

Semantics of Spatio-Directional Geometric Terms of Indian Languages

Sukhada¹, Soma Paul², Rahul Kumar³, Karthik Puranik⁴

¹IIT (BHU), Varanasi, ²IIT-Hyderabad, ³University of Delhi ⁴IIT Tiruchirappalli
sukhada.hss@iitbhu.ac.in, soma@iiit.ac.in,
sadarpurrahull@gmail.com, karthikp18c@iiitt.ac.in

Abstract

This paper examines widely prevalent yet little-studied expressions in Indian languages which are known as geometrical terms because “they engage locations along the axes of the reference object”. These terms are *andara* (inside), *bāhara* (outside), *āge* (in front of), *sāmane* (in front of), *pīche* (back), *ūpara* (above/over), *nīce* (under/below), *dāyem* (right), *bāyem* (left), *pāsa* (near), *dūra* (away/far) in Hindi. The way these terms have been interpreted by the scholars of Hindi language and handled in the Hindi Dependency treebank is misleading. This paper proposes **an alternative analysis of these terms focusing on their triple – nominal, modifier and relational - functions and presents abstract semantic representations** of these terms following the proposed analysis. The semantic representation will be explicit, unambiguous abstract and therefore universal in nature. The correspondence of these terms in Bangla and Kannada are also identified. Disambiguation of geometric terms will facilitate parsing and machine translation especially from Indian Language to English because these geometric terms of Indian languages are variedly translated in English depending on context.

1 Introduction

Geometric terms that “engage locations along the axes of the reference object” (Landau, 2017) play multiple roles - relational, nominal, modifier - in Indian languages and their equivalents in English can vary based on their functional role. For example, the geometric term *nīce* in (1-a) and (1-b) is translated differently into English depending on the function of the word in the given two contexts:

- (1) a. meja ke nīce
table.SG GEN under
cyūṁgama
Chewing gum.SG.3.NOM

- cipakī hai
stick.PR.SG.3
‘There is a chewing gum stuck **under**
the table’.
b. nīce jāo
down go.IMP.SG
‘Go **down**’.

In (1-a), the term *nīce* indicates a spatio-directional¹ relation between *meja* (table) and *cyūṁgama* (chewing gum) while in (1-b), *nīce* denotes downward location. Thus, in (1-a), *nīce* (under) is a relational marker that specifies the geometric position of ‘chewing gum’ with respect to the ‘table’ while in (1-b) *nīce* (down) indicates location, thus fulfilling a nominal function. Interestingly there are contexts where these terms inadvertently appear to be relational markers which they are not. For such cases we attribute a third role to them, the role of a modifier. For example, in (2) the term *nīce* is a spatial modification of the location *peṭī* (box). The location of ‘books’ is the box and the position of the box is specified by the geometric term *nīce*. The reference object *meja* (table) with respect to which *nīce* has to be interpreted is present in (2-a). In (2-b), we show that *nīce* can even take the adjectival suffix, thus making its modifier role clearer.

- (2) a. meja ke nīce peṭī meṁ
table.SG GEN under box.SG LOC
pustaka rakhī hai
book.SG.NOM keep.PR.SG.3
‘The book is kept in the box under the
table.’
b. meja ke nīcevālī peṭī meṁ
table.SG GEN under.ADJ box.SG LOC
pustaka rakhī hai
book.SG.NOM keep.PR.SG.3
‘The book is kept in the box that is
under the table.’

¹Spatio-directional refers to spatial cum directional semantics

Thus, we identify three roles of the geometric terms in Indian languages:

1. A spatial noun denoting a place as in (1-b)
2. A geometric spatio-directional relation marker as in (1-a).
3. A geometric spatio-directional modifier of a noun as in (2-a)

In this paper, we propose semantic interpretation of geometric terms and present an abstract semantic representation for them. The semantic representation will be explicit, unambiguous, abstract and therefore universal in nature. Such enquiry is significant not just for Natural Language Understanding, but also in the context of language transfer. We have shown in (1) that English, for example, uses a preposition when the geometric term is a relational marker, otherwise the language uses adverbs as in (1-b) while Hindi and other Indian languages such as Bangla and Kannada use the same lexical item in both (1-a) and (1-b).

The paper is divided into the following sections. Section 2 introduces the equivalents of Hindi spatio-directional geometric terms in Bangla and Kannada. Section 3 presents the semantic interpretation of these terms. Section 4 presents semantic representation of these terms and shows how such representation captures information explicitly and unambiguously, which are the characteristic features of any efficient semantic representation system. Finally, we will present the design of a Geometric terms Search Interface with annotation facility integrated in section 5.

