr. Anil Thakur
Banaras Hindu University, Varanasi India
E-mail: mi_anil@yahoo.com

ABSTRACT
Interrogation is marked by several devices in a natural language. Some of these devices are linguistically well-defined. However, many of them are not so well-defined and hence their dealing depends on several factors including some extra-grammatical factors. Furthermore, different languages adopt different strategies to mark different types of interrogation and there is no one-to-one map between strategies adopted in one language and those in another language. For instance, English uses different wh-words to denote content-type interrogative sentences whereas yes/no-type interrogation is denoted by auxiliary-fronting. In Hindi, on the other hand, an interrogative marker particle kyaa can be used to denote both content-type and yes/no-type interrogation besides several other roles. Likewise, some other markers of interrogation in Hindi are also used in multiple senses and have multiple mapping patterns in English. In this paper, we discuss different types of interrogative sentences in Hindi and formulate strategy for the purpose of their translation into English.

KEY WORDS
Machine translation, Natural language Processing, Sense disambiguation, Interrogative sentence, Hindi to English

1. Introduction

Interrogative sentences in natural languages are of different types and they are denoted by different devices. A question-mark ('?') denotes an interrogative sentence in written text. There cannot be any punctuation mark in speech and it is the stress pattern that may give clue to an utterance being interrogative. Although here we are only concerned with written text and it is easy to determine whether a sentence is interrogative, it is the type of interrogative sentence that becomes important in translation, question-answering and other applications. Two broad types of interrogative sentences across-languages are: content-question and yes/no-question. In English, content-question interrogative sentences are denoted by the use of different wh-words, which occur in the sentence-initial position whereas yes/no-question interrogative sentences are denoted by fronting of the auxiliary element [4, 7]. In both the cases the interrogation marking elements occur in the sentence-initial position. On the other hand, Hindi like other South Asian languages, is a relatively free word order language. Here the particles denoting interrogation can occur at different places in the sentence. Moreover, a Hindi particle *kyaa* is used to denote both the content-question type interrogative sentences and yes/no-question type interrogative sentences [8, 9].

For the purpose of machine translation from Hindi to English, we need to identify the types of the interrogative sentences in Hindi because different types of interrogative sentences have different mapping patterns in English. A few works on English to Hindi machine translation [2, 4] have been reported. However, there has been very little work on taking Hindi as a source language for machine translation to English and the aspect of interrogatives have not been explicitly considered in any system to the best of our knowledge. Levie et al. [6] describe an interesting experiment on Hindi-to-English machine translation, using a trainable transfer-based machine translator trained with limited aligned Hindi-English text. Divergence patterns for English-Hindi language pair have been presented in [2, 10] and divergence with respect to interrogative sentences have been pointed out. In this paper, we examine the different types of the interrogative sentences in Hindi and their mapping patterns in English. In section 2, we discuss the structure of different interrogative sentences in Hindi. In section 3, we discuss the disambiguation strategies used in our translation system. In section 4, we conclude the paper.

2. Structure of Interrogatives in Hindi

2.1 Content-Question Type

Content-question interrogative sentences are denoted by different interrogative words. It may be noticed that while the English interrogative words mostly begin with wh- (who, what, why, etc) and hence are called wh-words, their Hindi counterparts mostly begin with a *k-* (*kOn*, *kyaa*, *kyoN*, etc) and hence they can be called *k*-words in Hindi [1]. Below we examine some of the major interrogative words in Hindi with respect to their multiple uses and mapping patterns in English.

i. *kOn:* *kOn* has two types of uses in Hindi. In one, it is an interrogative pronoun in unmarked Case form that interrogates the human subject NP and can occur in different positions in a sentence (1a-c) and can also occur in a reduplicative form (1d) that denotes plurality. We look at the different forms of *kOn* in this section. In its second use, *kOn* is used as an interrogative adjective and is mapped in English by *which*.

(1) a. *kOn aayaa?* {who came} 'Who came?'

b. *kOn yah kaam kar-egaa?* {who this work do-FU}

c. *yah kaam kOn kar-egaa?* {his work who do-FU} 'Who will do this work?'

d. *mere saath kOn-kOn cal-egaa?* {my with which go-FU} 'Who (all among you) will come with me?'

