

Constellations, Polysemy, and Hindi –ko

SHRAVAN VASISHTH
University of Saarland

BRIAN JOSEPH
Ohio State University

1. Background: General Issues

Much of linguistic analysis rests on a single key question: given entities X and Y as objects for analysis, are they the same or different? This issue pervades all components of grammar: phonetics, phonology, morphology, syntax, etc. Moreover, in addressing this issue, one often needs to recognize the relevance of different levels of analysis, especially underlying versus surface, since underlying sameness can be surface difference, and vice versa. For example, phonemic analysis takes phones with decidedly different phonetic realizations (e.g. aspirated vs. unaspirated stops in English) and treats them as the same at the phonemic level if their distribution does not overlap. But at the same time, segments that seem to be the same phonetically on the surface and even phonemically as well, e.g. the [d] of *recede* and the [d] of *invade*, might need to be treated as different from a morphophonemic standpoint, since, in this example, the former alternates with [ʃ] in the related noun *recession* whereas the latter alternates with [ʒ] in *invasion*, both nominal formations having ostensibly the same suffix. In syntax, too, patterns that are alike on the surface, such as control constructions (e.g., *Skippy tried not to mind*) and raising constructions (e.g., *Skippy seemed not to mind*), can show some unlike properties that lead, in most current theoretical frameworks at least, to structural differentiation in some way, e.g., in underlying structure (cf., **It tried not to be raining* vs. *It seemed not to be raining*). As these examples indicate, an answer to the above key question regarding sameness often involves a recognition of differences too. Thus, the issue becomes one of measuring similarities and differences against one another and weighing the relative importance of one or the other, as well as deciding how to represent the sameness or difference that one ends up positing.

2. Same vs. Different in Morphology

The few examples in section 1 involve phonology and syntax, but, as noted at the outset, the same-vs.-different question pertains to morphology also, where the issue is rather: When are two morphs to be considered related to one another?

While there is the purely conditioned sort of difference that can be appealed to, as with phonemics (i.e., where there are allomorphs of a single morpheme), at the level of morphology, meaning is also involved; therefore one has to balance purely formal similarities and differences with semantic/functional similarities and differences. The traditional approach to this issue in morphology has been essentially all or nothing relatedness: one looks for recurring elements of form matching with recurring elements of meaning and posits a single underlying form that unites the variant realizations. For example, *repose* [ripoz] and *repository* [ripazitori] share [ripVz-] in form and 'put to rest' (or the like) in meaning, and this sameness can be expressed through an underlying root form having (roughly) the shape /ripoz-/.

But this discovery procedure is not without some difficulties. One the formal side, suppletion, as in *go/went*, presents a problem as there is usually little or no formal sameness to draw on (e.g., is the [+back] specification of [g] and [w] enough to allow for *go* and *went* to be connected formally?). On the semantic side, a problem arises with words that are etymologically, and still formally, related but have come to show different meanings, e.g., *suppose/suppository* (which as a pair is formally similar to, but semantically different from, *repose/repository*). More generally, then, any sort of drift, especially in form but also in meaning or even both, poses problems. For example, in *two/twelve* there is (some) formal separation (#t- vs. #tw-, and different vowels) but still a plausible semantic connection, while in *two/twine* there is both formal and semantic separation; other parallel cases can be easily found.

3. Other Solutions

In a sense, the null case from the formal standpoint comes when the forms one is considering relating are identical, differing only in meaning. Such a situation typically lends itself to a different type of solution. In particular, when the items under consideration show no formal differences, generally one can invoke polysemy as the means of representing the connection, and see the differences in meaning as a series of extensions, possibly metaphorical in nature, along one of the dimensions of the meaning of a given form. An example is the classic analysis by Brugman (1988) of *over* in English (where the link has to do with mapping the relation of a "trajector" to a "landmark", i.e. ABOVE x ACROSS) or with a putative connection linking English prepositional *to* with infinitival *to*, with directional *to*, and with indirect object marking *to* (in terms of movement towards some goal), etc.

However, when the forms themselves are not identical, in order to capture unity-in-diversity as well as diversity-in-unity, a different approach has been argued for, the CONSTELLATIONAL APPROACH of Janda and Joseph (1986). This approach involves two constructs, defined in (1):

- (1) a. THE CONSTELLATION: A group of elements which share at least one characteristic property of form but are distinguished by individual idiosyncrasies of both form and function that prevent their being collapsed with one another.
- b. Meta-template: A meta-level redundancy statement, which ranges over all relevant candidates and equates instances of a particular formal configuration that meet certain criteria of uniting properties.

