Cutting and Breaking Events in Telugu

Vasanta Duggirala, and Y. Viswanatha Naidu

Abstract—This paper makes a contribution to the on-going debate on conceptualization and lexicalization of cutting and breaking (C&B) verbs by discussing data from Telugu, a language of India belonging to the Dravidian family. Five Telugu native speakers' verbalizations of agentive actions depicted in 43 short video-clips were analyzed. It was noted that verbalization of C&B events in Telugu requires formal units such as simple lexical verbs, explicator compound verbs, and other complex verb forms. The properties of the objects involved, the kind of instruments used, and the manner of action had differential influence on the lexicalization patterns. Further, it was noted that all the complex verb forms encode 'result' and 'cause' sub-events in that order. Due to the polysemy associated with some of the verb forms, our data does not support the straightforward bipartition of this semantic domain.

Keywords—Cluster analysis, Cutting and breaking events, Polysemy, Semantic extension, Telugu.

I. INTRODUCTION AND BACKGROUND

The continuous stream of activity we witness in everyday world is parsed into specific events containing particular objects. Like objects, events have beginnings, the middles and ends, but unlike objects, events are structured in time. In an insightful discussion of the role of bodies in the categorization of events, Tversky, Morrison and Zacks [1] commented that bodies, events and objects are intimately connected and intertwined with categories of scenes that are the settings for events. Their main argument is that cognition of central categories is not just embodied, but is also embedded. In recent years the question of how people decide on category membership and draw inferences about members of a particular category has attracted the attention of researchers from different disciplines including psychology, linguistics and cognitive sciences. Much of the already published cross-linguistic data on this topic has focused on the cognitive variables that contribute to event conceptualization and expression and the way events are structured.

Theoretical linguistic studies dealing with verb meanings have paid attention to the semantic fields, semantic contrasts, syntactic properties and the way all these map onto each other. They typically deal with intensional aspects of the verbs encoding events and sub-events. In English, the cut and break verbs (C&B verbs henceforth) fall into two broad classes: one, the break-class, the members of which participate in causative-inchoative alternation (e.g. John broke the vase vs. the vase broke) and two, the cut-class verbs, the members of which do not participate in causative-inchoative alternation (e.g. John cut the bread; * the bread cut). C&B verbs however reportedly participate in middle constructions as shown in the English examples below:

The vase broke easily.
The bread cut easily.

The inchoative constructions unlike the middles are thought to refer to events under specific time reference, and do not introduce cause or an agent into the discourse representation. Bohnemeyer [2] also pointed out that the break-verbs do not always specify the instrument involved whereas the cut-verbs do; that the causative forms of break verbs constitute the basic forms in English, whereas, the inchoative forms are derived by argument structure operations in general, and two specific principles connected to the argument structure, viz., morpholexical transparency and complete linking. Bohnemeyer concluded that not all languages have a binary distinction between cut and break verbs; in some languages, complex predicates that are semantically specific on both properties of an instrument and state change inflicted on the theme object serve as bipolar verbs (neither cut-class nor break-class) and that other languages have severance, that is break type, but no cut-type verbs.

Bohnemeyer, Bowerman and Brown [3] at the Max Planck Institute for Psycholinguistics at Nijmegen developed an elicitation tool and field guide for collecting data on C&B verbs using a set of 61 short video-clips. Using this tool, 24 researchers around the world collected data from 28 genetically, typologically and geographically unrelated languages including two languages spoken in India – Tamil and Hindi. The rationale for this research being that if categories associated with everyday words (in this case cutting and breaking verbs) are largely universal, then they should be shaped by non-linguistic cognition guided by perceptual and cognitive predispositions. On the other hand, if lexical categories are shaped by language/ culture specific constraints, then there should be considerable variability in the use of C&B verbs across languages in describing the same visual scenes contained in everyday events. The cross-linguistic study involving 28 languages sought to look at the verb types, tokens, meaning extensions, and on how state change is encoded (in mono-morphemic verbs or in compound verbs) by having a small group of participants from each language describe what is happening in the video-clip, and by analyzing each predicate produced. A special issue of Cognitive Linguistics (Volume 18, 2007) contains details of the results from some of the languages studied as part of this project. Some salient observations based on this project are

Vasanta Duggirala is Professor of Linguistics at the Department of Linguistics, Osmania University, Hyderabad, Andhra Pradesh, India Pin code 500007 (e-mail: vasantad@gmail.com).

