



# International Journal of Linguistics & Computing Research

## Problem of extracting case relations from sentences of a morphologically rich Indian language

Ranjan Kumar Das

Centre for Development of Advanced Computing, Pune

[dasranjan257@gmail.com](mailto:dasranjan257@gmail.com)

*Abstract: Case relation is a semantico-syntactic relationship in which abstract case features (roles) are mapped with their surface (morphological and syntactic) forms. Semantic parameters such as agency and affectedness are known to affect the realization of case-relationships. Arguments of a predicate (e.g. verb) play some semantic role (= thematic role) by virtue of the meaning of the predicate and they are assigned 'Case' in the sentence. The nouns as arguments of verbs bear case-roles reflected through case-forms. Case-markers, post-positions or prepositions represent case-forms. This paper examines the formal realization of case relationships in Hindi and Odia from the perspective of an automated analysis by a natural language processing system. It highlights how multiple mapping of case-forms with the case-relations pose a problem of extraction of case-relationship without ambiguity. Further it attempts to find a solution to this problem. It takes examples from Hindi and Odia languages.*

*Key words: case relation, case form, predicate, argument, thematic role*

### I. INTRODUCTION

The notion of case is a universal feature of language, as if it forms a part of the human language faculty. Case theory provides an approach to the representation of sentence meaning and case is important in accounting for the way the structure of sentences relates to that of meanings. In other words, case relates the deep semantic-logical structure of sentence to the surface formal (syntactic) structure of language. Although different languages have different case systems, case as a grammatical property/principle is universal.

Fillmore (1968) characterized the relationships between the verb and noun phrase as "semantically relevant syntactic relationships" and called them *cases*. [1][2] The case assignment comes from the deep structure, even though the surface structure is different. So, for example, the sentence "John opened the door with the key" is semantically equivalent to "The door was opened by John with the key."

*Case relations* [Cook 1979; Bruce 1975] describe the relation between a verb and the other components (typically nouns) of a single *proposition*. [3][4] The semantic content of a sentence is called its *proposition* and it consists of a *predicate* denoting an activity or event and arguments on which the predicate says something. Argument is a term borrowed by linguistics from philosophy (predicate calculus) to describe the role played by particular types of referring expressions (i.e. a thing, person etc.) in the semantic structure of sentences.

For example, in the sentence *Rama killed Ravana*, *killed* is the predicate and *Rama* and *Ravana* are the arguments. Arguments of a predicate (e.g. verb) play some semantic role (also called thematic role or  $\theta$ -role) by virtue of the meaning of the predicate. Case relation is a middle layer function between semantic and syntactic relationships in which abstract case features (roles) are mapped with their surface (morphological and syntactic) reflex.

Principles and Parameters theory (Chomsky 1981) considers case-theory and theta-theory as the essential components of grammatical principles which play an important role in assigning semantic and structural wellformedness to the expressions of

language. [5] *Case* is considered as a general property of language to be associated with a syntactic phenomenon — the grammatical function (relations) of NPs, and reflects semantic relationships. Thus, all nouns in a sentence indeed get case – overtly or covertly.

In the case theory of Chomsky (1981), NPs are given *Case* if and only if they appear in specific positions in the sentence. In particular, nominative case is assigned in the specifier of finite verb, and accusative case is assigned to a complement/sister of the verb. Prepositions also assign what is often called ‘dative’ or other cases to their complement NPs. [5]

The importance of case-relation extraction (= case analysis) is that it forms an essential part of interpretation (or parsing) of source language sentences which only enables an appropriate mapping into target language case forms represented through prepositions or postpositions. The problem is nontrivial even in case of simple sentences, and needs attention.

## II. THEORETICAL BACKGROUND

Nouns of sentence bear some particular grammatical relations like *subject*, *object*, *indirect object*, etc. These are *not* the same as *thematic relations* (theta-relations), but have some correspondence with them. Grammatical relations represent how an NP (noun phrase) is functioning in the sentence syntactically. The morphology associated with grammatical relations is called *case*. For example, the *nominative case* is found with subjects, and the *accusative case* is found with objects.

