Some issues in Part of Speech Tagging of Marathi

Prof. Malhar Kulkarni

Anuja Ajotikar

Dept. Of Humanities and Social Sciences
Indian Institute of Technology, Bomaby
Powai, Mumbai 400076
(malharku, anujaajotikar)@gmail.com

Abstract

POS tagging is an important module required as a base for various kinds of processing and tools like Chunker, Parser etc. in NLP. In fact the approach in doing the POS tagging determines the approach one adopts in developing various tools mentioned above. POS Tagging could be done following Lexical category Assignment approach as well as Functional Category Assignment approach.

For the time being in India, we have adopted the Lexical Category Assignment approach for POS tagging. This poses many challenges and questions in the minds of the Annotators. IITBombay is involved in annotating Marathi data with POS tags.

This paper presents and discusses such challenges faced during the annotation process and what linguistic insights we could gain from this experience. The paper also discusses the BIS tagset adopted for Marathi critically.

References

Akshar B. Mishra, D., Lakshmi Bai, Sangal, R. (2006). *AnnCorra: Annotating Corpora Guidelines For POS and Chunk Annotation For Indian languages*, IIIT: Hyderabad (Unpublished reference document)

Arjunvadkar, K. S. (edi.) (1970). *Shastriya Marathi Vyakarana* of Damale M. K. Pune: Deshmukh ani Company

Bapat, M. (2010). *Development of High Accuracy Morphological Analyzer for Marathi,* IIT Bombay*:* Unpublished MTECH Dissertation.

Bhattacarya, (Dr.) P. et. al. *Marathi POS tagger*, CFILT, IIT Bombay

Chaitanya, V. Sangal, R. BHarati, A. (1995) *Natural Language Processing: A Paninian Perspective New Delhi: Prentice Hall of India Publication*

Dash, N.S. (2010) *Corpus Linguistics: A General Introduction* http://www.ldcil.org (August 26, 2011)

Dash,N.S. (2004) L*anguage Corpora: Present Indian Need* http://www.elda.org (August 25, 2011)

Dixit, Veena. (2005) Marathi *Verb Morphology and POS tagging.* PPT presentation, IIT Bombay: CFILT

Gadge, Dinesh. (2007) Common Part of Speech Tagger for Hindi and Marathi, IIT Bombay: unpublished BTech Dissertation.

Mohanty, R. (2005) *Part Of Speech Tagging* PPT presentation:2nd Asian Regional Training on Local Language Computing Cambodia: Siem Reap

Pandharipande, R. (1997) *Marathi*, London:Routledge

Pawar, S. S. (2008) *Shallow Parsing of Indian Languages*, IIT Bombay: Unpublished MTECH Dissertation.

Singh, Smriti. (2006) *Hindi Mrphological Analysis: Applications In Various Language Processing Tools* Annual Progress Report – 2, IIT Bombay: Dept. of Humanities and social Sciences

Waghmare Rajkumar. (2009) Morphological Analysis and Part of Speech Tagging, IIT, Bombay: Unpublished M. Tech Dissertation