2 Geometric Terms in Hindi, Bangla and Kannada

Geometric terms under considerations are *andara* (inside), *bāhara* (outside), *āge* (in front of), *sāmane* (in front of), *pīche* (back), *ūpara* (above/over), *nīce* (under/below), *dāyeṃ* (right), *bāyeṃ* (left), *pāsa* (near), *dūra* (away/far) in Hindi. There are some more geometric terms which are used in Hindi such as *āra-pāra* (across), *ora* (towards) which are purely relational markers and hence we keep them out of scope of this paper.

Table-1 presents equivalent lexical terms in Hindi, Bangla, Kannada and English for the relational variant of these terms and their morpho-lexical properties. Most of the English terms apart

from ‘right’, ‘left’, ‘inside’ and ‘outside’ are spatial prepositions which are indeclinable.

Table-2 gives a quick comparative study of these lexical items in Hindi, Bangla and Kannada. The gloss for each example given in Table-2 is given in Appendix-1.

Fortis and Fagard following Levinson et al. have shown that relational nouns (spatial nominals) in Japanese and Korean follow a structure [Ground-GEN Spatial_Nominal-PostP] which, we see, is quite similar to Indian languages.

3 Semantic Interpretation of Geometric Terms

Talmy (1983) has introduced Figure-Ground geometric sense⁶ “to refer to the located vs locating entity”. Figure is “the object which is considered as moving or located with respect to another object” and Ground is “the object with respect to which a first is considered as moving or located” (Talmy, 1983) in the context of spatial configuration. Ground is alternatively referred to as a Reference object.

- (3) laḍakā_{ng} ghara_{gr} ke bāhara
 boy.SG.NOM house GEN outside
 khaḍā hai
 stand.PR.SG
 ‘The boy is standing outside the house.’

In Indian Grammatical Tradition, the relation is described between *ādhāra* (Ground) and *ādheya* (Figure). Two kinds of relations between *ādhāra* and *ādheya* have been identified in the literature which is very much relevant for our interpretation of geometric terms. This analysis also explains data recorded by Talmy in a different light that fits to our multirole analysis of these terms. The two relations are *saṃyoga-sambandha* (contact by touch) and *sāmīpya-sambandha* (contact by proximity) Shastri (1926); Subramania Iyer (1971). When there is a temporary physical contact (*saṃyoga*) between the *ādhāra* and the *ādheya* then the spatial relationship is said to have a *saṃyoga-sambandha*.

- (4) pakṣī peḍa ke ūpara baiṭhā hai /
 bird.SG.NOM tree GEN above sit.PR.SG.3 /
 pakṣī peḍa para baiṭhā hai
 bird.SG.NOM tree.LOC sit.PR.SG.3
 ‘The bird is sitting on the tree.’

²The concept of Figure and Ground are borrowed from gestalt psychology to linguistics.

Hi.	Hi. form	Ba.	Ba. form	Ka.	Ka. form	En.	Example	En. Translation
andara	Indcl ²	bhitare	bhitare ³ + e ⁴	oḷage	oḷa +akke	inside	ḍibbā thaile ke andara hai.	The box is inside the bag.
bāhara	Indcl	bāhira, bāire	bāira ⁵ +e	horage	hora +akke	outside	ghara ke bāhara nāma paṭṭī lagī hai.	The nameplate is outside the house.
āge	Indcl	āge	āga + e	mumde	Indcl	In front of	gāḍī ke āge nambara pleṭa nahīm hai.	There is no number plate in front of the car.
sāmāne	Indcl	sāmāne	Indcl	edhuru	Indcl	In front of	ye pirāmiḍa itihāsa kī eka alaga hī duniyā ko hamāre sāmāne rakhate haiṃ.	These pyramids put a different kind of world before us.
pīche	Indcl	pichane	pichana +e	himdhe	Indcl	behind	gāḍī ke pīche nambara pleṭa nahīm hai.	There is no number plate behind the car.
ūpara	Indcl	opara, opore	opara +e	mele	Indcl	Above/ over/ on	meja ke ūpara kapa hai	The cup is on the table.
nīce	Indcl	nīce	nīca +e	keḷage	keḷa +akke	under	meja ke nīce cyūimḡama cipakī hai.	The chewing gum is stuck under the table.
dāyem	Indcl	dāine	Indcl	balakke	bala +akke	right	paidala yātrī saḍaka ke dāyīm ora caleṃ	Pedestrians should walk on the right side of the road.
bāyem	Indcl	bāmye	bām +e	edakke	eda +akke	left	gāḍiyām saḍaka ke bāyīm ora caleṃ	Cars should go on the left side of the road.
cārom ora	Indcl	cāradike, cāridike	cāradika +e, cāridika +e	suttalu	sutta +alu	around	ribana moma-battī ke cārom ora baṃdhā huā hai.	The ribbon is tied around the candle.
bīca	Indcl	mājha- khāne	mājha- khāna +e	madhya	Indcl	between	nāva samudra ke bīca taira rahī hai.	The boat is floating in the middle of the ocean.
pāsa	Indcl	pāše, kāche	pāśa +e, kācha +e	hattira	Indcl	near	hara mausama meṃ govā ke pāsa āpako parosane ke lie kucha khāsa hai.	In every weather Goa has something special to serve you.
dūra	Indcl	dūre	dūra +e	dūra	Indcl	far	-	-