Thus, diversity is characterized by the Constellation, where differences in the elements are recognized, and unity by the Meta-template, which “parses”, and thus identifies, all relevant equitable instances of a determined form. Moreover, the Constellation and Meta-template are related, in that, e.g., morphological constellations are ensembles of word-formational elements (e.g., morphemes) united by meta-templates which express the formal and functional identities shared by a set of distinct morphemes or, alternatively, uncollapsible morphological rules or constraints. The Constellation and the Meta-template together provide a mechanism that allows a realistic, non-procrustean approach to sameness in linguistic analysis – a recognition of how elements can simultaneously be same but also different (uncollapsible). Examples in the literature include Sanskrit reduplication (Janda & Joseph 1999), Arapesh plurals (Dobrin 2001), and Finnish definitives (Välismaa-Blum 1989).

4. Our Goal: A Constellational Account of *-ko*

We argue here that invoking constellations is appropriate even when identical forms are involved, and use constellations to clarify the relationship among several interconnected elements in Hindi, all with the shape *ko*, that have preciously been misanalyzed as merely polysemous. Moreover, we extend the range of evidence available to confirm constellational status, going beyond distributional facts and syntactic behavior, and bring in relevant experimental results from language comprehension studies that bear on the similarities and differences among these elements *ko*.

5. The element(s) *-ko* in Hindi

The relevant elements are the Hindi postposition markers with the shape *ko*. Most relevant analyses (e.g., McGregor 1995, Wunderlich 2000) either conflate them or claim several distinct uses for this single element:

- (2) a. Rīṭa Sita-ko akhbaar de-gii
 Rita Sita-ko newspaper give-fut
 ‘Rita will give (a/the) newspaper to Sita.’
- b. Ram-ko bhuukh lagii hai
 Ram-ko hunger feel is
 ‘Ram is feeling hungry.’

- c. Kavita kitaab(-ko) paṛh-rahii hai
 Kavita book(-ko) read-cont is
 'Kavita is reading a (the) book.'
- d. Kalika-ne Seema-ko kitaab khariidne-ko bolaa
 Kalika-erg Seema-dat book buy-inf told
 'Kalika told Seema to buy a book.'
- e. Hari shaam-ko Ravi-ke ghar gayaa
 Hari evening-ko Ravi-gen house went
 'Hari went to Ravi's house in the evening.'

These *-ko*'s appear to be a single polysemous morpheme, with a common form *-ko*, and identical positioning vis-à-vis topicalizer *-to* and focalizer *-tak*, occurring only to the left of the marker with *-to*, and only to the right with *-tak*.

- (3) a. Riṭa Sita-ko-to/*to-ko akhbaar de-gii
 Rita Sita-ko-top newspaper give-fut
 'As for Sita, Rita will give her (a/the) newspaper.'
- b. Ram-ko-to/*-to-ko bhuukh lagii hai
 Ram-ko-top hunger feel is
 'Ram is feeling hungry.'
- c. Kavita kitaab-ko-to/*-to-ko paṛh-rahii hai
 Kavita book-ko-top read-cont is
 'As for the book, Kavita is reading it.'
- d. Kalika-ne Seema-ko kitaab khariidne-ko-to/*-to-ko bolaa
 Kalika-erg Seema-dat book buy-inf-top told
 'Kalika did tell Seema to BUY a book.'
- e. Hari dopehar-ko-to/*-to-ko jaaye-gaa
 Hari afternoon-ko-top go-fut
 'Hari will go in the AFTERNOON (but maybe not any other time).'
- (4) a. Riṭa Sita-tak-ko/??ko-tak akhbaar de-gii
 Rita Sita newspaper give-fut
 'Rita will give even Sita (a/the) newspaper.'
- b. Ram-tak-ko/??ko-tak bhuukhq lagii hai
 Ram hunger feel is
 'Even Ram is feeling hungry.'
- c. Kavita kitaab-tak-ko/??ko-tak paṛh-rahii hai
 Kavita book read-cont is
 'Kavita is reading even the book.'
- d. Kalika-ne Seema-ko kitaab khariidne-tak-ko/??ko-tak bolaa
 Kalika-erg Seema-dat book buy told
 'Kalika told Seema to even buy a/the book.'

- e. Hari dopehar-**tak-ko**/?**ko-tak** jaaye-gaa
 Hari afternoon-ko-even go-fut
 'Hari will go even in the afternoon.'

Even though these *ko*'s all behave alike here, we claim that they do not represent mere polysemy. In particular, a polysemy account cannot explain the semantic spread for *ko*: even if some functions are relatable, e.g., indirect objects and infinitivals as goal-oriented, no semantic principles or pathways can derive definiteness from any of the other functions. Moreover, these *ko*'s always show differentiated behavior.

