

Issues with the Unergative/Unaccusative Classification of the Intransitive Verbs

Nitesh Surtani, Khushboo Jha, Soma Paul

Language Technologies Research Center

IIIT Hyderabad

Hyderabad, Andhra Pradesh-500032

{nitesh.surtaniug08, khushboo.saumya}@students.iiit.ac.in, soma@iiit.ac.in

Abstract—The paper abandons a strict two-way sub-classification of intransitive verbs into unaccusative and unergative for Hindi and proposes a distribution plotting of the same in a diffusion chart. The diagnostics tests that Bhatt (2003) applied on Hindi data are ranked for their efficiency of attributing correct sub-class to verbs. The diffusion chart shows that a tripartite classification handles the issue of classification of intransitive verbs in a better manner than the classical binary approach. The tripartite classification is as follows: (1) Verbs that take animate subject and are compatible with adverb of volitionality; (2) Verbs that take animate subject but are not compatible with adverb of volitionality; and (3) Verbs that take inanimate subject. The classification is of immense advantage for various NLP tasks such as machine translation, natural language generation.

I. INTRODUCTION

The Unaccusativity Hypothesis, originally proposed by Perlmutter (1978) distinguishes between the two classes of intransitive verbs; one sub-class of intransitive verbs, known as unergative verbs, entails “willed or volitional acts” (e.g. ‘someone dances, runs, walks’ etc.) and the other sub-class, unaccusative verbs, that implies “unwilled or non-volitional acts” (e.g. ‘something burn, melt, fall, happen’ etc.). The unergative/unaccusative distinction has been shown to exist cross linguistically as illustrated below with examples from Hindi:

- (1) a. *raam-ne khub dauR-aa.*
3P.M.Sg-Erg very Run-Pfv.M.Sg
‘Ram ran a lot.’
b. *glaas TuT-aa.*
M.Sg break- Pfv.M.Sg
‘(The) glass broke.’

The verb “run” in (1a) is a volitional act and it takes a subject that can be marked for ergative¹; while the subject of unaccusative verb never licenses ergative marking on the subject as shown in (1b). Thus sub-classification of intransitive verbs is a significant task for the analysis of Hindi language. The endeavor has bearing on various NLP tasks such as machine translation, natural language generation etc. We take here a case from Bangla-Hindi MT system. Bangla does not mark for ergative case while Hindi does as exemplified in (1). The equivalent of the

sentences in (1) is given in (2) for Bangla where the subject in both sentences remains unmarked.

- (2) a. *raam khub douRo-lo.*
‘Ram ran a lot.’
b. *glaas bhAM-lo.*
‘(The) glass broke.’

Classifying intransitive verbs of (2a) and (2b) into subclasses can result in producing right case marking on the subject in the target language Hindi. Although the Unergativity/Unaccusativity classification has widely been applied for many languages and the distinction is based on a number of linguistic phenomena (called unaccusative diagnostics) that differentiate between the two subclasses of intransitive verbs (Levin & Hovav 1995), a closer look at Hindi data shows that the distinction is not clear-cut for the language. A detailed study by Bhatt (2003) supports the observation for Hindi. This paper argues that a clear-cut two way distinction does not work for Hindi. Instead, a tripartite classification is tried out. We observe that the class of verbs taking animate subjects i.e. unergatives and unaccusative animates are hard to classify because even though their subjects are volitional in nature (being animate), they do not respond in a similar manner for all unaccusativity tests. On the other hand, unaccusative inanimates perform quite well for these tests and are well separated. This observation motivates the following three-fold proposals in the paper:

- (i) Each diagnostic test is assigned a score depending on its authenticity in terms of how efficiently it can diagnose the behavior of a verb.
- (ii) Each verb is given a unaccusativity score depending on its response to the diagnostic tests.
- (iii) Verbs are plotted on a diffusion graph based on their score.