Table 1: Morpho-lexical properties of relational variants of Geometric Terms in Hindi, Bangla, Kannada and English

³Indeclinable : indcl

⁴All the roots to which -e suffix has been added has an

independent nominal occurrence in present Bangla

⁵-e is the locative marker in Bangla

⁶bāira ; bāhira (sadhu form of Bangla) does not have lexical

Morpho-syntax of geometric terms	Hindi	Bangla	Kannada
Nominal use	Frozen expressions: <i>ladakā pīche se āyā</i>	Bare forms with exceptions <i>dāine</i> and <i>sāmane</i> which are frozen: <i>chele-ṭā pechon theke elo</i>	Bare forms to which suffixes can be added: <i>huḍuga himḍininḍa baṁdanu</i>
Relational marker use	Frozen expressions. ‘ke’ variant of the genitive marker precedes the relational marker: <i>meja ke nīche cyūiṁgama cipakī hai</i>	The ‘-e’ suffix is added to the root form. Genitive suffix is added to the reference object: <i>ṭebil-er nīc-e cuiṁg-gum āṭake āche</i>	The ‘-akke’ suffix is added to the root form. Genitive suffix is added to the reference object: <i>ṭebal-ina keḷage cyūiṁgama aṁṭide</i>
Modifier	The genitive marker optionally follows the geometric form: <i>ghara ke pīche (ke) bāgīce meṁ baḍe baḍe peḍa haiṁ</i>	<i>bāriṛ pechone(r) bāgāne baro baro gāch āche</i>	The genitive marker precedes and follows the geometric form: <i>maneya hindina tōṭadalli doḍḍa maragaḷive</i>

Table 2: A morpho-syntactic properties of geometric items in Hindi, Bangla and Kannada

When the *ādheya* does not touch the *ādhāra* but stays in proximity with *ādhāra* then that spatial relation is said to have *sāmīpya-sambandha*.

- (5) *peḍa ke ūpara cāṁda*
tree.SG GEN above moon.SG.NOM
ā cukā hai
appear.PR.SG.3
‘The moon has appeared above the tree.’

In order to localize a Figure with reference to Ground, Talmy has identified three kinds of expressions:

1. The expression indicates the Figure in touch with the Ground – same as *samyoga-sambandha*. Talmy observes that “the part of the Ground thus named is treated a regular noun” and usually occurs after ‘the’ in such cases,

- (6) The mosaic is on the back of the church.

- (7) The boy is in the front of the line.

2. The expression indicates the Ground’s part to indicate ‘immediate adjacency’ – similar to *sāmīpya-sambandha*. Talmy has observed that in such cases in English, there is no ‘the’ before the geometric terms:

- (8) The bike is in back of/behind the church.

- (9) The police officer is in front of the line.

We omit the third type of expression here because they are poorly represented in English as Talmy has pointed out and do not come under the scope of this paper.

Given the above understanding, we propose two possible situations:

1. The Figure is located in a locus that is in ‘part-whole’ relation with the reference object,

- (10) *peḍa ke ūpara pakṣī hai*
tree GEN on bird.NOM be.PR.SG
‘The bird is on the tree.’

- (11) The bike is behind the church (sentence (8) is repeated here)

2. The Figure is located in a locus that indicates a space denoted by the geometric term with respect to the reference object.

