

# WSD of To-Infinitive into Hindi : An Information Based Approach

Akshar Bharati  
R Vaishnavi Rao  
Rashtriya Sanskrit Vidyapeetha  
Tirupati A P  
r.vaishnavi.rao@gmail.com

[Presented at Platinum Jubilee Conference of Linguistic Society of India,  
Central University Campus, Hyderabad, 6-8 Dec 2005]

Word Sense Disambiguation (WSD) is a major problem in Machine Translation (MT). There have been several attempts to handle WSD<sup>1</sup>. To develop WSD rules manually is laborious and time consuming. Moreover, if the rules are developed for bilingual WSD, they may be language-pair specific. If the rules are monolingual, it is difficult to decide the granularity for different senses. While the statistical method may be helpful in handling large volumes of language data, it does not give any linguistic insight about the languages involved in MT. There have been attempts to semi-automate the task of WSD<sup>2</sup>. However the rules which the machine learns are in hundreds and it becomes again difficult to gain any linguistic insight from these methods.

Rule-based WSD, on the other hand, helps in linguistic analysis, but mostly works with the syntax of a sentence and is hence surface structure dependent. Moreover, in this method, rules are written to suit the systems running on available technology and may have to be changed if a better technology comes about. Of course, we cannot do without writing rules for WSD for running the MTS, but a deeper approach for language analysis dealing the language semantically, which would enlighten us about where the information about the language phenomenon is available would be of greater advantage. The results of such analysis can be always used along with further development in technology as the analysis is independent of technological constraints. Such a method of WSD may be called as Information-based approach. We illustrate information-based WSD with an example of sense disambiguation of to-infinitive in English into Hindi. First we look at few English sentences with to-infinitives and their Hindi

translations. For example –

E: I would love to live in New York.

H: maiM New York meM rahanA pasAMda karUMgI.

E: He tried to leave quietly.

H: usane chupachApa jAne kA prayatna kiya.

E: I am going there to see my sister.

H: maiM apanI bahana ko dekhane ke liye vahA.N jA rahI hU.N. E: She refused to accept that there was a problem.

H: usane mAnane se inakAra kara diyA ki koI samasyA thI.

E: She is the next person to speak.

H: vaha hai agali bolane vAllI vyakti.

From the above illustrations we can see that to-infinitive has more than one mapping in Hindi. English has a single infinitive marker 'to' whereas Hindi requires a marker 'nA' optionally followed by nominal suffixes (vibhaktis) such as kA, ke\_liye, se, vAllI, etc. The kAraka relation of the to-infinitive with the other words in the sentence is not explicitly known from the surface of the sentence in which the to-infinitive is used. However, Hindi requires an explicit vibhakti to specify the kAraka relation of the infinitive with the other words in the sentence. Realisation of this fact that English does not code something which Hindi requires explicitly is an important step in information-based approach. So as a first cut, we may say that to-infinitive can be translated into Hindi as 'nA \*' where '\*' stands for the parasargas (post-positions) such as '0', 'kA', 'ke liye' etc. Thus, English is leaving something unspecified and Hindi obligatorily requires the unspecified information for the sentence to be grammatical.

From this information we would first try to abstract what is common among the disparities and heuristically make a rule for MT. In this case we have fixed that to-infinitive may be translated, giving room for a few exceptions, as 'nA\*'. This information may be used further for making rules for WSD and where disambiguation becomes too obscure, this minimal information may give a broad hint about the sense of to-infinitive in the English sentence, which the user may be able to interpret from the expectancy (AkA.NkshhA) etc., context, world knowledge and so on. With this information we can make rules for WSD of the to-infinitive for humans by further semantic analysis, if not for machines immediately. This information may pay us in the long run; as and when we get more data in this regard; as and when technology would support us to use such rules of WSD. While trying to make rules of WSD for humans from the available information, the

question is, is it possible to "guess" the unspecified information just from the language strings? It is here that the Paninian model gives us insight about how to frame the rules for a language phenomenon keeping in mind both its syntactic and semantic considerations. Panini has explicitly specified the semantics of infinitives in his monumental work ashhTAdhyAyI<sup>3</sup>. We observe that some of the cases of the English to-infinitive may be translated as the affix 'tumun' in Sanskrit. Therefore we look at the rules dealing with tumun from ashhTAdhyAyI to gain further insight.

There are five rules in the ashhTAdhyAyI stating the meaning of 'tumun'. They are as follows:

1. tumuNNvulau kriyAyAM kriyArthAyAm [bhavishhyati kAle](3-3-10).  
'tumun' and 'Nvul' are the affixes attached to the verb denoting an action which is to happen in the FUTURE and which is the PURPOSE of another action.

S: kR^ishhNaM drashhTuM yAti.