The constellational approach predicts the occurrence of non-semantic differentiation. This is found among the various *ko*'s. Consider the interaction of the *ko*'s with *bhii* 'also, even'¹ and *hii* 'only': only the Indirect Object and Subject *ko*'s can be freely ordered; the other *ko*'s must occur to the left of these particles. (5) shows the cooccurrence patterns with *bhii*; those with *hii* are identical.

- (5) a. Rīṭa Sita-**ko-bhii/bhii-ko** akhbaar de-gii
 Rita Sita-ko-even newspaper give-fut
 'Rita will give even Sita (a/the) newspaper.'
- b. Ram-**ko-bhii/bhii-ko** bhuuk lagii hai
 Ram-ko-even hunger feel is
 'Even Ram is feeling hungry.'
- c. Kavita kitaab-**ko-bhii**/?**bhii-ko** paṛh-rahii hai
 Kavita book-ko-even read-cont is
 'Kavita is reading even the book.'
- d. Kalika-ne Seema-ko kitaab khariidne-**ko-bhii**/?**bhii-ko** bolaa
 Kalika-erg Seema-dat book buy-inf-even told
 'Kalika told Seema to even buy a/the book.'
- e. Hari dopehar-**ko-bhii**/?**bhii-ko** jaaye-gaa
 Hari afternoon-ko-even go-fut
 'Hari will go even in the afternoon.'

6. Evidence from language comprehension studies

The constellational approach assumes that there is a purely formal and grammatical basis for differentiating among the various *ko*'s, and that these *ko*'s really are different entities (though united by a meta-level redundancy statement).

We present evidence from language processing that further supports the constellational view by demonstrating the existence of undifferentiated as well as differentiated treatment in processing of the case-marker versus infinitival *-ko*.

Davison (1991) and Butt (1993) have argued that *ko*-marked infinitivals are simply nominals with ordinary case-marking: "...the constituent headed by the infinitive not only has the distribution of an NP, it can take case markers and

¹ See Schwenter & Vasishth 2000 for the distinction between *-tak* and *-bhii*.

undergo some further morphological processes that only apply to NPs. The entire infinitival ‘clause’ must therefore be analyzed as an NP” (Butt 1993:52).

They provide linguistic evidence that this is the case: (1) The inflections on the infinitivals are similar to nominal inflections; (2) In an infinitival, *-ko* can be replaced by clear postpositions, like *keliye* ‘for’; and (3) Coordination of two NPs versus two infinitivals behaves identically. By contrast, Mohanan (1994:13-14), Bickel & Yadava (2000), and others suggest that infinitivals are verbal.

If infinitivals are merely NPs with *-ko* case marking, then the expectation is that these will behave like NPs during real-time sentence comprehension (assuming that there is a close connection between linguistic constructs and their mental representation). In particular, they should be stored in short-term or working memory as NPs. Center embeddings are a good way to test these hypotheses. Consider the double self-center embedding below:

- (6) Siitaa-ne Hari-ko Ravi-ko kitaab khariid-neko bol-neko kahaa
 Sita-erg Hari-dat Ravi-dat book buy-inf tell-inf told
 ‘Sita told Hari to tell Ravi to buy a/the book.’

Pre-theoretically, parsing such a sentence in real time involves (a) **storing** each NP as it is encountered; and (b) **integrating** the NPs with verbs as the verbs are encountered. However, when NPs are stored in memory, they are encoded in some way rather than being stored as-is, and there is much evidence that NPs in the context of a sentence generate predictions (Lewis & Nakayama 2001, Gibson 2000, Vasishth 2002). These predictions are expectations of verbs and of sentence structures. In self-paced reading experiments involving Dutch (Dickey & Vonk 1997, Kaan & Vasić 2000), we find an invariant pattern: arriving at a verb after seeing an array of NPs results in faster reading time – the integration of NPs with a verb reduces local processing load, perhaps because the NPs and verb are now stored in working memory as one unit, a “chunk” (Miller 1956, Lewis 1996).

If infinitivals are really NPs, they should not be involved in any integration-related speedups, which (we are assuming) is a property of verbs. Thus, there should be no speedup at the innermost infinitival if it is an NP. Psycholinguistic experiments show that the infinitival behaves just like a verb – there is a significant fall in reading time (RT) at the infinitival. This is in contrast to the monotonically nondecreasing RT observed with the successive appearance of NPs (a reflection of storage costs).