In Hindi, the subject NP can occur in different Case forms and depending on that, interrogative subject pronoun *kOn* also occurs in different Case forms. However, all of them are mapped in English by *who* when they function as the subject NP. If they function as an object NP, they are mapped by *whom* and other relevant forms. *kisane* ('who') is the ergative Case form of *kOn* (2a) and *kisako* ('who') is the dative (subject) Case form of *kOn* (2b). Likewise, *kisase* ('who') (2c), *kisak-ii/-aa/-e* ('who') (2d), *kisameN* ('who') (2e) can also occur as interrogative subject pronoun. As we notice, they all are mapped by *who* in English. The identification of these subject interrogative pronouns is also dependent on the identification of subject and object NPs in a sentence. The oblique forms of *kOn* are *kis* (singular) and *kin* (plural).

(2) a. *kisane raam-ko maaraa?* {who Ram-ACC beat}
'Who beat Ram?'

b. *kisako jaanaa caahiye?* {who go-GER OBL}
'Who should go?'

c. *kisase calaa nahiiN jaataa?* {who walk not PASS}
'Who cannot walk?'

d. *kisakii ek hii santaan hE?* {who one FP issue be.PR} 'Who has only one child?'

e. *kisameN himmat nahiiN hE?* {who courage not be.PR} 'Who does not have courage?'

kOn and *kisane* are used exclusively in subject position in a sentence and are always mapped by *who* in English. However, *kisako* is largely used in object position and in that case, it is mapped by *whom/who*. *kisako* is always used for human object. In case of non-human object, *kyaa* is used as interrogative pronoun. Given below are some of the illustrative examples:

(3) a. *usane kitaab kisako dii?* {he book whom gave}
'Whom/who did he give the book?'

b. *usane vahaaN kisako dekhaa?* {he there whom saw}
'Whom/who did he see there?'

(4) a. *usane kisase baat kii?* {he whom talk did}
'With whom did he talk?'

b. *aapane use kisase maaraa?* {you him with what beat} 'With what did you beat him?'

(5) *aapako kisameN visvaas hE?* {you in whom faith be.PR} 'In whom do you have faith?'

ii. *kyaa*: As has been referred earlier, *kyaa* in Hindi is used both as an interrogative pronoun in content-question interrogative sentences and as a question particle in a yes/no-question interrogative sentences. In the former case, it is mapped by *what* whereas in the latter case, it is not directly mapped but different strategies are used to denote yes/no-question. Another use of *kyaa* is as an interrogative adjective. In this case, too, it is mapped by *what* but it has different structure than the one in the case of an interrogative pronoun. In fact, *kyaa* also has multiple functions (such as a marker of exclamation, temporal and negation adverbial, emphatic assertion, etc) other than these three interrogative functions but we cannot go into the details here due to constraint on space. However, it may be pointed out that in the identification task of the interrogative functions of *kyaa*, we need to differentiate its other functions, too. Below we examine the use of *kyaa* as an interrogative pronoun in content-question interrogative sentences (6).

(6) *kyaa:* Interrogative Pronoun (IP)

a. *kyaa ho rahaa hE?* [Subject]
{IP be PROG be.PR} 'What is happening?'

b. *aap kyaa caahate hEN?* [Object]
{you IP want be.PR} 'What do you want?'

As we notice, *kyaa* as an interrogative pronoun can occur both in subject and object positions. Hindi, like most of the South Asian languages and unlike English, has been categorized as an "in-situ" language in which the question word occurs in the same position as the NP which it interrogates, rather than moving to the sentence-initial position. For instance, unlike Hindi (8), in English (7), a question word is obligatorily fronted to the clause-initial position. The contrast is illustrated in (7b-8b).

(7) a. Ram ate a mango.
b. What did Ram eat?

(8) a. *raam-ne aam khaayaa*. {Ram-ERG mango ate}
'Ram ate a mango.'

b. *raam-ne kyaa khaayaa?* {Ram-ERG what ate}
'What did Ram eat?'