E: goes to see Krishna.

2. samAnakartR^ikesshu tumun [ichchhArtheshhu](3-3-158).

'tumun' is affixed to the verb where the other verb denotes the action of DESIRE and where the kartA (AGENT) IS SAME FOR BOTH ACTIONS.

S: bhoktum ichchhati.

E: wishes to eat.

3. shaka-dhR^ishha-GYA-glA-ghaTa-rabha-labha-krama-saha-arha-asti-artheshhu tumun (3-4-65).

'tumun' is affixed to the verb in a sentence where the other verb is from the list of shaka etc. (This list is specific to Sanskrit language).

S: bhoktuM shaknoti.

E: can eat.

4. paryAptivacaneshhu alamartheshhu [tumun](3-4-66).

'tumun' is affixed to the verb in a sentence where the related word is 'paryApta' etc. that occur in the sense of CAPABILITY.

S: bhoktuM paryAptaH.

E: capable to eat.

5. kAlasamayavelAsu tumun (3-3-167).

'tumun' is affixed to the verb in a sentence where the related words such as kAla etc., referring to TIME, occur. Here the time to perform an action is indicated.

S: bhoktuM kAlaH.

E: time to eat.

Here we see that except in rule 3, which only lists the verbs with which 'tumun' can occur, all the rules describe the context and semantics of the affix 'tumun'. 'tumun' can occur as a modifier of a verb (as in 1 and 2), or forming a verb group (as in 3), or modifying a noun or an adjective (as in 4 and 5).

An attempt to write the rules of WSD of the to-infinitive into Hindi on the same lines actually throws light on the semantics of the to-infinitive, which is to be otherwise inferred from the context in which it is used. To put it explicitly, in the effort to get the information about the contexts in which Hindi takes a particular vibhakti for the otherwise inexpressive to-infinitive, we get to know as to what are the various meanings of the to-infinitive. The various senses of the to-infinitive can be broadly classified into the following categories –

1. To-infinitive semantically connected to another verb
2. To-infinitive semantically connected to a noun or an adjective,
3. ECM constructions
4. To-infinitive forming part of a verb-group, as in 'have to go'.
5. Exception of translation of to-infinitive into nA\*

Let us discuss the WSD of the to-infinitive based on the above classification. 1. To-infinitive semantically connected to another verb –

a. To-infinitive denoting the purpose as in

E: I am going there to see my sister.

H: maiM apanI bahana ko dekhane ke liye vahA.N jA rahI hU.N.

Hindi uses 'ke\_liye' to indicate the purpose, and hence in such cases, to-infinitives is to be translated as 'nA~ke\_liye'.

b. To-infinitive as a modifier of the main verb meaning 'desire' -

E: I would love to live in New York.

H: maiM New York meM rahanA pasaMda karUMgI.

Here the object of the main verb 'love' is another action itself; in this case 'living in New York'. Here, the non-finite verb indicates only action (bhAva). Hindi uses 'nA' to indicate the bhAva and hence in such cases, to-infinitive is to be translated as 'nA'.

c. To-infinitive modifying the main verbs such as 'refuse', 'fear', 'loath' etc. -  
E: She refused to accept that there was a problem.

H: usane mAnane se inakAra kara diyA ki koI samasyA thI.

We can notice that the parasarga 'se', which indicates paJNchamI vibhakti (ablative case) is occurring where the main verb has the sense of disapproval to it. Two sUtras in ashhTAdhyAyi could explain this phenomenon. 'dhruvam apAye apAdAnam (1-4-24)' and 'apAdAne paJNchamI (2-3-28)' explain that, that which is the source of separation is apAdAnam; and apAdAnam takes paJNchamI vibhakti. In the example, mental separation from the act of 'acceptance' is intended to be conveyed by the verb 'refuse', which makes the act of 'acceptance' the 'apAdAna'. Hence, it takes paJNchamI vibhakti. The phenomenon of mental separation is common to all the verbs indicating the sense of disapproval. Where the source of mental separation is another activity indicated by an infinitive, it takes paJNchamI vibhakti in Hindi.

d. To-infinitive related with verbs whose Hindi translations are 'kriyAmUla'-

E: He tried to leave quietly.

H: usane chupachApa jAne kA prayatna kiya.

The verbs such as 'promise', 'try' etc. are translated into Hindi as 'vAdA karanA', 'prayatna karanA' etc. These verbs are technically called as 'kriyAmUla' verbs which retain an element of nominal tendency. Here the infinitive is literally connected with the noun such as 'vAdA' etc. In Hindi the relation between nouns is indicated by shhashhThI vibhakti with the parasarga 'k[AI]' as in 'rAma kA kalama'. Hence, it acquires shhashhThI vibhakti.