- (12) *peḍa ke ūpara cāṁda*
tree GEN above moon.SG.NOM
hai
be.PR.SG
‘The moon is above the tree.’

- (13) The mosaic is on the back of the church (sentence (6) is repeated here)

attestation in the language

In case of 1, the geometric term mainly denotes the spatio-directional relation between the Figure and the reference object. The directional configuration of the Figure ‘bike’, for example in (8) , with respect to the reference object ‘church’ can be front/back/behind/ near and so on. Hence, this is the relational marker interpretation of the geometric terms.

In case of 2, on the other hand, the geometric term specifies a location. The location, although underspecified, indicates either directional or spatial position of the Figure. For example, the position of ‘cāmda’ (moon) in (12) is in a space which is above the ‘peḍa’ (tree). This is the nominal use of the geometric terms. Such terms can be modified just like any other noun as shown in (14)-(17).

- (14) **peḍa ke thīka ūpara** cāmda
tree.SG GEN right above moon.NOM
hai
be.PR.SG
‘The moon is right above the tree.’
- (15) **ghara se 4 kilomītara dūra** eka
house.SG ABL 4 kilometre away a
mandira hai
temple.NOM be.PR.SG
‘There is a temple 4 kilometres away from the house.’
- (16) **mandaura se eka mīla nīce** eka
mandaura.NNP ABL a mile below a
choṭā saṃgha hai,
small sangha.NOM be.PR.SG
‘There is a small sangha a mile below the mandaura.’
- (17) **kāmkera se 22 kilomītara āge**
Kanker.NNP ABL 22 kilometre ahead
śurū hotī hai keśakāla kī manorama
start.PR.SG Keshkal.NNP GEN charming
ghāṭī
valley.NOM
‘The charming valley of Keshkal starts 22 kilometres ahead of Kanker.’

4 Abstract Semantic Representation of expressions with Geometric terms

Keeping in tune with our analysis given in section 3, we will present here an abstract semantic representation of the geometric terms. Before we present our representation, we will examine how these terms have been represented in the Hindi dependency treebank. In Hindi grammar, these geometric terms have been considered as part of complex post-positions (Bharati et al., 1995; Kachru,

2006; Koul, 2008; Bharati et al., 2009). In Hindi dependency treebank, however, they are lexically marked as Noun-Space-Time (NST) thus distinguishing them from other post-positions which are tagged as PSP⁷ (Bharati et al., 2007). The tag NST indicates that these lexical items are nouns denoting space and time, but their role in the context is not that of a noun. The semantic relation that has been annotated for the sentence “*ladakā ghara ke bāhara khadā hai.*” where ‘bāhara’ is an NST as shown in Figure-1.

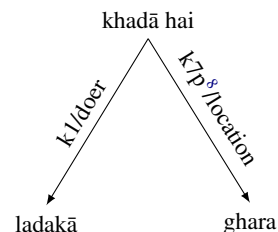


Figure 1: Basic dependency tree for sentence: *ladakā ghara ke bāhara khadā hai.*

In dependency tree structures the content words are represented as nodes, the verb is the head of the tree (in case of simple sentences) and relational markers are all semantic labels on the edges connecting the nodes as shown in Figure-1. The convention of Hindi Dependency treebank is that the syntactico-semantic relations are marked among the chunks. Each chunk has a head and the relation is to be understood between the heads of the chunk. The meaning of this structure is that ‘*ke bāhara*’ (which is annotated as PSP_NST at the POS level) is a relational marker and conveys ‘*deśādhikaraṇa*’ (locative) information. This analysis implies ‘*ghara*’ (house) the location where the boy stands. This is a misleading analysis because the boy does not stand at the house rather he stands in a place which is outside the house. ‘*bāhara*’ conveys that information but the analysis in the treebank does not explicitly capture that.

According to our interpretation of geometric terms, the representation of relational and nominal/modificational variants are postulated differently. Figure-2 presents the relational marker variant of geometric terms. For the sentence ‘*pakṣī peḍa ke ūpara / para baiṭhā hai*’, both the reference object and the Figure are shown as dependents

⁷postposition

⁸k1 and k7p are tags used for marking syntactico-semantic relations in Hindi Dependency Treebank.

of the sentential head which is the verb:

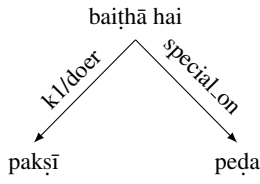


Figure 2: Basic dependency tree for sentence (4)

The semantic label *spatial_on* indicates that *peḍa* is a reference object for the action ‘sitting’ whose *kartā* is *pakṣī* (bird). The labels for geometric terms are given in Table-3. Semantic classes in terms of positive and negative values for features are also specified.