However, Hindi also exhibits variation with respect to word order of different constituents in a sentence. In Hindi, the object NP can also occur in the sentence-initial position and likewise, the interrogative pronoun in the object position can also often occur in the sentence-initial position. Thus we can have the following (9) variations of the example in (6b). Notice that change of meaning with change in the order of *kyaa* in (9). When it occurs between the main verb and the auxiliary, it is always an IP.

(9) a. *aap kyaa caahate hEN?* [from (6b)]
{you IP want be.PR}

b. *aap caahate kyaa hEN?* {you want IP be.PR}
'What do you want?'

c. *kyaa aap caahate hEN?* {IP you want be.PR}

d. *aap caahate hEN kyaa?* {you want be.PR QP/IP}
i. 'Do you want (this)?'
ii. 'What do you want?'

This shows that the word order and the position of occurrence of an interrogative particle is an important factor in the identification of the type of interrogation in Hindi.

iii. *kyoN/ kisaliye*: *kyoN* ('why') is an interrogative pronoun that denotes reason adverbial in Hindi, as in (10a-b). It can occur in different positions in a sentence but has single mapping as *why* in English. Thus it is not an ambiguous interrogative pronoun.

(10) a. *aap yahaaN kyoN aaye hEN?* {you here why come be.PR}

b. *aap kyoN yahaaN aaye hEN?* {you why here come be.PR}

'Why have you come here?'

iv. *kEs-aa/-ii/-e (kis tarah)*: *kEs-aa/-ii/-e* is used to denote interrogative manner adverbial in Hindi. *kEs-aa/-ii/-e* is used as an interrogative adjective also (see 2.3 iii) below.

(11) a. *yah kEse ho-gaa?* {this how be-FU}

'How will this happen?'

b. *vah kEsaa khelataa hE?* {he how play be.PR}

'How does he play?'

c. *aap kEse aaye?* {you how came}

'How did you come?'

d. *aam kEse de rahe ho?* {mango how give PROG be.PR} 'At what price are you selling the mangos?'

Besides the major interrogative particles examined above, there are also some other interrogative adverbial elements such as (*kab, kahaaN, kidhar,* etc) in Hindi which are usually unambiguous and hence are easily identifiable except in cases of idiomatic uses. In the following section, we examine the structure of the yes/no-question interrogative sentences in Hindi.

2.2 Yes/No-Question Type

Yes/no-question interrogative sentences in Hindi are formed by different methods. Besides *kyaa*, which is the most common yes/no-question marker particle in Hindi, there are other devices. As we have referred above, *kyaa* is also used as an interrogative pronoun and as an interrogative adjective. The contexts of these different uses of *kyaa* need to be identified for the correct mapping of *kyaa* in English. We examine some of the representative examples of yes-no type of sentence constructions and their mapping to English.

i. *kyaa:* *kyaa* as QP has different structural patterns (related to the position of occurrence) in different types of sentences such as transitive sentences (12), intransitive sentences (13), and copula sentences (14). It can occur in different positions making it highly ambiguous. For instance, in (13), if we change the position of *kyaa* from sentence-initial or sentence-final to a sentence-medial position (for instance, at the second place in the sentence), the sentence becomes a content-question interrogative sentence.

(12) a. *aap kyaa vahaaN gaye the?* {you QP there went be.PST}

b. *kyaa aap vahaaN gaye the?* {QP you there went be.PST}

c. *aap vahaaN gaye the kyaa?* {you there went be.PST QP} 'Did you go there?'

(13) a. *kyaa raam paDh rahaa hE?* {QP Ram study PROG be.PR} 'Is Ram studying?'

b. *raam paDh rahaa hE kyaa?* {Ram study PROG be.PR QP} 'Is Ram studying?'

(14) a. *raam kyaaa vidyaarthii hE?* {Ram QP student be.PR} 'Is Ram a student?'

b. *dilli kyaa ek shahar hE?* {Delhi QP a city be.PR}

i. 'Is Delhi a city?'

ii. 'What a city is Delhi!'

ii. *ki*: *ki* in one of its multiple uses also marks (alternative) interrogative sentence of yes/no-question type, as in (17).

