

---

# Tag Disambiguation of Discourse Particles

Parul Upadhyay

Department of Linguistics

Banaras Hindu University

Email: [parul\\_ujjain@yahoo.com](mailto:parul_ujjain@yahoo.com)

---

# Discourse Particles

- Discourse particles such as English *yes/ya*, *well*, *ok* and Hindi *to*, *haaN*, *thiik hE*, *calo* (*choRo*) are quite frequent in spontaneous speech
- Despite their significant role in information structure in dialogue/conversational speech, they have not been given due attention in tag disambiguation, unlike major words such as nouns, verbs, adjectives, etc.

---

# Focus

- To study
  - What these discourse particles contribute to the conversation/dialogue with the aim to determine their functions
  - In order to make use of the information they provide, they need to be disambiguated

---

# Tags of *to*

- *to*
  - *to* in Hindi is one of the most complex and interesting discourse particle
  - *to* can be tagged as
    - Sentential Connective (SC)
    - Discourse Marker (DM)
    - Topic Marker(TM)
    - Pragmatic Question Marker (PQM)
    - Conditional Question Marker (CQM)
    - Emphatic Marker (EM)
    - Contrastive Connective(CC)

---

# Sentential Connective (SC)

- *to*: SC

1) *to* rotiyon kaa sahaaraa bhii gayaa

SC bread of support also went

‘The support for livelihood too disappeared.’

Context: *to* occurs in the initial position

---

# Discourse Marker (DM)

- *to*: DM

2) *to*, aNdherii raat thii aur ham haath meN DanDaa liye saTaasaT cale aa rahe the, tabhii ek peR ke piiche se kisii ne kahaa, 'khabardaar'.

{DM dark night was and we hand in sticks taking quickly moving then one tree behind from someone said stop

“It was a dark night and we were walking quickly carrying sticks in our hands when suddenly from behind a tree someone said, 'Stop'.”

Conext: *to* occurs between one or more sentences and one sentence connecting a idea in a narrative.

---

# Topic Marker (TM)

- *to*: TM

3) maiN **to** apnaa kismat kaa phaislaa kiye baithaa thaa.

I      TM self    fate    of    decision    did    sat      was

“I had already decided my fate.”

Context: *to* occurs immediately after subject noun. It can also occur after a verb too with same effect of topicalization.

---

# Pragmatic Question Marker (PQM)

- *to*: PQM

4) tumhe isne kaaT **to** nahiiN liyaa?  
you it bite-pr-3p negative  
“It did not bite you, I hope.”

Context: *to* occurs between verb and negative particle forming a question.

- This has a pragmatic overtone. This question can be said only in the context where the predicate event is contrary to the speaker's expectation.



---

# Conditional Question Marker (CQM)

- *to*: CQM

5) dekhnaa, koi jaag gayaa **to**?

see, someone wake up CQM

“Be careful, someone may wakes up.”

Context: *to* occurs sentence-finally after the verb in a question sentence.

---

# Emphatic Marker (EM)

- *to*: EM

6) nahiiN **to**

no

‘NO’

Context: *to* occurs with the negation particle in a reply-sentence for yes-no question.

---

# Contrastive Connective (CC)

- *to*: CC

7) ek viir hE **to** dusaraa mahaaviir.

one brave is CC other very brave

‘When the one is brave the other one is much more so.’

Context: *to* occurs between two clauses and functions as contrastive connective.

---

# Concluding remarks

- Here, only one particle *to* has been illustrated with discourse functions and their tag disambiguation
- The discourse particles such as *calo*, *choRo*, etc in Hindi are much more interesting and require study with a view to identify their tags
  - *calo choRo jaane do*
  - *dekho, dekhate hEN*
  - Etc.

---

# Concluding remarks

- All the examples presented here are taken from different Hindi texts.
- Tags and their disambiguation contexts are not exhaustively worked out.
- Requirements: A Very Large Corpus

---

# References

## References

- Brown, G., and Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Byron, D. K., and Heeman, P. A. (1997). Discourse marker use in task-oriented spoken dialog. In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech)*.
- DeRose, S. J. 1988. Grammatical category disambiguation by statistical optimization. *Computational Linguistics* 14(1):31–39.
- Jurafsky, D., J.H. Martin, (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice- Hall.
- Martina Möllering (2001). Teaching German Modal Particles: A Corpus-based Approach. In *Language Learning & Technology*, Vol. 5, No. 3, pp. 130-151.

---

# Thanks