Thus the paper contends that intransitive verbs do not belong to clear cut classes. The position of a verb in the graph will very interestingly signify syntactic behavior of the verb. Thus the plotting of verbs in the graph becomes an innovative measure of classifying intransitive verbs.

The paper is organized as follows: Section 2 discusses the related work. Section 3 discusses various diagnostic tests for Unergativity/Unaccusativity distinction in Hindi and gives observation made for the same. Section 4 discusses the approach for the classification of intransitive verbs. The diffusive distribution of verbs using diffusion chart is shown in Section 5 and discussed in Section 6. Section 7 talks about the conclusion and future work.

¹ In Hindi the ergative marker generally appears on the subject of transitive verbs when the verbs are inflected for perfective tense. The occurrence of ergative marker on the subject of intransitive verbs is context situated and the subject has to be volitional actor. The discussion of contextual factor that licenses the ergative marked subject of intransitive verbs is beyond the scope of the paper.

II. RELATED WORK

The unergative/unaccusative distinction has been shown to exist cross linguistically including many languages e.g. German, Dutch, Hindi etc. Various language specific tests have been proposed as diagnostics for the distinction of the verbs of these two classes. But most of the languages like Turkish (Acartürk and Zeyrek, 2010), German (Sorace, 2000; Kaufmann, 1995), Urdu (Ahmed, 2010) etc. have reported the failure of the unaccusativity tests and the absence of clear-cut distinction between unaccusatives and unergatives. Various diagnostics have been proposed in Bhatt (2003) for Indian languages. We will review those tests in the next section in the light of Hindi data maintaining the original unergativity/unaccusativity distinction.

III. APPLYING DIAGNOSTICS AND REVIEWING THE EFFECTS

A. Diagnostic Tests:

We have examined the seven tests that Bhatt (2003) has proposed in his work and counterexamples are also discussed.

1) *Ergative Subjects*: Unergatives sometimes allow ergative subjects with an ergative case marker ‘*ne*’ esp. when paired with the right adverbials and compound verbs (as in (3a)). On the other hand, Unaccusatives do not allow ergative subjects (as in (3b)).

- (3) a. *raam-ne* *bahut* *naach-aa*.
3P.M.Sg-Erg a lot dance-Pfv
‘Ram danced a lot.’
- b. **raam-ne* *bahut* *ghabraaya*.
3P.M.Sg-Erg a lot panic-Pfv
‘Ram panicked a lot.’

2) *Cognate objects*: These are simply the verb’s noun form. Unergatives verbs sometime allow for Cognate objects (as in (4a)) whereas Unaccusatives do not allow for cognate objects.

- (4) a. *raavan-ne* *bhayaanaka* *hasii* *has-ii*.
3P.M.Sg-Erg horrifying laugh laugh-Pfv
‘Ravan laughed a horrifying laugh.’

3) *Impersonal Passives*: The impersonal passive deletes the subject of an intransitive verb and reduces its valency to zero. Unergatives allow for the impersonal passive (as in (5a)) whereas unaccusatives do not.

- (5) a. *thodii* *der* *aur* *jhool-aa* *jaaye*.
Some time more swing-Pfv go-Sbjv
‘Swing for some more time.’

4) *Past Participial Relatives*: Past participial relatives target the internal/theme argument of the verb, if there is one. The past participial relatives on Unaccusatives have an active syntax taking *hua* ‘be-Pfv’/ *gaya* ‘go-Pfv’ (as in (6b)) whereas unergatives are ungrammatical with past participial relatives (as in (6a)).

- (6) a. **kal* *dauR-aa* *huua* *chhaatra*
yesterday run-Pfv be-Pfv student
‘The student who ran yesterday’
- b. *vahaan* *bandh-aa* *huua* *ladkaa*
there tie-Pfv be-Pfv boy
‘The boy who is tied there’

5) *Inabilitatives*: Inabilitatives describe the inability of the agent towards an action which applies to the class of verbs that undergo the transitivity alternation. Unaccusatives enter the inabilitative with active syntax (as in (7b)) whereas Unergatives do not (as shown in (7a)).