(15) a. *ve aa-yeNge ki nahiiN?* {they come-FU QP not}

'Will they come or not?'

b. *aapane use dekhaa hE (yaa) ki sunaa hE?*

{you him saw be.PR or QP heard be.PR}

'Have you seen him or heard him?'

iii. (*hE*) *na*: *hE na* is used as a question tag marker in Hindi [9]. It is the counterpart of the different question tag marking strategies in English. Some examples are presented in (16).

(16) a. *aap vahaaN gaye the na?* {you there went be.PR not} 'You went there, didn't you?'

b. *aap yah kitaab paDh rahe hEN na?*

{you this book read PROG be.PR not}

'You are reading this book, aren't you?'

2.3 Interrogative Adjectives

Interrogative sentences also include those sentences that have an interrogative adjective in them. Some of the interrogative pronouns are also used as interrogative adjectives and to identify the contexts of their different uses becomes important for the purpose of mapping them in English. Some examples are presented below:

i. *kOn* or ***kOn-s-aa/-ii/-e*:** *kOn* or *kOn-s-aa/-ii/-e* is also used as an interrogative adjective and is mapped by *which* or *which one* in English. *kOn* also occurs in reduplicative form *kOn-kOn* and in this case it denotes plural meaning.

(17) a. *kOn-sii kitaab acchii hE?*

{IA book good be.PR} 'Which book is good?'

b. *aap-ko kOn-saa makaan pasand hE?*

{you-DAT IA house like be.PR}

'Which house do you like?'

c. *aap-ko kOna-kOnasii kitaabeN caahiye?*

{you-DAT which books want}

'Which books do you want?'

ii. *kyaa*: *kyaa*, besides being an interrogative pronoun (see example (8) above) and a question particle (see example (14-16) above), is also used as an interrogative adjective in Hindi. *kyaa* as IA has two types of mapping patterns in English; *what* (18a) and *what kind of* (18b). In the latter case, *kyaa* is similar to (and seems to be a replacement of) *kEs-aa/-ii/-e* ('what kind of').

(18) a. *aapakaa naam kyaa hE?*

{your name IA be.PR} 'What is your name?'

b. *aapako kyaa kaam pasand hE?*

{You-DAT what work like be.PR}

i. 'What kind of work do you like?'

ii. 'Do you like the work?'

iii. *kEs-aa/-ii/-e*: *kEs-aa/-ii/-e* is also used as an interrogative adjective in Hindi and agrees with the head noun. It is mapped by '*what kind/type of*' phrase. There are several factors that need to be examined to identify the correct context of this interrogative particle, such as gender, number and person information, the nature of the element that immediately follows it in a sentence and also the type of sentence. For instance, when *kEs-aa/-ii/-e* is

immediately followed by a verb (in a sentence other than a copula verb sentence), it is an adverb. In (19), *kEs-aa/-ii/-e* occurs as an interrogative adjective. *kEs-aa/-ii/-e* as IA is similar to *kis taraha k-aa/-ii/-e* 'what kind/type of'. In both adverbial and adjectival uses, *kEs-aa/-ii/-e* occurs in reduplicative forms.

(19) a. *aapake paas kEsii kaar hE?* {you with IA car be.PR} 'What kind of car do you have?'

b. *aapakaa makaan kEsaa hE?* {your house IA be.PR} 'What type of house do you have?'

c. *vahaaN ke log kEse hote hEN?* {there of people IA be be.PR} 'What kind of people are/live there?'

iv. *kitan-aa/-ii/-e*: This is an interrogative quantitative adjective (agrees with the GNP of the head noun) and is mapped by *how many* (20a) or *how much* (20b) or *what* (20c) in English depending on the nature of the head noun.

(20) a. *kitane log aaye the?* {how many people come be.PR} 'How many people had come?'

b. *aapako kitanii miThaaii caahiiye?* {you-DAT how much sweets want} 'How much sweets do you want?'

c. *unake aane kii kitanii sambhavanaa hE?*
{his coming of what possibility be.PR}
'What is the possibility of his coming?'

v. *kis*: *kis*, besides being an interrogative pronoun (oblique form) is also used as an interrogative adjective and gets mapped by *which* in English. The plural form of *kis* is *kin*.