However, representation of case relations differs from language to language. In morphologically rich languages like Hindi, nouns take special forms and there are special markers. On the other hand, English being a morphologically poor language, NPs in *sentences* bear no obvious case markings. Grammatical relations are represented by the *position* of the noun in the sentence. Following sentences (1) and (2) demonstrate case marking in English.

- 1) Ram slapped Shyam.  
Ram-subject slapped-PAST Shyam-object  
[Ram-NOM slapped-PAST Shyam-ACC]
- 2) Shyam slapped Ram.  
Shyam-subject slapped-PAST Ram-object  
[Shyam-NOM slapped-PAST Ram-ACC]

Both the *nominative* and accusative cases are present semantically/logically, but are marked by null (zero) marker. The information is coded in the positions of these NPs. This is called *abstract Case*. (Abstract Case normally has a capital C to distinguish it from morphological case).

In morphologically rich languages like Sanskrit, Hindi or Odia, nouns which bear particular grammatical relations (or case relations) take special forms (called as case forms) marked with overt suffixes (case endings) as in sentence (3).

- 3) Ram *ne* Shyam *ko* mara. (Hindi)

Ram Shyam-*ku* marila. (Odia)

Ram-subject slapped-PAST Shyam-object

[Ram-NOM slapped-PAST Shyam-ACC]

In Hindi, subjects are marked with zero suffix or an ergative marker ‘*ne*’, objects are marked with *-ko* /zero suffix and indirect objects and certain adjuncts with *-ko*. Similarly, in Odia, subjects are marked with zero suffix, objects are marked with *-ku* /zero and indirect objects and certain adjuncts with *-ku*.

## II.I ARGUMENT STRUCTURE, THETA (Θ)-THEORY, CASE THEORY AND PANINIAN NOTION OF KARAKA

Argument structure is the pattern of underlying relations between a predicate and its arguments / dependents. It is at the base of syntactic theory and syntax-semantics interface.

Theta theory is concerned with the assignment of *thematic roles* (= semantic roles) to sentential constituents. By using lexical information like theta roles, we can stop the X-bar rules from generating ungrammatical sentences. One way of encoding selectional restrictions is through the use of *thematic relations*. These are particular semantic terms that are used to describe the role that the argument plays with respect to the predicate.

The *predicate* defines the relation between the individuals being talked about and the real world as well as with each other. The *arguments* are the entities who are participating in the relation. *Argument Structure* defines the number of arguments that a predicate takes. Thus, an intransitive predicate takes only one argument, a transitive predicate takes two arguments, and a ditransitive predicate takes three arguments. Some common thematic relations are:

- Agent*: The doer of an action (under some definitions must be capable of volition.)
- Experiencer*: The argument that perceives or experiences an event or state.
- Theme (patient)*: The element undergoing the action or change of state.
- Goal*: The end point of a movement.
- Recipient*: A special kind of goal, found with verbs of possession (e.g., give).
- Source*: The starting point of a movement.
- Location*: The place an action or state occurs.
- Instrument*: A tool with which an action is performed.
- Benefactive*: The entity for whose benefit the action is performed.

In the sentence (4), we show argument structure and case.

- 4) Ram slapped Shyam.  
Ram-Agent slapped-Predicate Shyam-Theme/Patient  
SLAP <Agent, Theme>  
Ram-NOM slapped-PAST Shyam-ACC

There is no one-to-one relationship between thematic relations and arguments. However, a *theta role* (or  $\theta$  role) maps one-to-one with arguments. Theta roles are bundles of thematic relations that cluster on one argument. Often particular theta roles are referred to by the most prominent thematic relation that they contain. The thematic relations are things like agent, theme, goal, etc. but theta roles are bundles of thematic relations assigned to a particular argument.

The *Theta-Criterion* of Principles and Parameters theory (Chomsky 1981) defines that

- a) Each argument is assigned one and only one theta role.
- b) Each theta role is assigned to one and only one argument.

The thematic relation of an NP with the verb plays a role in semantic interpretation of the sentence. [5]

The **case theory** of Principles and Parameters theory (Chomsky 1981) defines that

- (a) Case is assigned under government, and
- (b) No NP which is lexically realized will be without a case (case-filter). [5]

Thus, all NPs in a sentence/clause must be assigned case.