2. To-verb semantically connected to a noun/adjective:

a. To-infinitive modifying the words indicating 'time' –

E: It is time to go home.

H: yaha ghara jAne kA samaya hai.

This instance is covered by the Paninian rule 'kAlasamayaveLAsu tumun' (3-3-167). Here the infinitive describes the 'time' and literally acts as an adjective to it. In Hindi the relation between an adjective and its corresponding noun is indicated by shhashhThI vibhakti with the parasarga 'k[AI]', as in case of 'safeda raMga kI chAdara'. Hence, the infinitive takes shhashhThI vibhakti here.

b. To-infinitive modifying the nouns such as 'ability', 'tendency', 'opportunity' etc.

E: I have a tendency to tease.

H: mujhe chheDane kA svabhAva hai.

The phenomenon explained in (2a) applies here also. This instance is covered by the Paninian rule 'paryAptivachaneshhu alamartheshhu tumun' (3-4-66).

c. to-infinitive related with adjectives such as 'afraid' etc. –

E: I was afraid to go home.

H: maiM ghara jAne se bhayabhIta thI.

The same phenomenon mentioned in rule (1c) above applies here also. Here, however mental separation is not indicated by a verb with a sense of disapproval, but by the adjective 'afraid'. Here one more sUtra, viz., 'bhItrArthAnAM bhayahetuH [apAdAnam](1-4-25)' may be quoted in support of the occurrence of paJNchamI vibhakti. According to the sUtra the cause of fear is apAdAna in case of words used in the sense of 'fear'. In the example, the act of 'going' is the cause of fear indicated by the word 'afraid'. Hence, it is the apAdAna and it takes paJNchami vibhakti.

d. To-infinitive modifying the nouns –

E: She is the next person to speak.

H: vaha hai agalI bolane vAlI vyakti.

Here the infinitive actually modifies the noun, which is the agent of the action denoted by the non-finite verb. This is indicated by the affix 'vAl[AI]' in Hindi as in case of 'jAnevAlI gADI' etc.

e. To-infinitive related with adjectives such as 'anxious' –

E: I am anxious to know.

H: maiM jAnane ke liye utsuka hU.N.

Here the adjective 'anxious' has the expectancy 'for what' (tAdarthyā). Hindi takes the chaturthI vibhakti (with the parasarga as 'ke liye') for expressing 'tAdarthyā' as in the case of 'laDake ke liye miThAI'. Hence the infinitive takes 'nA~ke\_liye' in this case.

f. To-infinitive related with the adjective modified by 'too' –

E: It was too hot to go out.

H: bAhara jAne ke liye kuchha zyAdA hI garmI thI.

The same phenomenon as in (2e) applies here.

g. To-infinitive related with adjectives such as 'easy', 'good' etc. –

E: The path was easy to follow.

H: mArga anusaraNa karane meM AsAna thA.

Though the same phenomenon as in (2e) and (2f) are applicable here and hence the translation 'mArga anusaraNa karane ke liye AsAna thA' would not be wrong, it is the idiosyncrasy of the Hindi language to use the parasarga 'meM' in case of adjectives such as 'AsAna', 'achchhA' etc.

3. ECM constructions –

Certain verbs, such as 'want', exhibit a special behaviour in English.

For example, in the sentence -

E: I want him to go.

'He' which is in the subject position of the verb 'go' gets case assigned by the verb 'want'. Here the kartA of the verb 'want' is 'I' and the kartA of the verb 'go' is 'he'. However, in Hindi, as in Sanskrit, the verbs denoting desire share their agents with the infinitive. This poses a problem from translation point of view which is discussed elsewhere<sup>4</sup>. We can notice that English also allows sentences with the main verb denoting 'desire' sharing its agent with the infinitive, as in -

E: I want a pen to write.

The translation for this follows from rule (1a) and hence the translation for this sentence would be -

H: maiM likhane ke liye kalama chAhatI hU.N.

whereas, the translation of the above sentence will be -

H: maiM usakA jAnA chAhatI hU.N.

Hence we can see clearly that the ambiguity is difficult to resolve in such cases and some amount of knowledge of context is required to disambiguate the senses of these sentences. Hence, it is left to the user to decide the sense of the sentence in such cases.

4. To-infinitive forming part of a verb-group –

These are to be treated case by case. Hindi equivalents of each of the tam groups e.g. have\_to\_0, ought\_to\_0, etc. need to be given separately.

E: I have to go home.

H: mujhe ghara jAnA hai.

E: That child ought to be in bed.

H: usa bachche ko bistara meM honA chAhiye.

5. Exception of translation of to-infinitive into nA \* -

There are a few cases which are exceptions to the default rule that to-infinitive is 'nA \*'.

a. To- infinitive related to adjectives such as 'glad', 'sorry' etc. –

E: I am glad to hear that news.

