WSD of To-infinitive into Hindi : An Information Based Approach

Akshar Bharati
R Vaishnavi Rao
Rashtriya Sanskrit Vidyapeetha
Tirupati A P

Overview and Background of Information Based Approach

- Rules of bilingual WSD language-pair specific
- Rules of monolingual WSD – difficult to decide the granularity for different senses

Overview and Background of Information Based Approach

- Statistical WSD lacks linguistic insight
- Semi-automated WSD – machine learns many rules & difficult to gain linguistic insight

Overview and Background of Information Based Approach

- Rule based WSD helps in linguistic analysis
- but laborious, time consuming and dependent on
  - surface structure
  - available technology

Overview and Background of Information Based Approach

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 constrains. Such a method of WSD may be called as Information-based approach.

Illustration of information based WSD with an example of sense disambiguation of to-infinitive in English into Hindi

Hindi translation of English sentences with to-infinitive

1. E: I would love to live in New York.
   H: maiM New York meM raha**nA** pasaMda karUngI.
2. E: He tried to leave quietly.
   H: usane chupachApa jAne kA prayatna kiyA.
3. E: I am going there to see my sister.
   H: maiM apane bahana ko dekhane ke liye vahA.N jA rahI hU.N
4. E: She refused to accept that there was a problem.
   H: usane mAnane se inakAra kara diyA ki koI samasyA thI.
5. E: She is the next person to speak.
   H: vaha hai agalI bolane vAlI vyakti.

Observations

- 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, vAlI, etc.

Observations

- To-infinitive - kAraka relation with other words not explicitly known from surface of sentence.
- 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

Heuristic rule as a first cut

English	Hindi
To-infinitive	nA *[1]

* - Vibhaktis such as 0, ke liye, se, k[AI], meM, vAl[AI].

1 - with a few exceptions.

Advantage of this information

- Useful for framing further rules.
- If disambiguation too obscure minimal information for the user to interpret with the help of expectancy (AkA.NkshhA), context, world knowledge and so on.
- Rules for humans, if not for machines immediately.

Advantage of this information

- It 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.

Paninian model of language analysis and rule framing gives the guiding principles
Rules of ‘tumun’ – an instance

Rules of ashhTAdhyAyI stating meaning of ‘tumun’

- tumuNNvulau kriyAyAM kriyArthAyAM
- samAnakatR^ikeshhu tumun
- shhaka-dhR^ishha-GYA-glA-ghaTa-rabha-labha-krama-saha-arha-asti-artheshhu tumun
- paryAptivachaneshhu alamartheshhu
- kAlasamayavelAsu tumun

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^ishNam drashhTum yAti.

E: goes to see Krishna.

samAnakartR^ikeshhu tumun [ichchhArtheshhu] (3.3.158)

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

S: bhoktum ichchhati.

E: wishes to eat.

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 shhaka etc. (This list is specific to Sanskrit).

S: bhoktum shaknoti.

E: can eat.

paryAptivachaneshhu 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.

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.

Context and semantics of ‘tumun’

- Rule 1 & 2 – ‘tumun’ as modifier of a verb
- Rule 3 – ‘tumun’ forming a verb group
- Rule 4 & 5 – ‘tumun’ modifying a noun or an adjective

Senses of to-infinitive

- To-infinitive semantically connected to another verb
- To-infinitive semantically connected to a noun or an adjective
- ECM constructions
- To-infinitive forming part of a verb-group as in ‘have to go’
- Exceptions of translation of to-infinitive into nA *

To-infinitive semantically connected to another verb

a. To-infinitive denoting purpose

E: I am going there to see my sister.

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

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

E: I would love to live in New York.

H: maiM New York meM rahanA pasaMda karUMgI.

To-infinitive semantically connected to another verb

c. To-infinitive modifying main verbs such as 'fear', 'refuse', 'loath' etc.

E: She refused to accept that there was a problem.

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

d. To-infinitive related with kriyAmUla verbs

E: He tried to leave quietly.

H: usane chupachApa jAne kA prayatna kiyA.

To-infinitive semantically connected to nouns/adjectives

a. To-infinitive modifying words indicating 'time

E: It is time to go home.

H: yaha ghara jAne kA samaya hai.

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

E: I have a tendency to tease.

H: mujhe chheDane kA svabhAva hai.

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

E: I was afraid to go home.

H: maiM ghara jAne se bhayabhIta thI.

d. To-infinitive modifying nouns

E: She is the next person to speak.

H: vaha hai agalI bolane vAlI vyakti.

To-infinitive semantically connected to nouns/adjectives

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

E: I am anxious to know.

H: maiM jAnane ke liye utsuka hU.N.

f. To-infinitive related with adjectives modified by 'too'

E: It was too hot to go out.

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

g. To-infinitive with adjectives 'easy', 'good' etc.

E: The path was easy to follow.

H: mArga anusaraNa karane meM AsAna thA.

ECM constructions

E: I want him to go.

H: maiM usakA jAnA chAhatI hU.N.

To-infinitive forming part of a verb-group as in ‘have to go’

E: I have to go home.

H: mujhe ghara jAnA hai.

E: That child ought to be in bed.

H: usa bachche ko bisatara meM honA chAhiye.

Exceptions of translation of to-infinitive into nA *

a. To-infinitive related with adjectives 'glad', 'sorry' etc.