Y. Viswanatha Naidu, Senior Research Fellow, is with the Department of Linguistics, Osmania University, Hyderabad, Andhra Pradesh, India Pin code 500007 (e-mail: viswanath.iiit@gmail.com).
summarized below:

Brown [4] reported that Tzeltal, a free-argument ellipsis language spoken in parts of Mexico had more than 50 distinct C&B verb roots that were differentiated depending on the theme object properties (long, thin, hard, flexible, round, soft, brittle); spatial properties (across or along the axis or part-whole relations); type of result (action completed or incomplete); manner of action and type of instrument involved. The other extreme case was Yeli Dnye, the Papuan language, which according to Levinson [5] required only three different C&B transitive verbs to describe the caused events of material separation. In this language, the type of action or the manner (force) of action made little difference.

Majid, Bowerman, Van Staden and Boster [6] commented that the events of ‘opening’ were encoded most distinctly across all the 28 languages, that C&B verbs did hang together distinctly from the other verbs of separation such as tearing, peeling and taking apart and that the control over the locus of separation distinguished cut verbs from break verbs. Narasimhan’s [7] data based on 3 Hindi speakers and 2 Tamil speakers revealed that there were 50 tokens of kāṭ ‘cut’ in Hindi, highest number compared to the other verb forms used. The comparable verb in Tamil was veḻu ‘cut’ which occurred 34 times. The verb ‘break’ was rendered as tōḍ in Hindi with a frequency of 42 times and as oḍi in Tamil with a frequency count of 20. Narasimhan stated that it was the degree of control an agent has over the locus of separation in the material integrity of the object that determined the C&B verb choice. The cut-verbs had higher control compared to the break-verbs that had less precise degree of control and the tear-verbs were applied primarily to the 2-D flexible objects such as paper and cloth. Speakers from both Tamil and Hindi reportedly grouped snapping and smashing events by labeling them with the same verb.

A closer look at the categorization of C&B event data collected from four genetically related languages, viz., English, German, Dutch and Swedish by Majid, Gullberg, Van Staden and Bowerman [8] revealed that English verb ‘break’ picks out a super-ordinate category that subsumes finer distinctions encoded through ‘snap’ and ‘smash’ verbs and that ‘break’ is used to describe destruction of a wide variety of objects such as sticks, rope, plates, yarn etc. In English, chopping goes with cutting events involving knives and scissors that are relatively high in predictability with respect to the locus of separation. In contrast, in the other three languages, chopping events go with breaking events such as smashing that are relatively low in predictability in relation to the locus of separation.

A recently published study based on Aiwoo by Naess [9] an oceanic language spoken in Reef islands in the Pacific Ocean sheds more light on the complexity of lexicalization patterns in the domain of cutting and breaking that poses certain challenges to the clear-cut dichotomy reported by the other researchers in this field. Specifically, Naess noted that the acts of cutting and breaking are systematically encoded in Aiwoo not through single lexical verbs, but through use of bimorphemic expressions specifying both the causing event and its effect on the object. The author also reported that neither of the C&B morphemes could be used independently outside the semantic domain of cutting and breaking in Aiwoo. It was reported that the strategy of simply leaving the effect (on the object) unspecified is unavailable in Aiwoo and that cause-effect integration in this language results in a cline of event integration.