The argument-verb relation (related to Chomskyan  $\theta$ -role) is captured in Paninian notion of *Karaka* (case) relation and expressed in terms of *Vibhaktis* (or case marking) which include case suffixes and post-positions in Indian languages. A nominal in Hindi or Odia is obligatorily inflected for case (*karaka*) along with number (Das 2003: 96). Of course, both can have a null value. In Odia, these are expressed by means of case markers or *Vibhaktis* which include both case-suffixes (bound forms) and post-positions (free forms) (Das 2003: 97-98). [6]

The  $\theta$ -roles are like *agent, theme, experiencer, recipient, goal, source, locus, instrument* and many others, *not fixed in number*. These universal relations are expressed in surface structure of a particular language by some specific mechanism. In Hindi, these are expressed by means of case markers or *Vibhaktis* as post-positions. In Odia, case markers or *Vibhaktis* include both case-suffixes and post-positions (Das 2003: 97-98). [6]

Altogether there are six groups of case markers in traditional grammars to reflect the case relations operative in Hindi and Odia following Pāṇinian framework which is more suitable for these languages than Fillmorean (which is deeper) and Chomskyan (where structural case is assigned to positions) approaches for describing case system. 'Case' in the sense of Paninian *Karaka*, as a syntactico-semantic role, is a bit shallow (surface structure oriented) than  $\theta$ -role.

A tentative correspondence between Karakas and  $\theta$ -roles (table-1) and then Karakas and case markers (*Vibhaktis*) in Odia and Hindi (table-2) are shown below.

(1)

Correspondence between Case-roles and Theta-roles			
Case-Role		index	$\theta$ -Role
Karta	Nominative	(K <sub>1</sub> )	Agent, experiencer, force
Karma	Accusative	(K <sub>2</sub> )	Theme, patient, content, result, goal
Karana	Instrumental	(K <sub>3</sub> )	Instrument
Sampradana	Dative	(K <sub>4</sub> )	Beneficiary
Apadana	Ablative	(K <sub>5</sub> )	Source
	Genitive	(K <sub>6</sub> )	
Adhikarana	Locative	(K <sub>7</sub> )	Location in time and space

Genitive is not a case-role in Paninian sense. It shows a relationship between two nominals. On the basis of surface similarity of genitive suffixes with case suffixes (being a relational marker and different from number suffixes) it is included as a *Vibhakti*.

In this paper, the term 'K-role' is used for *Karaka* role or case role. In Hindi and Odia sentences, any argument of any verb has to have any one of these six K-roles (leaving K<sub>6</sub>). These six K-roles are the only grammatically (formally) relevant relations in these languages, although we can take a larger set of relations leading to more ambiguity for classification. The situation may be different for other languages.

(2) Correspondence between Case-roles and Case markers in Odia

Case-Role	Case-Markers of Odia	Case-Markers of Hindi
(K <sub>1</sub> )	Φ	Φ, ne
(K <sub>2</sub> )	Φ, -ku, -ki	Φ, ko
(K <sub>3</sub> )	-re, द्वारा, dei	se, ke द्वारा
(K <sub>4</sub> )	-ku, -ki, paī	ko, ke lie
(K <sub>5</sub> )	-ru, - <sup>th</sup> aru, - <sup>th</sup> ū	se
(K <sub>6</sub> )	-rə, -kə(rə)	ka, ke, kii
(K <sub>7</sub> )	-re, - <sup>th</sup> are, - <sup>th</sup> i	me, pər

### III. THE MAPPING PROBLEM

*Mapping problem type-1:*

From the table-2, we can point out that some of the case suffixes are mapped with multiple case-relations and thus pose a problem for analysis. For instance, in Hindi /ko/ is mapped with K<sub>2</sub> and K<sub>4</sub>; /se/ is mapped with K<sub>3</sub> and K<sub>5</sub>; and in Odia /-ku/ (and /-ki/) is mapped with K<sub>2</sub> and K<sub>4</sub>; /-re/ is mapped with K<sub>3</sub> and K<sub>7</sub>, as in the following sentences.