The two experiments presented here are noncumulative moving window self-paced reading tasks (Just, Carpenter, & Woolley 1982). The procedure is as follows. First, a set of blank lines appear on the computer screen, each corresponding to a word in the sentence to be read. Then, with each press of the space bar, the subject sees each phrase successively, and the previous phrase disappears. The dependent measure is RT for each phrase. After the sentence ends, a yes/no comprehension question is presented to ensure that subjects are attending to the sentence.

The first experiment uses single center embeddings containing ditransitive (lexical causative) embedded verbs; although the critical comparison here is between presence versus absence of specificity/definiteness marking on direct object (1 x 2 design), in the present discussion we are interested in the difference between RTs at NPs versus verbs. The sentence types of interest are shown in (7); each phrase separated from others by white spaces is a separate region.

- (7) a. Siitaa-ne Hari-se Ravi-se kitaab le-neko kahaa
 Sita-erg Hari-abl Ravi-abl book take-inf told
 ‘Sita told Hari to take a book from Ravi.’
 b. Siitaa-ne Hari-se Ravi-se kitaab-ko le-neko kahaa
 Sita-erg Hari-abl Ravi-abl book-acc take-inf told
 ‘Sita told Hari to take the book from Ravi.’

As Figure 1 shows, there is a significant speedup at the infinitival element. This is consistent with integrative (“verb”-like) processes occurring at the infinitival, but not with storage (“noun”-like) processes. Therefore, the results support the assumption in Mohanan 1994 and Bickel & Yadava 2000 that infinitivals are verbs.

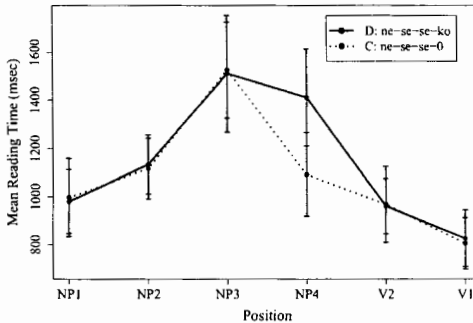


Figure 1: Results for Experiment 1 (95% confidence intervals)

A second experiment provides independent support for the conclusion that in real-time processing infinitivals behave like verbs.

Research on working memory suggests that human parsing involves a predictive component: possible sentence completions are anticipated as a sentence is being processed. Coupled with the fact (Mahajan 1990:87-88) that an adverb in Hindi must attach to a verb-projection, we can manipulate the degree of confidence in the prediction that the next word is a verb. A higher degree of

confidence would mean a faster RT at the verb. By contrast, if the infinitival is an NP, then there is no reason to expect a speedup at the infinitival as a consequence of adverb-insertion; if anything, there should be a slowdown, since the infinitival NP would then have to be “reconstituted” as a verb for parsing purposes.

In order to test these hypotheses, a self-paced reading study was conducted with a 1 x 2 design. The factors were presence or absence of adverb between final NP and first verb.

- (8) a. Siitaa-ne Hari-ko [Ravi-ko [kitaab-ko
 Sita-erg Hari-dat Ravi-dat book-ko
 khariid-neko] bol-neko] kahaa
 buy-inf tell-inf told
 ‘Sita told Hari to tell Ravi to buy the book.’
- b. Siitaa-ne Hari-ko [Ravi-ko [kitaab-ko
 Sita-erg Hari-dat Ravi-dat book-acc
jitnii-jaldi-ho-sake khariid-neko] bol-neko] kahaa
 as-soon-as-possible buy-inf tell-inf told
 ‘Sita told Hari to tell Ravi to buy the book as soon as possible.’

As Figure 2 shows, RT is significantly faster at the infinitival when an adverb is present. This is consistent with the assumption that the infinitival is a verb and not with it being an NP.

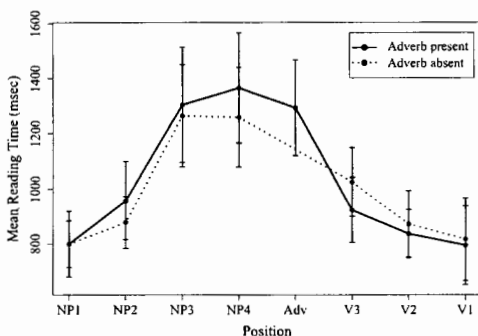


Figure 2: Double embeddings in Experiment 2 (95% confidence intervals)