The nominal/modificational variant of geometric terms are represented in the form of constructions. Constructions in Construction Grammar (Fillmore, 1988) are linguistic patterns in which some aspect of its form or its meaning cannot be predicted from its component parts. In Indian Grammatical Tradition, it is called *vṛtti*. *Vṛtti* is defined as ‘*parārthābhidhānam vṛtīḥ*’ – an aggregated word-form that gives a sense which is different from the literal sense of its constituents (Joshi et al., 1990; Sharma and Sharma, 1982).

In our construction, geometric terms entail a space which is not lexically expressed but whose geometric relation to the reference object is conveyed by the geometric terms. The linguistic patterns can be represented as a template and the template is assigned a meaning. For example, the following template defines the form-meaning pair for the nominal/modificational usage of the geometric terms:

$$[X]_{\text{ref-obj}} \text{ke_Term}_{\text{spatial/directional-ground}} [ke/vālā][Y]_{\text{loc}} \quad (1)$$

where *X* and the space denoted by the Term are not in a part-whole relationship.

The subscripts define the meaning and *X*, *ke_Term* and *Y* are variables for physical expressions in 1.

For the sentence (5), the template-1 can be instantiated as follows:

$$(18) \quad \text{peḍa}_{\text{ref-obj}} \text{ke_ūpara}_{\text{spatial/directional-ground}} \text{cāṁda}_{\text{fig}} \text{hai}$$

Following are more examples:

$$(19) \quad \text{sāmane}_{\text{spatial/directional-ground}} /$$

$$\text{cārom} \quad \text{ora}_{\text{spatial/directional-ground}} / \\ \text{pīche}_{\text{spatial/directional-ground}} \text{baḍe} \text{baḍe} \\ \text{peḍa}_{\text{fig}} \text{haiṁ}$$

$$(20) \quad \text{ghara}_{\text{ref-obj}} \quad \text{ke} \\ \text{sāmane}_{\text{spatial/directional-ground}} / \\ \text{cārom} \quad \text{ora}_{\text{spatial/directional-ground}} / \\ \text{pīche}_{\text{spatial/directional-ground}} \text{baḍe} \text{baḍe} \\ \text{peḍa}_{\text{fig}} \text{haiṁ}$$

In the above sentences, ‘big trees’ are the Figure/*ādheya* and the space denoted by the geometric terms *sāmane* / *pāsa* / *pīche* is the Ground / *ādhāra*. In sentences (20) *ghara* is a reference point with respect to which ‘*ādhā*’ has to be interpreted. In (19), the reference point is not explicitly mentioned, and it is from the context the reference point has to be determined. For the above sentences the final part of the template $[ke/vālā][Y]_{\text{loc}}$ is null, and therefore they represent nominal variants. For (21) and (22) below, the construction is used as a modifier and the full template is instantiated as shown below:

$$(21) \quad \text{ghara}_{\text{ref-obj}} \quad \text{ke} \\ \text{sāmane}_{\text{spatial/directional-ground}} / \\ \text{cārom} \quad \text{ora}_{\text{spatial/directional-ground}} / \\ \text{pīche}_{\text{spatial/directional-ground}} \text{ke} \text{bagīce}_{\text{loc}} \\ \text{meṁ} \text{baḍe} \text{baḍe} \text{peḍa}_{\text{fig}} \text{haiṁ}$$

$$(22) \quad \text{ghara}_{\text{ref-obj}} \quad \text{ke} \\ \text{sāmane}_{\text{spatial/directional-ground}} / \\ \text{cārom} \quad \text{ora}_{\text{spatial/directional-ground}} / \\ \text{pīche}_{\text{spatial/directional-ground}} \text{vāle} \text{bagīce}_{\text{loc}} \\ \text{meṁ} \text{baḍe} \text{baḍe} \text{peḍa}_{\text{fig}} \text{haiṁ}$$

5 Developing a Corpus Search cum Annotation Interface for Geometric Terms

From the above discussion, it has become clear that the semantics of geometric terms are quite chequered. We have done a corpus study to understand how these terms are translated into English. We present here example sentences for 4 geometric terms in Table-4.