5) Ram *ne* Shyam *ko* mara. (Hindi)

Ram Shyamə-ku marila. (Odia)

Ram-K<sub>1</sub> Shyam-K<sub>2</sub> slapped-PAST

Ram-subject slapped-PAST Shyam-object

6) Ram *ne* Shyam *ko* *əpani* kitab dia. (Hindi)

Ram Shyamə-*ku* *nijə* bəhi dela. (Odia)

Ram-K<sub>1</sub> Shyam-K<sub>4</sub> his-book-K<sub>2</sub> give-PAST

Ram-subject give-PAST Shyam-object\_indirect his book-object\_direct

Here there is ambiguity of dative and accusative case marking.

7) Ram chaku *se* seo kat-rəha-hai. (Hindi)

Ram-K<sub>1</sub> knife-K<sub>3</sub> apple-K<sub>2</sub> cut-PRE-PROG

‘Ram is cutting apple with a knife.’

8) Ram skul *se* a-rəha-hai. (Hindi)

Ram-K<sub>1</sub> school-K<sub>5</sub> come-PRE-PROG

‘Ram is coming from school.’

Here the ambiguity is of instrumental and ablative case marking in Hindi.

9) Ram chhuri-*re* seo kat-u-chhi. (Odia)

Ram-K<sub>1</sub> knife-K<sub>3</sub> apple-K<sub>2</sub> cut-PRE-PROG

‘Ram is cutting apple with a knife.’

10) Ram rasta-*re* bəs-i-chhi. (Odia)

Ram-K<sub>1</sub> road-K<sub>7</sub> sit-PRE-PRF

‘Ram is sitting on the road.’

Here the ambiguity is of instrumental and locative case marking in Odia.

To devise an automated case analyzer, we need to take care of these ambiguities. Any analyzer that considers only the surficial case-forms will fail to resolve case ambiguity of this type. The problem can be addressed by considering both the case-forms and the argument structure of the verb. It is demonstrated with examples from Odia.

For a sentence in Odia, six K-roles mentioned above are the semantico-syntactic requirements of a verb. However, all verbs do not require all and the same K-roles. The degree of requirement or relevance varies according to the internal meaning of the verbs. Sometimes, the verb phrase as a group decides the K-role of some arguments. Moreover, some K-roles like K<sub>1</sub>, K<sub>2</sub> are more relevant (even necessary) than others which are optional. Accordingly, if we sub-categorize verbs, case-role assignment can be done more precisely. For illustration consider the examples in (11) and corresponding K-role structures in (12). [6]

11)

a) ma            dɔɔɔtable|e       piɔdare       so-il-e

mother 10 o'clock-LOC veranda-LOC sleep-TAM-AGR

‘Mother slept on veranda at 10 o'clock’.

b) ma            dɔɔɔtablele       piɔdare       pila ku sua-il-e  
mother 10 o'clock-LOC veranda-LOC child-ACC  
sleep-CAUSE-TAM-AGR

‘Mother caused the child sleep on veranda at 10 o'clock.’

c) ma            b<sup>h</sup>atɔ            k<sup>h</sup>a-il-e  
mother rice-ACC eat – TAM-AGR  
‘Mother ate rice.’

d) ma            pila ku            b<sup>h</sup>atɔ            k<sup>h</sup>ua-il-e  
mother child-DAT rice-ACC eat- CAUSE- TAM-AGR  
‘Mother fed the child with rice’.

e) se dɔɔɔta bele       tebul re            bɔɔhi            rɔk<sup>h</sup>-il-e  
he 10 o'clock-LOC table-LOC book put-  
TAM-AGR

‘He put the book on the table at 10 o'clock’.

f) se tebul re            cakɔɔɔ dwara bɔɔhi            rɔk<sup>h</sup>a-il-e  
he table-LOC servant-INST book put-  
CAUSE-TAM-AGR

‘He made the servant put the book on the table’.