In an increasingly multilingual world we have today, comprehension and expression of C&B verbs might be subject to the influence of the language use histories and language practices of speakers. The authors of this paper have been working with a larger group on a project dealing with multilingualism in the context of users of Telugu-Hindi-Urdu-English in the city of Hyderabad, India. The present study could be considered as an offshoot of that larger investigation in which language use history and use patterns were examined using a specially developed questionnaire. Using a shorter web-version of that questionnaire [available on http://valrg.in/], we have now identified five speakers of Telugu who while being exposed to and familiar with Hindi, Urdu and English languages, provided relatively high ratings (4 or 5 on a 5-point rating scale) for Telugu (their native language) in proficiency and language choice sections of the questionnaire. Their descriptions of the original 61 videos (developed at the MPI, Netherlands) containing cutting breaking, opening and peeling events provided the database for our investigation. Even in Telugu, events involving opening and peeling in Telugu (teracu/fiyu’open’, olacu’peel’) were verbalized with little, if any variation. All the participants used the same form to describe events involving these actions.

In this paper, we discuss results obtained from 43 videos depicting animate agents causing separation / destruction in the material integrity of different objects such as carrot, stick, rope, cloth, watermelon, pot, plate, branch of a tree using instruments such as knife, machete, hammer, axe, scissors and hand.

II. THE PRESENT STUDY

Telugu is one of the major languages of India belonging to the Dravidian family, the other three languages of this family being, Kannada, Tamil and Malayalam. All these four languages are spoken primarily in the southern part of India. Telugu is one of the state languages and spoken by a majority of people living in the state of Andhra Pradesh. The canonical word order of Telugu is SOV. Every Telugu verb has both finite and non-finite forms. Finite verbs carry tense/aspect and person, number, and gender (PNG) suffixes in agreement with the noun/pronoun used as the subject. For a detailed discussion of Telugu verb paradigms see Krishnamutti and Gwynn [10].
III. PARTICIPANTS

Of the five participants, one was male and four were females. Their ages ranged from 20-31 years (average: 26.6 years). Two of them had under graduate level education while the other three had post graduate degrees. Except for one, all the others had their early education in Telugu medium. All the participants spoke Telugu at home and were fluent readers and writers of the language. They rated their ability to read, write, speak and understand Telugu with ‘Good’ or ‘Excellent’.

IV. DATA COLLECTION PROCEDURE

Each participant was seated in a comfortable chair in front of a laptop in a quiet room. They watched the video-clips presented in a fixed order and described what was happening / or what the agent was doing in each event. A short break of 10 minutes was given after they have watched 30 videos. The participants were paid for their participation in this study. Their verbalizations were tape-recorded using a digital audio-recorder (Sony ICD-MP3) and transcribed later for analysis. The 215 (43 events X 5 speakers) verbalizations were examined for number of unique verbs used, types of verbs, and cognitive linguistic variables underlying the lexicalizations. When they used two different verb forms for describing an event occasionally, only the first form was taken as a response.

V. RESULTS

Some of the Telugu verb forms generated by the 5 speakers for the 43 videos of caused separation/destruction are listed in Table I.

It should be noted that the verbs listed in Table I are not exhaustive of all C&B verbs in all dialects of Telugu in that the 43 video-clips did not represent all possible C&B events in this culture. In day to day life, people engage in many events involving separation/destruction of ordinary objects such as paddy, dhaaL, berries etc. Verbalizations of such events require lexicalizations. When they used two different verb forms for describing an event occasionally, only the first form was taken as a response.

VI. THE FORMAL PROPERTIES

All the verb forms noted in this study include simple verbs, explicator compound verbs, and complex predicates; the frequency count for each is shown in Table II:

Simple verbs: Consisted of verb root + tense /aspect + PNG markers accounted for over 30% of the lexicalizations used to encode the action of objects being cut or broken by an agent. For instance, in response to the video where a woman cuts a small branch of a tree (a twig) with a knife, a typical response was:

(1) āme komma-nu katti-to kosindi

'She cut the twig with knife'

This category also includes forms such as kosēsindi that has aspectual information in addition to gender marker.
Explicator Compound verbs: A small proportion of the lexicalizations (~10%) could be classified as Explicator Compound Verbs (ECVs), a term used by Abbi [11] in the context of discussion of how manner is encoded in a number of languages spoken in South Asia. ECVs are formed by a non-finite form of a verb followed by a second delexicalized explicator verb. To illustrate with an example:

(2) oka vyakti tāḍu-ni goḍqali-to nariki-vesāṭu
one man rope-acc axe-with break(non)-put.past.png
He broke the rope with an Axe.