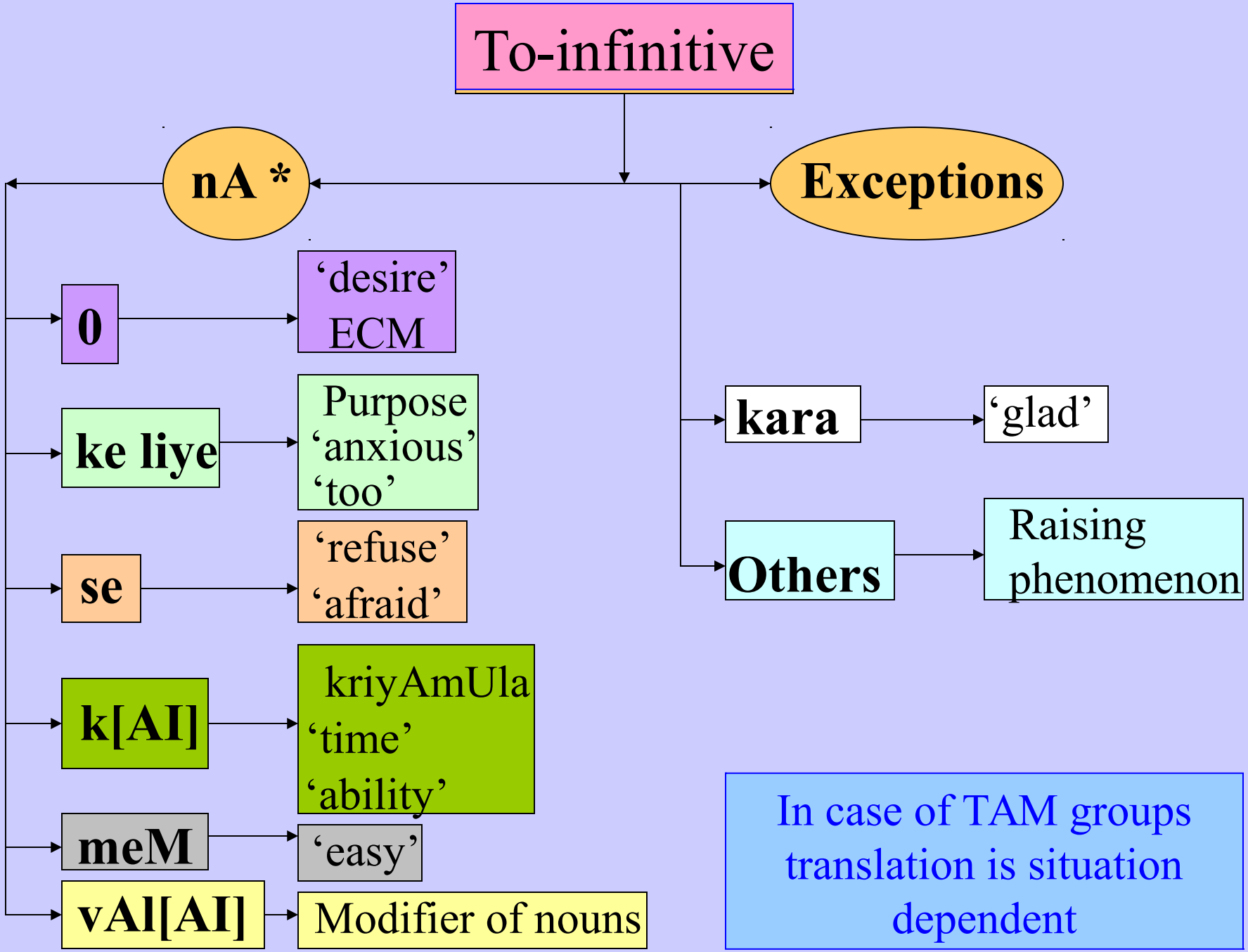
E: I am glad to hear that news.

H: maiM vaha khabara sunakara khusha huI.

b. To-infinitive in case of raising phenomenon

E: John seems to have left.

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


To-infinitive
nA *
Exceptions
0
'desire'
ECM
ke liye
Purpose
'anxious'
'too'
se
'refuse'
'afraid'
k[AI]
kriyAmUla
'time'
'ability'
meM
'easy'
vAl[AI]
Modifier of nouns
kara
'glad'
Others
Raising
phenomenon
In case of TAM groups translation is situation dependent

Implementation and Results

Translation	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

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.

Conclusion

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.

Conclusion

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

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-> {tam:nA}
   # I want him to pass the exam
3. from_list(previous_root,begin_to.dat) &&
   following_category=v-> {tam:nA}
   # Anne began to cry.
4. previous_word=going-> {tam:ne}
   #He is going to start it.

Appendix

S.No Condition Hindi meaning

5. from_list(previous_root,ability_to.dat) &&
   following_category=v -> {tam:ne~kI}
   #He has ability to cross the river.
6. from_list(previous_verb,promise_to.dat) &&
   following_category=v -> {tam:ne~kA}
   #They promised her to go there.
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.

Appendix

S.No	Condition	Hindi meaning
9.	from_list(previous_word,happy_kara_to.dat) && following_category=v -> #We were glad to see him.	{tam:0_kara}
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.

Namaste

