Hindi Verb Knowledge Base and Noun Incorporation in Hindi

Debasri Chakrabarti
Computer Science and
Engineering Department
Indian Institute of Technology,
Bombay,
Mumbai, 40076, India.
debasri@cse.iitb.ac.in

Vaijayanthi Sarma
Department of Humanities
and Social Sciences
Indian Institute of Technology,
Bombay,
Mumbai, 40076, India.
vsarma@hss.iitb.ac.in

Pushpak Bhattacharyya
Computer Science and Engineering
Department
Indian Institute of Technology,
Bombay,
Mumbai, 40076, India.
pb@cse.iitb.ac.in

Abstract

The work reported in this paper deals with the Hindi verbs. This paper can be divided into two parts. The first part is a description of a lexicon where the verbs are arranged hierarchically according to their super-ordinate terms. In this hierarchy, verbs are first listed according to their specific senses. This lexicon is named as Hindi Verb Knowledge Base (HVKB). HVKB uses constructs and Knowledge Base of Universal Networking language (UNL), an interlingua, CIIL (Central Institute of Indian Languages) Corpora and partially Hindi Wordnet (henceforth HWN). In the second part of the paper we have presented the structural description of the Hindi verbs. A type of complex verb, namely, Noun+Verb combinations are studied in detail. The motivation behind this study was to look for a principled approach in storing the Noun+Verb combinations in HWN. Different syntactic tests are applied to identify such verbs. Various syntactic and semantic properties of this group emerged through these tests. These properties help identifying this group as a process of lexical compound. These lexical compounds are then stored in the hierarchy. Finally, the paper shows how HVKB will prove beneficial for HWN.

Keywords: Simple Verbs, Conjunct Verbs, Hindi Wordnet (HWN), Universal Networking language (UNL) and UNL Knowledge Base, Noun+Verb (N+V) combination, Noun Incorporation (NI).

1 Introduction

Although the lexical tradition in Hindi is old [1], the lexicons of the language do not record idiosyncratic properties of a category or give example sentences. Moreover, the dictionaries do not mark the *meaning change* or *meaning shift* that the language has undergone. In this scenario, we felt the necessity of dealing with the Hindi verbs more elaborately. The enriched lexicon of Hindi verbs, that we have constructed, is termed as Hindi Verb Knowledge Base (henceforth, HVKB).The need for creating this HVKB is to integrate it with Hindi WordNet (HWN) [2] and enrich the latter.

Recent lexicography researches attempt to classify the verbs and place them according to their syntactic and semantic structure. WordNet [3; 4], EuroWordNet [5], Verbnet [6] are some examples of such researches. As per our knowledge, HVKB is the first attempt to organize the verbs in Hindi in a hierarchical structure.

In creating the HVKB, CIIL corpora [7] was of immense help as it helps in identifying various senses of a verb through different usages of it. These senses are then verified by the native speakers and kept in the lexicon. According to the lexicography theories a lexicon should contain all probable semantic and syntactic features of a given lexical category. This has been taken care of while building HVKB. Presently, HVKB deals only with the simple verbs, as complex verbs need thorough investigation. After identifying the senses of simple verbs, they are arranged in a hierarchy

verbs in the VKB. The coverage of the verbs is increasing everyday.

In future we would like to extend the analysis by studying the other ways of forming complex verbs and also would provide the case marking information in the sentence frame.

References

[1] Jha Vishwanath. (1975) *Amarkosha by Amarsingha*, Motilal Banarasidas Publications, Varanasi.

[2] `http://www.cfilt.iitb.ac.in/wordnet/webhwn/`

[3] Miller G., Beckwith R., Fellbaum C., Gross D., and Miller K. (1990) *Five Papers on WordNet.* CSL Report 43, Cognitive Science Laboratory, Princeton University, Princeton.

[4] Fellbaum C (ed.). (1998) *WordNet: An Electronic Lexical Database,* The MIT Press.

[5] Vossen Piek. (1998) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks,* ed *Dodrecht: Kluwer Academic Publishers.*

[6] `http://www.cis.upenn.edu/~mpalmer/verbnet.html`

[7] `http://www.cfilt.iitb.ac.in/indiancorpora/web/hindi/index.html`

[8] `http://www.unl.ias.edu/unlsys/unl/UNL%205specifications.html`

[9] Verma Ramchandra and Kapur Badarinath. (1998) *Lokbharti Pramanik Hindi Kosh.* Lokbharati Prakashan, ed. Varanasi.

[10] Shree Naval Jee (2000) *Nalanda Vishal Shabd Sagar*, Adish Book Depot, New Delhi.

[11] R. S. Mcgregor, *The Oxford Hindi English Dictionary*, Oxford University Press, New Delhi, 1997.

[12] S. S. Gupta, Suresh Agarwal compiled and Edited, *Standard Illustrated Advanced Dictionary*, Ashok Prakashan, Delhi-6, 2003.

[13] `http://framenet.icsi.berkeley.edu/papers/acl98.pdf`

[14] Chakrabarti D Pushpak Bhattacharyya, *Creation of English and Hindi Verb Hierarchies and their Application to Hindi WordNet Building and English-Hindi MT*, Proceedings of the Second Global Wordnet Conference, Brno, Czech Republic, 2004.

[15] BNC Consortium: *British National Corpus.* The Humanities Computing Unit of Oxford University. (2000) `http://www.hcu.ox.ac.uk/BNC`.

[16] George Miller, Wordnet 2.0. (2003), `http://wordnet.princeton.edu/`

[17] A. S. Hornby, *Oxford Advanced Learner's Dictionary of Current English*. Oxford University Press, Oxford.(2000)

[18] *The Universal Networking Language (UNL) Specifications*, Version 3.0, UNL center, UNDL Foundation, 2001.

[19] Andrew Radford, *Transformational Grammar A First Course,* Cambridge Textbooks in Linguistics, 1988.

[20] Ray, Jackendoff, *Semantics and Cognition*, Cambridge, MIT Press, 1983.

[21] Mohanan, T. 1995. "Wordhood and Lexicality", *NLLT* 13, 75-134.

[22] Hook, P. E. 1979. *Hindi Structures: Intermediate Level*. Michigan Papers on South and Southeast Asia, The University of Michigan Center for South and Southeast Studies, Ann Arbor Michigan.

[23] Veneeta Dayal, *A Semantics for Pseudo Incorporation*, Rutgers University, November, 2003.

[24] Mohanan, Tara Warrier. 1990. *Arguments in Hindi,* A Dissertation, Dept. of Linguistics, Stanford University.

[25] Baker M. 1995. *The polysynthesis parameter*, New York: Oxford University Press.

[26] M., Mithun, *The Evolution of Noun Incorporation*, Language 60, 1984.