- (7) a. **raam-se* *ramaa* *nahii* *has-ii*.
3P.M.Sg-Instr 3P.F.Sg neg laugh-Pfv.f
‘Ram couldn’t make Rama laugh.’
- b. *raam-se* *ghar* *nahii* *banaa*.
3P.M.Sg-Instr house neg build-Pfv
‘Ram couldn’t build the house.’

6) *Compound Verb Selection*: There seems to be a kind of selection between compound verbs and main verbs. The unaccusative compound verb *jaa* ‘go’ appears most naturally with unaccusatives while Unergatives tend to take transitive compound verbs like *le* ‘take’ / *de* ‘give’ / *daal* ‘did’ and seem unhappy with *jaa* ‘go’ (as in (8a)).

- (8) a. *raam-ne* *pahaar* *chaD* *liyaa*.
3P.M.Sg-Erg mountain climb take-Pfv
‘Ram climbed the mountain.’

7) *Unmarked Subjects for Non-Finite Clauses*: Non-Finite clauses in Hindi do not permit overt unmarked subjects (as in (9a)). But inanimate subjects of the Unaccusative verbs can appear without an overt genitive.

- (9) a. [*raam-ka*/**raam* *tez* *bhaagna*] *zaruurii* *hai*.
3P.M.Sg-Gen/*Nom fast run necessary is
‘It is necessary for Ram to run.’

B. Review of the diagnostics for Hindi

We have made the following observation from the study of diagnostics.

- (i) There is no single distinguishing criterion for subclassifying Hindi intransitive verbs. Some diagnostic tests, however, perform better than the other giving more accurate results.
- (ii) Verbs that take non-volitional inanimate subject perform quite well for the diagnostics.
- (iii) Verbs with animate subject respond fuzzily to the unaccusative tests. For example, let us look at the verb *gir* ‘fall’. It is originally being classified as an unaccusative verb because the subject of the verb is an undergoer undergoing some kind of change of state. When the subject is inanimate, the unaccusativity feature holds as in (10a):

- (10) a. *peD-se* *patta* *giraa*.
tree.M.Sg-Abl leaf fall-Pfv
‘The leaf fell from the tree.’

With such subject, the verbal construction will not be compatible with adverb of volitionality as is true for other unaccusative verb. Therefore the following sentence is illegitimate:

- b. **patta* *jaan-buijh-kar* *giraa*.
leaf deliberately fall-Pfv
‘The leaf deliberately fell.’

However the situation changes when the verb takes an animate human subject. The construction licenses adverb of volitionality as illustrated in (10c):

- c. *raam* *jaan-buijh-kar* *giraa*.
ram.M.Sg deliberately fall-Pfv
‘Ram deliberately fell.’

The above observation guides our approach of handling sub-classification of intransitive verbs which we discuss in the second subsection of the next section.

IV. OUR APPROACH

We propose a two step procedure to handle the sub-classification issue of intransitive verbs.

1) *Ranking the Diagnostics*: Since some diagnostics performed better than the other in capturing the syntactic behavior of the intransitive verbs, we ranked the diagnostic tests in the first step of our approach. Tests for which maximum number of verbs show expected output are attributed higher rank. In order to rank the tests, we applied the diagnostic tests on 134 instances of intransitive verbs and manually checked the performance. We calculate a diagnostic score for each diagnostic test in the following manner:

- (i) Count the correctly and incorrectly identified Unaccusative and Unergative verbs for each diagnostic.
- (ii) Set positive score for each correctly classified and negative score for each incorrectly classified verb for each diagnostic. These scores are manually given depending upon the performance of the diagnostic.
- (iii) Calculate the total score and normalize it by dividing it by the *maximum achievable score*.²