(21) a. *kis aadamii ko aapane bulaayaa hE?*
{which man ACC you called be.PR}
'Whom have you called?'

b. *kin logoN ko aapane bulaayaa hE?*
{which people ACC you called be.PR}
'Which persons have you called?'

3. Disambiguation strategy

From the foregoing discussions, it is evident that Hindi uses a number of interrogative particles have multiple functions and their contexts need to be disambiguated for their correct mapping/translation in English. Our Hindi to English machine translation system uses a hybrid strategy with hybridization of rule-based and example-based approaches. During the development phase, rules are used to develop an interpretation and if it fails then the sentence pattern is entered into the example-base for further processing. During translation phase, it is the example-base which is matched first and when no match is found, rule-base is invoked.

We use a shallow parser for analyzing Hindi sentences. The input sentence is grouped into logical grammatical chunks such as noun-phrase, verb-phrase, adverbials etc. We use a Hindi lexical data-base, a Hindi morphological analyzer and a set of rules for chunk formation. These chunks are analyzed for interpretation. We have formulated a number of rules for interpretation of the role of the interrogative particles based on the nature of the identified chunks. Where it is not possible to formulate rules, examples with chunks are used. The detailed strategy has been omitted due to constraints on space.

4. Concluding Remarks

In this paper, we have discussed different types of interrogative sentences in Hindi with a view to devise strategies for their identification and disambiguation. We have also discussed their mapping/translation patterns in English. For this, we examined a parallel Hindi-English corpus of approximately a size of 150,000 sentences. The corpus is 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 3029 interrogative sentences in this selected corpus. These sentences have been examined for their multiple usage and translation patterns in English. The judgment on the multiple interpretations of various interrogative sentences is based on both written text corpus and the native speakers' intuition. On the basis of different linguistic markers available in a sentence, we have formulated strategies to identify the nature of interrogative sentences and also disambiguate their different mapping patterns in English. Our strategy yielded an accuracy of 91% tested over the selected corpus. Most of the errors have been in interpretation of the particle *kyaa*.

The results presented in this work forms part of the Hindi-English machine aided translation system. The strategy outlined here are applicable to most of the languages of Indo-Aryan family and other languages of South Asia that have relatively free word-order.

Abbreviations/Symbols: ACC: Accusative, **DAT:** Dative, **FU:** Future Tense, **GNP:** Gender, Number, Person, **IA:** Interrogative Adjective, **IP:** Interrogative Pronoun, **QP:** Question Particle, **PASS:** Passive Particle, **PR:** Present Tense, **PROG:** Progressive Aspect, **PST:** Past Tense

References:

[1] Veneeta Dayal. Locality in WH Qantification: Questions and Relative Clause in Hindi. Kluwer Academic Publishers. (1996).

[2] S. Dave, J. Parikh and P. Bhattacharyaa. Interlingua-based English-Hindi Machine Translation and Language Divergence. Machine Translation 16(4) (2001) 251-304.

[3] Rodney Huddleston and Geoffrey Pullum. The Cambridge Grammar of the English Language. Cambridge University Press, UK. (2002).

[4] http://shakti.iiit.net/

[5] Yamuna Kachru. Aspects of Hindi Grammar. Manohar, Delhi. (1980).

[6] Alon Lavie, S. Vogel, L. Levin, E. Peterson, K. Probst, A. F. Llitjós, R. Reynolds, Jaime Carbonell, R. Cohen. Experiments with a Hindi-to-English transfer-based MT system under a miserly data scenario, ACM Transactions on Asian Language Information Processing (TALIP), 2(2):143-163. (2003).

[7] Randolph Quirk, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. A Comprehensive Grammar of the English Language. Longman, London. (1985).

[8] Michael C. Shapiro. A Primer of Modern Standard Hindi. Motilal Banarasidass, Delhi. (1989).

[9] Ayendra Sharma. A Basic Grammar of Modern Hindi. Central Hindi Directorate, Government of India. (1958).

[10] R.M.K. Sinha and Anil Thakur, Divergence Patterns in Machine Translation between Hindi and English, 10th Machine Translation summit (MT Summit X), Phuket, Thailand, September 13-15, (2005), 346-353.