In sum, the all-or-nothing assumption leads to the infinitival-as-NP debate, and raises two questions; (1) Is the *ko* marking on the infinitival ordinary case marking or not? (2) How do we resolve the apparent contradiction that these infinitival embedded clauses behave like NPs (purely linguistic evidence) as well

as verbs (reading comprehension studies)? By contrast, the Constellational approach suggests that the infinitival-as-NP debate is asking the wrong questions. Under this view, it is completely acceptable if the *ko* of ...*neko* is not segmentable and categorizable in exactly the same way as in *Sita-ko*—the meta-level redundancy statement will “pull out” a *ko* and equate the *ko* of ...*neko* with the *ko* of *Sita-ko*, due, e.g., to their parallel behavior vis-à-vis *to* and *tak* (as a feature that unites them). Moreover, under the Constellational view, these infinitival embedded clauses are *expected* to be similar to NPs in some respects (linguistic tests) and different in other respects (their behavior in real-time sentence comprehension).

7. Conclusion

The Hindi *ko*'s, therefore, show that constellations achieve finer granularity in empirical coverage than pure polysemy accounts, thus allowing the analyst to have his/her cake and eat it too: elements can be same and different, but in a principled manner.

Moreover, the experimental evidence independently motivates the unity in diversity that is absent from the strictly unifying polysemy approach. The experimental results also provide a good example of the empirical consequences of adopting the Constellational approach; without it, debates like (Butt, 1993) *versus* Mohanan (1994) (“are infinitivals NPs or not?”) are bound to arise, but these debates (although useful) are raising the wrong question.

References

- Bickel, B. & Y.P. Yadava. 2000. A fresh look at grammatical relations in Indo-Aryan. *Lingua* 110: 343-373.
- Brugman, C. 1988. *The Story of Over: Polysemy, Semantics, and the Structure of the Lexicon*. New York: Garland Publishing.
- Butt, M. 1993. Hindi-Urdu infinitives as NPs. *South Asian Languages Review* 3(1): 73-84.
- Davison, A. 1991. Finiteness and case in Hindi-Urdu complements. *Proceedings of the Thirteenth South Asian Languages Roundtable*.
- Dickey, M.E. & W. Vonk. 1997. Center-Embedded Structures in Dutch: An On-Line Study. Ms.
- Dobrin, L. 2001. The morphological status of Arapesh plurals. Talk presented at the 2001 Linguistic Society of America Annual Meeting, San Francisco, CA.
- Gibson, E. 2000. Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (eds.) *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*. Cambridge, MA: MIT Press.
- Janda, R. & B. Joseph. 1986. One rule or many? Sanskrit reduplication as fragmented affixation. In B. Joseph (ed.) *OSUWPL* 34: 84-107.

- Janda, R. & B. Joseph. 1999. The modern Greek negator *mi(n)(-)* as a morphological constellation. In G. Babiniotis (ed.) *Greek Linguistics: Proceedings of the 3rd International Conference on Greek Linguistics*, 341-351. Athens: Elinika Gramata Press.
- Just, M.A., P.A. Carpenter, J.D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* **111**(2): 228-238.
- Kaan, E. & N. Vasić. 2000. Cross-serial dependencies in Dutch: Evidence for a discourse-based metric? Poster presented at CUNY Sentence Processing Conference, University of California, San Diego.
- Lewis, R.L. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* **25**(1): 93-115.
- Lewis, R.L. & M. Nakayama. 2001. Syntactic and positional similarity effects in the processing of Japanese embeddings. In M. Nakayama (ed.) *Sentence Processing in East Asian Languages*, 85-113. Stanford, CA.
- Mahajan, A.K. 1990. The A/A-bar Distinction and Movement Theory. Unpublished doctoral dissertation, MIT, Cambridge, MA.
- McGregor, R.S. 1995. *Outline of Hindi Grammar*. New York: OUP.
- Miller, G.A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**: 81-97.
- Mohanan, T. 1994. *Argument Structure in Hindi*. Stanford, CA: CSLI.
- Schwenter, S.A. & S. Vasisht. 2000. Absolute and relative scalar particles in Spanish and Hindi. *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA.
- Väkunaa-Blum, R. 1989. Finnish word order as a set of syntactic constructions. In J. Powers & K. de Jong (eds.) *Proceedings of ESCOL 1988 (Eastern States Conference on Linguistics)*, 500-511. Columbus, OH.
- Vasisht, S. 2002. An Abductive Inference Based Model of Human Parsing. Unpublished Master's thesis, Ohio State University.
- Wunderlich, D. 2000. Optimal case in Hindi. Ms.

Department of Linguistics
Ohio State University
222 Oxley Hall
Columbus, OH 43210

vasishth@ling.osu.edu; bjoseph@ling.osu.edu