TABLE I. a) Results of the diagnostics b) Score for each +/- instance

Tests	(Unergatives)		(Unaccusatives)		Tests	(Unergatives)		(Unaccusatives)	
	Correct	Wrong	Correct	Wrong		Correct	Wrong	Correct	Wrong
Ergative case	31	16	85	2	Ergative case	2	-1	2	-3
Cognate objects	12	35	86	1	Cognate objects	2	-1	2	-3
Impersonal Passive	38	9	83	4	Impersonal Passive	2	-1	2	-3
Past Participial	31	16	69	18	Past Participial	2	-2	2	-2
Inabilitaives	41	2	56	31	Inabilitaives	2	-2	2	-3
Compound verb	43	4	82	5	Compound verb	2	-1	2	-3
Overt Genitive	47	0	77	13	Overt Genitive	2	-3	2	-3

The Table (1a) shown above gives the number of correctly and incorrectly identified instances for each diagnostic and the scores for the same are assigned in Table (1b). In Hindi, unergatives allow ergative marker *ne* whereas unaccusatives do not. We see that, 31 out of 47 unergatives are correctly marked as they take ergative marker '*ne*'. On the other hand, 85 out of 87 unaccusatives do not take '*ne*' marker and are marked as correct instances. We have assigned a score of (+2) for correctly identified instance for both unaccusative and unergative whereas the negative instances have been penalized with negative scores, (-3) for unaccusative and (-1) for unergatives. Unaccusatives have been penalized more as they do not allow ergative subjects than unergatives which often allow for ergative subjects but do not force it. These scores are set to balance the final score obtained by the diagnostic tests. The final *normalized diagnostic score*³ is calculated using the formula:

$$score = \frac{(N1 \cdot X1 + N2 \cdot X2 + N3 \cdot X3 + N4 \cdot X4)}{(N1 + N2) \cdot X1 + (N3 + N4) \cdot X4}, \text{ where }^4$$

² Maximum achievable score is secured when the result for all the verbs given by the diagnostic matches the expected result of the class of the verbs.

³ Normalised diagnostic score is the score secured by the diagnostic divided by the maximum achievable score.

⁴ N1, N2 = Correct and incorrect instance for unergative

N3, N4 = Correct and incorrect instances for unaccusative

X1, X2 = Score for each correct and incorrect instance of unergative

X3, X4 = Score for each correct and incorrect instance of unaccusative

The diagnostics are then ranked according to their final scores with one having the highest score assigned as Rank 1 and so on as shown in Table 2. The rank of a diagnostic reveals its significance for a particular language and tells how correctly it maps the semantics of the verb to its syntactic behavior. Since the top ranking diagnostic has the highest score, it has the maximum contribution in deciding the Unergative/Unaccusative nature of the intransitive verb.

2) *Tripartite Classification of Verbs*: The performance of the diagnostic tests has been evaluated in terms of the classical unaccusative/unergative distinction of verbs because those tests were designed to capture that binary classification of intransitives. The paper, however, highlights the fact (see examples in (10) for illustration) that when the subject of a popularly known unaccusative verb takes a human subject, the behavior in terms of its response to various tests changes. Therefore, at the second step of our approach, we start with a tripartite classification of intransitives:

Class 1. Verbs that take animate subject and agree with adverb of volitionality⁵.

Property: [+Vol +Ani]

Class 2. Verbs that take volitional animate subject but are not compatible with adverb of volitionality.

Property: [-Vol +Ani]

Class 3. Verbs that take non-volitional subject.

Property: [-Vol -Ani]

V. DISTRIBUTION OF INTRANSITIVE VERBS

The motivation for tripartite classification of verbs needs to bring into focus with its proper merit. We will find out in this section that the neat binary classification of intransitive verbs gets totally upset especially due to verbs of Class1 and Class2. We propose in this paper to attribute a Unergativity/Unaccusativity score to each verb and plot the verb accordingly in a diffusion chart as described in this section. This amounts to predict the distribution of intransitive verbs on a scale of unaccusative-unergative.