12)

a. K<sub>1</sub> (K<sub>7</sub>.Temporal)(K<sub>7</sub>.Spatial)[so-]<sub>v</sub> \*K<sub>2</sub>\*K<sub>3</sub>\*K<sub>4</sub>\*K<sub>5</sub>

b. K<sub>1</sub> (K<sub>7</sub>-T) (K<sub>7</sub>-S) K<sub>2</sub> [sua-]<sub>v</sub> \*K<sub>3</sub>\*K<sub>4</sub> \*K<sub>5</sub>

c. K<sub>1</sub> (K<sub>7</sub>-T) (K<sub>7</sub>-S) K<sub>2</sub> [k<sup>h</sup>a-]<sub>v</sub> (K<sub>3</sub>) \* K<sub>4</sub> (K<sub>5</sub>)

d. K<sub>1</sub> (K<sub>7</sub>-T) (K<sub>7</sub>-S) K<sub>4</sub> K<sub>2</sub> [k<sup>h</sup>ua-]<sub>v</sub> (K<sub>3</sub>) (K<sub>5</sub>)

e. K<sub>1</sub> (K<sub>7</sub>-T) K<sub>7</sub>-S K<sub>2</sub> [rɔk<sup>h</sup>-]<sub>v</sub> \*K<sub>3</sub> \* K<sub>4</sub> \*K<sub>5</sub>

f. K<sub>1</sub> (K<sub>7</sub>-T) K<sub>7</sub>-S K<sub>3</sub> K<sub>2</sub> [rɔk<sup>h</sup>-]<sub>v</sub> \* K<sub>4</sub> \*K<sub>5</sub>

From the illustrations (11) and (12) it is evident that all the verbs do not have same argument structure (case-frame or K-role structure). Even different arguments of a verb do not have same status. Some are mandatory, some are optional (with different degree), and some are even not-possible (marked by *asterisk*). Optionality is marked by simple *brackets*.

The locative K-role (K<sub>7</sub>) with varieties of temporal and spatial can optionally occur almost with all verbs. But it is mandatory in case of verbs like /rɔk<sup>h</sup>-/ ‘put’. A causative verb gets one more mandatory K-role than the root verb which may be intransitive or transitive.

Another observation we can make is that although the case markers for accusative (K<sub>2</sub>) and dative (K<sub>4</sub>) are same /-ku/ and its variant /-ki/, by syntactic behaviour they are different. An

accusative /-ku/ can be deleted, (11 c,d,e,f), whereas it is not for dative /-ku/ (11d). This shows how this case analysis problem can be solved by considering argument structure of verb.

#### Mapping problem type-2:

Another type of case-role ambiguity as a bigger problem arises when one verb form is mapped with multiple argument structures as illustrated in sentences below.

- 13) (a) Ram *ne* dərwaja khol-a. (Hindi)

Ram-K<sub>1</sub> door-K<sub>2</sub> open-PAST

‘Ram opened the door.’

- (b) dərwaja khol-a / khul-gaya. (Hindi)

door-K<sub>1</sub> open-PAST

‘The door opened.’

- 14) (a) The boy hit the wall with a hammer. (English)

- (b) The hammer hit the wall.

Similar phenomenon is seen in Odia. Here there is a mismatch in mapping of logical case role (semantic role) and formal case role. To resolve this problem, more information is required than the simple K-role structure.

Other types of problems other than the above two may include the instances of structural ambiguity.

#### CONCLUSION

The above analysis ultimately demonstrates the problem of unambiguous analysis of case-role in sentences. A case analyzer is useful for semantic analysis required for developing a natural language understanding system.

One problem faced here is that some of the case suffixes have multiple mappings with case-roles. To devise an automated case analyzer, we need to take care of these ambiguities. We need additional information like the behavior of verb-forms and their contexts in sentences/phrases. Another problem arises when a verb-form in a language selects multiple case-frames reflecting different semantic structures. An exhaustive study of the nature of verbs and their sub-classification can be helpful in resolving this ambiguity problem that will help in precise extraction of semantic role of the nominals in a sentence.

#### REFERENCES

- [1] Fillmore, C. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1-88). New York: Holt, Rinehart & Winston.
- [2] Fillmore, C.J. (1977). “The Case for Case Reopened.” *Studies in Syntax and Semantics* 8, 59-81.
- [3] Bruce, B. (1975). “Case Systems for Natural Language.” *Artificial Intelligence* 6, 327-360.
- [4] Cook, W.A. (1979). *Case Grammar: Development of the Matrix Model*. Washington: Georgetown University.

- [5] Chomsky, N. (1981). *Lectures on Government and Binding*. Foris, Dordrecht.
- [6] Das, R.K. (2003) *A Study of Oriya Language: An attempt towards a holistic approach*. Ph.D. thesis, Berhampur University, Odisha.