From Table-4, we find that *nīce* has been translated into as many as 5 different words while *sāmane* into 4 different words. There is also one case of idiom and one phrasal verb attested in the data. We get the adjectival form of *bāhara* as *bāharī*.

In order to be able to disambiguate the geometric terms so that we get appropriate translations into

Hindi	Bangla	Kannada	Semantic Class	Semantic Label
andara	bhitare	oḷage	(R ⁹) Loc: Interior; Con +, (N/M ¹⁰) Loc: Interior; Con -	Spatial inside
bāhara	bāhira, bāire	horage	(R) Loc: Exterior; Con - (N/M) Loc: Exterior; Con -	Spatial outside
āge	āge	murnde	(R) Loc: Anterior; Dir +	directional ahead
sāmane	sāmane	eduru	(R) Loc: Anterior; Dir +	directional front facing
pīche	pichane	himde	(R) Loc: Posterior; Dir +	directional behind
ūpara	opara, opore, upare	mēle	(R) Loc: Superior; Dir +; Con + (N/M) Loc: Superior; Dir +; Con -	Directional on
nīce	nīce	keḷage	(R) Loc: Interior; Dir +; Con + (N/M) Loc: Interior; Dir +; Con -	Directional under
dāyem	dāine	balakke	(R) Loc: Right_side; Dir +; Con + (N/M) Loc: Right_side; Dir +; Con -	Directional left
bāyem	bāMye	edakke	(R) Loc: Left_side; Dir +; Con + (N/M) Loc: Left_side; Dir +; Con -	Directional right
cārom ora	cāradik, cāridike	suttalu	(R) Loc: Circumferential; Con + (N/M) Loc: Circumferential; Con -	Directional around
bīca	mājkhāne	madhya	(R) Loc: Medial; Con + (N/M) Loc: Medial; Con -	Spatial between
pāsa	pāše, kāche	hattira	(R) Approx Spatial Proximity: +; Con + (N/M) Approx Spatial Proximity: +; Con -	Spatial near
dūra	dūre	dūra	(R) Approx Spatial Proximity: -; Con + (N/M) Approx Spatial Proximity: -; Con -	Spatial far

Table 3: Semantic Classification of the geometric terms. The binary feature ‘Con’ defines the physical ‘contact’ of the ādheya with the ādhāra; ‘Dir’ and ‘Loc’ specify ‘direction’ and ‘location’ respectively.

English, we need to understand the context of occurrence of these terms more thoroughly. One way of doing this is to study their usage in corpora, both monolingual and Indian language-English bilingual corpora. To facilitate this study we propose to design an interface for intelligent search of these terms in different contexts. For that purpose, we create a database of existing annotated monolingual corpora as well as a database of parallel corpora. For monolingual corpora, we use POS tags, actual lexical items and syntactic relations as cues for searching the database. When we set out to do that, we find that the geometric terms irrespective of their functional status are annotated as NST in the existing corpora (Jha, 2012). In order to make the search more meaningful, we have integrated an annotation facility to the search interface for annotating the geometric term for relation markers, nouns and modifiers. This will help in future to

search these terms for their different functions.

6 Conclusion

The paper has examined the spatio-directional geometric terms and their semantics in a great detail mainly for Hindi and also for Bangla and Kannada. We have observed that even though these geometric terms have some morpho-syntactic differences in these three languages they are very much in alignment in terms of interpretations. This paper is the beginning of an in-depth study of geometric terms of Indian languages. There remains much work to be done in laying out systematically the subtler differences among apparently close terms such as *āge*, *sāmane*; *bāhara*, *sāmane* and so on. For example, we can almost interchangeably say the following two sentences:

⁹‘R’: Relational

¹⁰‘N’: Nominal, ‘M’: Modifier

- (23) ghara ke bāhara gāḍī khaḍī hai
house GEN outside car.NOM park.PR.SG
'The car is parked outside the house.'
- (24) ghara ke sāmāne gāḍī
house.SG GEN in front of car.NOM
khaḍī hai
park.PR.SG
'The car is parked in front of the house.'

But that is not true for the following pair of sentences:

- (25) a. meja ke sāmāne kursī
table GEN in front of chair.NOM
hai
be.PR.SG
'A chair is kept in front of the table.'
- b. *meja ke bāhara kursī
table GEN outside chair.NOM
hai
be.PR.SG
'*A chair is kept outside the table.'