H: maiM vaha khabara sunakara khush huI.

This is an exception to the default rule accepted by us that to-infinitive is to be translated as 'nA \*' in Hindi. On observing the context in the example we can see that the act of being glad follows the act of hearing. The temporal precedence (pUrvakAlikatva) of the action denoted by the main verb in comparison with the action denoted by the infinitive is indicated by the verbal affix 'kara' in Hindi.

b. To-infinitive in case of raising phenomenon –

In case of raising phenomenon, English uses 'to' to convert the sentence with finite verb into a clause as in –

E: It seems that John has left.

By dropping 'it' the sentence 'John has left' is converted to a clause as in –

E: John seems to have left.

Hindi translation of this sentence, close to the English structure would be –



H: jA.cna, aisA lagatA hai ki, chalA gayA hai.

Here, as such 'to' does not have any sense of its own. The presence of 'to' after 'seem' type of words seems to trigger the commas. Further reference is available elsewhere<sup>4</sup>.

## Implementation and Results

The above rules have been implemented in the current E-H anusaaraka. 140 sentences were chosen covering different phenomenon stated above. Following table shows the performance of the system.

Translation of to-infinitive	Total instances	Correct WSD by system	%
ne ke liye	54	50	92.6
nA	31	27	87.1
ne kA	12	12	100
ne kI	3	3	100
ne se	3	3	100
ne meM	5	4	80
ne ke liye kuchha jZyAdA hI	3	3	100
TAM	27	16	59.3
Kara	2	2	100
Total	140	120	85.7

Further, another set of 35 randomly selected sentences with to-infinitive were tested, and 30 of them were found to be translated correctly.

## Conclusion

In the process of WSD of to-infinitive into Hindi we can observe that the vibhaktis in Hindi make the semantics of the infinitive quite transparent and explicit which is otherwise implicit in case of English. This is an example to show how comparison of different language structures can enlighten us more about the semantics of the languages involved. We can also observe that Paninian grammar serves two purposes for language analysis.

1. The Paninian model serves as an ideal with regard to comprehensive

presentation of systematically analysed semantic and syntactic data of language and thereby acts as a good tool for Information-based analysis of language data.

2. The Paninian grammar explains language phenomenon clearly, thereby acting as a tool for problem solving in language.

From the machine point of view, while trying to lead rule-based WSD to utmost possible perfection, mainly in case of words with high magnitude of ambiguity, we face a problem of managing and maintaining the rules. It becomes quite a complex exercise to modify any existing rule when its exception is found. With the information-based approach we get a good orientation about dealing with the rules given even for machines as we have a strong foundation of justified and logical causes for deriving such rules. This makes even voluminous rules fairly manageable.

## Appendix

Following are the rules for WSD of to-infinitive into Hindi for the E-H anusAraka.

S.No	Condition	Hindi meaning
1.	from_list(previous_root,subj_subj_raising_verb.dat) #John seems to have left.	-
2.	from_list(previous_verb,ecm_verb.dat) && following_category=v-> # I want him to pass the exam	{tam:nA}
3.	from_list(previous_root,begin_to.dat) && following_category=v-> # Anne began to cry.	{tam:nA}
4.	previous_word=going-> #He is going to start it.	{tam:ne}
5.	from_list(previous_root,ability_to.dat) && following_category=v -> #He has ability to cross the river.	{tam:ne~kI}
6.	from_list(previous_verb,promise_to.dat) && following_category=v -> #They promised her to go there.	{tam:ne~kA}

7. from\_list(previous\_root,fear\_to.dat) &&  
following\_category=v -> {tam:ne~se}  
#They fear to go there.
8. from\_list(previous\_word,afraid\_se\_to.dat) &&  
following\_category=v -> {tam:ne~se}  
#I was afraid to go home.
9. from\_list(previous\_word,happy\_kara\_to.dat) &&  
following\_category=v ->  
{tam:0\_kara}  
#We were glad to see him.
10. default\_sense && following\_category=v-> {tam:ne~ke\_liye}

## References

1. Thesis of David Mowatt: Semantics and Machine Translation: What types of Semantic Information are necessary for Lexical Disambiguation for Machine Translation?
2. Kilgarriff Adam, Tugwell David: WASP Bench A Machine Translation Lexicographer's Workstation Supporting the State of the Art Lexical Disambiguation, Machine Translation Summit VII, Santiagode Compostela, September 2001.
3. Panini, ashhTAdhyAyI.
4. Bharati Akshar, Amba Kulkarni, English from Hindi Viewpoint, LSI Platinum Jubilee, Central University, Hyderabad 2005.

To-infinitive

nA \*

Exceptions

0  
ke liye  
se  
k[AI]  
vAlA  
meM

kara  
others