TABLE II. Ranks of the diagnostics

Tests	Scores	Max	Normalised	Rank
Ergative case	210	268	0.78	3
Cognate objects	158	268	0.59	5
Impersonal Passive	221	268	0.82	2
Past Participial	132	268	0.49	6
Inabilitaives	97	268	0.36	7
Compound verb	0	268	0.86	1
Overt Genitive	209	268	0.78	4

1) *Calculating Scores for each Verb*: The Unergativity/Unaccusativity score gives the position of an intransitive verb on the unaccusative-unergative scale. The normalised diagnostic scores (see column 3, table 2) are used for the calculation of the score of the verb. Unaccusatives have been assigned with negative whereas unergatives with positive scores for convenience so that the verbs are well distributed. Following steps have been followed for scoring a verb:

- (i) Each verb gets $-1 * (\text{normalized diagnostic score})$ score for each diagnostic if the output of the test for that verb matches the expected output of the unaccusative. Similarly, it gets $+1 * (\text{normalized}$

⁵ Some examples of adverb of volitionality are *jaanbujhkar* 'deliberately', *apne marji se* 'intentionally' etc.

diagnostic score) score if the output matches the expected output of unergative.

- (ii) The final score for a verb is calculated by summing up the score given for a verb by each diagnostic. A more positive or negative score of a verb shows a more unergative or unaccusative nature of the verb respectively.

Table III. Scores for the Verbs

Verb	English Gloss	Ergative Case	Cognate Object	Impersonal Passive	Part Participial	Inabilitatives	Compound Verb	Overt Genetive	Verb Score
नाच	Dance	0.78	0.59	0.82	0.49	0.36	0.86	0.78	4.68
थक	tired	-0.78	-0.59	-0.82	-0.49	0.36	-0.86	0.78	-2.4
फैल	spread	-0.78	-0.59	-0.82	-0.49	-0.36	-0.86	-0.78	-4.68

A small snapshot from the experiment is presented in Table 3. We see that the high positive score of in verb *naac* ‘dance’ shows it’s highly Unergative nature as it behaves like a unergative for every diagnostic test. Similarly, verb *phail* ‘spread’ behaves like an Unaccusative for all the diagnostics and achieves a high negative score accounting for its highly Unaccusative nature. The verb *thak* ‘get tired’ on the other hand passes few tests of unaccusativity but fails in a few and therefore falls in the region between the unaccusatives and unergatives.

2) *Diffusion Chart*: The scores calculated for each verb in the previous section are plotted in a Diffusion Chart. Each diagnostic test is actually a syntactic operation that a verb can participate in. Therefore the plotting of a verb also signifies the syntactic behavior of that verb.

The x-coordinate in the graph represents the total score that a given verb acquires for all tests (see table 3). The y-coordinate indicates the number of verbs that achieve same score which in turn signifies the similarity of their syntactic behavior. For example, the (-4.68, 32) co-ordinate represents 32 verbs all of which pass all the diagnostics of unaccusative. The list contains verbs *bana* ‘get built’, *phail* ‘spread’ etc. On the other hand verbs of coordinate (4.68, 9) indicates a concentration of verbs of high unergative features such as *kood* ‘jump’, *naac* ‘dance’, *hans* ‘laugh’ etc. The x-coordinate of verbs such as *tair* ‘swim’, *chillaa* ‘shout’ etc. (3.5, 8) is very close to that of ‘high unergative feature verbs’ which also indicates the similarity in their syntactic behavior.

The plotting in the diffusion chart also focuses on one important factor namely binary classification of intransitive verbs do not really hold for Hindi. Instead, we prefer to call the verbs ‘more unergative’ and ‘more unaccusative’ depending on their position in the graph.