It appears that the semantics of reference objects are playing a role in licensing the geometric terms. This paper draws the conclusion that in Indian languages (at least for those under consideration), spatio-directional geometric terms play three roles: relational, nominal and modificational. We have proposed to design an interface for annotating geometric terms for their different interpretations. The information can be useful for Natural Language Understanding, Natural Language Generation and knowledge rich Machine Translation.

Acknowledgement

We thank Mr Prateek Saxena for his active participation in developing the Annotation cum Search Interface for the geometric terms presented in this paper.

References

- Akshar Bharati, Vineet Chaitanya, and Rajeev Sangal. 1995. *Natural Language Processing: a Paninian Perspective*. Prentice-Hall of India New Delhi.
- Akshar Bharati, Rajeev Sangal, and Dipti M Sharma. 2007. Ssf: Shakti standard format guide. *Language Technologies Research Centre, International Institute of Information Technology, Hyderabad, India*, pages 1–25.
- Akshara Bharati, Dipti Misra Sharma, Samar Husain, Lakshmi Bai, Rafiya Begam, and Rajeev Sangal. 2009. Anncorra: Treebanks for indian languages, guidelines for annotating hindi treebank.

Charles J Fillmore. 1988. The mechanisms of "construction grammar". In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 35–55.

Jean-Michel Fortis and Benjamin Fagard. 2010. Space in language. In *Leipzig school summe*.

Girish Nath Jha. 2012. The tdil program and the indian language corpora initiative. In *Language Resources and Evaluation Conference*.

Shivram Dattatray Joshi, J. A. F. Roodbergen, et al. 1990. *Patañjali's Vyākaraa-Mahābhāya, Samarthāhnikā*, volume II. Bhandarkar Oriental Research Institute, Pune.

Yamuna Kachru. 2006. *Hindi*, volume 12. John Benjamins Publishing.

Omkar N Koul. 2008. *Modern Hindi Grammar*. Dunwoody Press Springfield, VA.

Barbara Landau. 2017. Update on "what" and "where" in spatial language: A new division of labor for spatial terms. *Cognitive science*, 41:321–350.

Stephen Levinson, Sérgio Meira, The Language, and Cognition Group. 2003. 'natural concepts' in the spatial topological domain-adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, pages 485–516.

Girdhar Sharma and Parmeshwaranand Sharma. 1982. *Laghu Siddhant Kaumudi Part 02*. Motilal Banrassidas Pvt Ltd.

Bhimsen Shastri. 1926. *Laghusiddhāntakaumudī bhaimi-vyākhyā*. Delhi, Bhaimi Prakashan.

KA Subramania Iyer. 1971. *The Vakyapadiya of Bhartrhari, Chapter Iii, Pt. i*. Poona:[Deccan College Postgraduate and Research Institute.

Leonard Talmy. 1983. How language structures space. In *Spatial orientation*, pages 225–282. Springer.

A Appendix-1

Glosses for the Hindi, Bangla and Kannada examples given in Table-2.

Hindi: Nominal use:-

- (26) ladakā pīche se āyā
boy.SG.NOM behind ABL come.3.SG.PT
'The boy came from behind.'

Bangla: Nominal use:-

- (27) chele-ṭā pechon theke elo
boy.SG.NOM behind ABL come.3.SG.PT
'The boy came from behind.'

Kannada: Nominal use:-

- (28) huḍuga hiṁḍininṁḍa baṁḍanu
 boy.SG.NOM behind.ABL come.3.SG.PT
 ‘The boy came from behind.’

‘There exist big trees in the garden which is behind the house.’

Hindi: Relational marker use:-

- (29) meja ke nīche
 table.SG GEN under
 cyūṁgama cipakī hai
 Chewing gum.SG.3.NOM stick.PR.SG.3
 ‘There is a chewing gum stuck under the table.’

Bangla: Relational marker use:-

- (30) ṭebil-er nīc-e
 table.SG.GEN under
 cuing-gum āṭake āche
 Chewing gum.SG.3.NOM stick.PR.SG.3
 ‘There is a chewing gum stuck under the table.’

Kannada: Relational marker use:-

- (31) ṭebal-ina keḷage
 table.SG.GEN under+towards
 cyūṁgama aṁṭide
 Chewing gum.SG.3.NOM stick.PR.SG.3
 ‘There is a chewing gum stuck under the table.’