We observe a diffusive distribution of the intransitives at the marked area of the graph. However count of those verbs is low. Verbs that belong to this area show their response to diagnostic tests do not match to the expected outcome. This observation motivates us to bring in a tripartite classification of verbs as shown in Figure (1b).

VI. REVISION OF DIFFUSION CHART

The predictions from this model are compatible with Bhatt (2003), Sorace (2000) and Kaufmann (1995) who have proposed to abandon a strict two-way distinction between unaccusatives and unergatives. But the tripartite classification of the verbs yields a better result since it has less intra-class diffusion among the verbs. As it could be

seen from Figure (1b), the diffusion among the verbs of Class2 and Class3 (resembling unaccusatives taking animate and inanimate subjects respectively) is very less

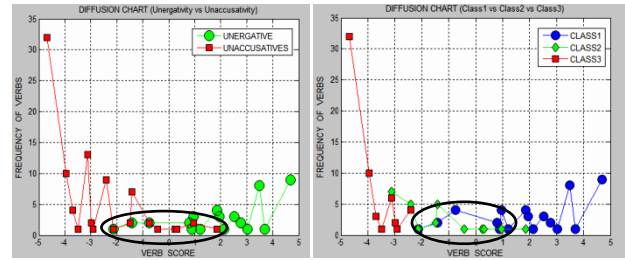


Figure 1. a) Diffusion Chart with two classes
b) Revised Diffusion Chart with three classes

and Class3 is very well separable with low intra-class diffusion. Therefore, a three-way classification approach gives better results than a two-way classification as it is able to separate the Class3 verbs and the complete class of Unaccusatives do not have to suffer because of the diffusion in the class of Unaccusative animates with unergatives. Also, the Unaccusatives having high intra-class diffusion are divided into two classes with less intra-class diffusion.

VII. CONCLUSION AND FUTURE DIRECTION

This paper introduces a more holistic approach towards the classification of intransitive verbs. The approach uses the combined results of all the diagnostic to capture the syntactic behavior of a verb and proposes a computational approach for the unaccusative/unergative distinction in Hindi. The work confirms that a split between unaccusative and unergative verbs indeed exists in Hindi but that the division is not clear-cut.

There are several dimensions in which the work can be improved. First, we can classify the intransitive verbs into semantic classes like motion verbs, stative verbs etc. and observe their syntactic behavior and try to capture the similarity among the classes. Secondly, we can use learning methods like neural networks to learn the scores for the diagnostics to get a better performance in the classification process.

REFERENCES

- [1] Acartürk, C. & Zeyrek, D. 2010. “Unaccusative/Unergative Distinction in Turkish: A Connectionist Approach”, the 23rd International Conference on Computational Linguistics. Proceedings of the 8th Workshop on Asian Language Resources. Beijing, China, 2010. pp. 111-119.
- [2] Ahmed, T. “The Unaccusativity/Unergativity Distinction in Urdu”. Journal of South Asian Linguistics, North America, 2010.
- [3] Bhatt, Rajesh. 2003. Causativization, “Topics in the Syntax of the Modern Indo-Aryan Languages”, handout.
- [4] Kaufmann, Ingrid. 1995. “O- and D-Predicates: A Semantic Approach to the Unaccusative-Unergative Distinction”, Journal of Semantics 12, 377-427.
- [5] Levin, Beth and Hovav, Malka Rappaport. 1995. “Unaccusativity: At the Syntax-Semantics Interface”. Cambridge, MA: MIT Press.
- [6] Perlmutter, D. 1978. “Impersonal passives and the unaccusative hypothesis”. Proceedings of the 4th Berkeley Linguistics Society, 157-189.
- [7] Sorace, Antonella. 2000. “Gradients in auxiliary selection with intransitive verbs”, Language 76, 859–890.
- [8] Van Valin, Robert D. Jr. 1990. “Semantic parameters of split intransitivity”. Language, 66:221-60