Hindi: Modifier use:-

- (32) gara ke pīche (ke) bāgīce
 house.SG GEN behind GEN garden.SG
 meṁ baḍe baḍe
 LOC big.PL.NOM big.PL.NOM
 peḍa haiṁ
 tree.PL.3.NOM be.PR.3.PL
 ‘There are big trees in the garden which is behind the house.’

Bangla: Modifier use:-

- (33) bāriṅ pechone(r)
 house.SG.GEN behind.GEN
 bāgāne baro baro
 garden.SG.LOC big.PL.NOM big.PL.NOM
 gāch āche
 tree.PL.3.NOM be.PR.3.PL
 ‘There are big trees in the garden which is behind the house.’

Kannada: Modifier use:-

- (34) maneya hindina
 house.SG.GEN behind.GEN
 tōṭadalli doḍḍa
 garden.SG.LOC big.PL.NOM
 maragaḷive
 tree.PL.3.NOM, exist.PR.3.PL

down, below, underground, under, underneath	nīce	Tag
In the same fashion Son river falls 300 metres down .	usī tarja meṃ sona nadī 300 phīṭa nīce giratī hai.	N
While going to Kedarnath snow mass are seen slipping here and there below the feet .	kedāranātha jāte samaya pairom ke nīce yatra-tatra hima rāśi khisakatī dikhāī paḍatī hai.	M
In Bucharest city there are underground trains as well which are called as 'Metro'.	bukhāresta śahara meṃ jamīna ke nīce calane vālī relem bhī haiṃ, jinheṃ 'metro' kahā jātā hai.	N
There are special kind of shoes for skiing, under which a long metal ski board is attached.	Skīṃga ke lie viśeṣa prakāra ke jūte hote haiṃ, jinake nīce dhātu kā banā lambā skī bleḍa lagā hotā hai.	R
Here underneath a peepal tree , Shree Krishna sat in a lugubrious pose.	yahāṃ eka pīpala ke peḍa ke nīce śrīkr̥ṣṇa viśādamaya mudrā meṃ baiṭhe the.	N
ahead, forward	āge	
The charming valley of Keshkal starts 22 kilometres ahead of Kanker .	kāṃkera se 22 kilomīṭara āge śurū hotī hai keśakāla kī manorama ghāṭī	N
Taking deep breaths making our breath regular, we were going forward .	gahare śvāsa bharakara apanī sāmsoṃ ko niyamita karate hue hama āge badha rahe the .	N
out, outer, outside	bāhara	
Picnic spots adorned with cedars and rivers and streams are out of population .	ābādī se bāhara nadīnāloṃ va devadāroṃ se saje pikanika sthala haiṃ.	N
Installed in the Atishay Kshetra the outer structure of this temple is extremely grand.	Atiśaya kṣetra meṃ sthāpita isa maṃdira kā bāharī svarūpa atyaṃta bhavya hai.	M
Diwan-e-Khas looks like a one-storey building from outside but from inside it is double storied.	dīvāna-e-khāsa imārata bāhara se dekhane meṃ eka maṃjilā pratīta hotī hai magara aṃdara se domaṃjilā hai.	N
before, in front of, to the fore, out	sāmāne	
Then Mahalaxmi had appeared before him with a lotus in her hand.	taba mahālakṣmī hātha meṃ kamala dhāraṇa kie hue unake sāmāne prakāṭa huī thīṃ.	R
An extremely attractive pillar is installed in front of the temple .	maṃdira ke sāmāne eka atyaṃta ākarṣaka staṃbha sthāpita hai	N
During the excavation of Vaishali it came to the fore that it has had an impressive history.	vaiśālī meṃ milī khudāī meṃ mile avāśeṣoṃ se yaha bāta sāmāne āī hai ke isakā eka prabhāvaśālī itihāsa rahā hai.	Idiom
At the confluence place of the rivers which is right before the temple are beautiful but small falls.	nadiyoṃ ke saṃgama sthala para jo maṃdira ke thīka sāmāne haiṃ suṃdara kintu choṭā prapāta hai.	N
Although quarrying of the Stupa is still not complete yet its 104 ft high structure has come out .	stūpa kā utkhanana-kārya yadyapi abhī pūrā nahīṃ huā hai, tathāpi, isakī 104 phīṭa ūṃcī saṃracanā sāmāne ā cukī hai .	Phrasal verb
The part in the front of it has fallen.	inakā sāmāne kā hissā gira gayā hai.	M

Table 4: Example sentences from the corpus for 4 geometric terms