Note that the form occupying the second place (V2), vēśāṭu is the explicator that contributes to the overall semantics of V1, nariku ‘break’. The participants had the choice of lexicalizing it using a simple verb, narikāṭu, carrying similar meaning as narikiwēśāṭu, but some of them chose to use ECVs, perhaps to indicate completion of the action. Together these two types of lexicalizations seem to draw participants’ attention more to the cause of the event than the effect on the theme object.

Complex Predicates (CPs): These include nominal expressions of separation and/or adverbs preceding the main verb such as:

renḍu mukkalugā cēyu ‘make/do it into two pieces’,
sagāniki kōyu ‘cut into half’, mad’yaki teggoṭtu ‘break it in the middle’, renḍu bāgālūgā cēyu ‘make/do it into two portions’, renḍugā viraggoṭtu ‘break it into two’ etc.

As mentioned earlier, these expressions reflect the fact that participants’ attention is drawn to the end result, that is, change in state in the object on which C&B operation was performed. An important aspect of these expressions is that in every case the result sub-event is followed by cause sub-event in that order. In a majority of these expressions, cēyu ‘make/do’, a general purpose verb is used to express the result of the action. This point is elaborated in the next section.

Abbi [11] has pointed out that South Asian languages make use of two varieties of complex predicates, viz., ECVs and of the action. This point is elaborated in the next section.

VII. SEMANTIC PROPERTIES OF C&B VERBS IN TELUGU

The distribution of Telugu C&B verbs could be schematically depicted as shown in Fig. 1:

As evident in the figure above, Telugu has verbs that are specific to cutting events (62 out of 215 lexicalization) breaking events (72 out of 215), and tearing events (12 out of 215). In addition, it also has verbs that are used for expressing cutting, breaking and tearing events. Of the 215 events, 69 verbalizations involved complex verb tokens that are used to express both C&B and tearing events. The equal distribution
of exclusive C&B tokens (62&72 respectively) and common tokens (69) suggest that Telugu does not show a clear-cut dichotomy of C&B semantic domain. To elaborate we would like to cite the data where cutting, breaking & tearing events were described by different users using both general expressions (e.g. mukkalu cēyu ‘make/do into pieces) as well as specific C&B and tearing verbs (kōyu ‘cut’, naraku ‘break’ and cimpu ‘tear’ respectively) while viewing the following specific events.

- A man cut the carrot into two pieces with a knife.
- A man chop a branch repeatedly with an axe.
- Tear cloth into two pieces by hand.

(5a) atanu katti-to kārī rēnu mukkalu cēsālu
he knife-with carrot-acc two pieces did.png

‘He did (cut) the carrot into two pieces with a knife’

(5b) atanu katti-to kārī rēnu mukkalugā kōsālu
he knife-with carrot-acc two pieces cut.past.png

‘He cut the carrot into two pieces with a knife’

(6a) oka vukti karru-nu goṣdali-to rēnu mukkalugā cēsālu
one person stick-acc axe with two pieces did.png

‘He did (cut) the carrot into two pieces with an axe’

(6b) atanu goṣdali-to komma-nu narikālu
he axe-with branch-acc broke.png

‘He broke the branch of a tree with an axe’

(7a) okammāyī bātta-ni rēnu mukkalugā cēsindi
one woman cloth-acc two pieces did.png

‘One woman did (tore) the cloth into two pieces’

(7b) āme bāttu-nu cimpinidi
she cloth-acc tore.png

‘He cut the carrot into two pieces’

This data illustrates that Telugu speakers choose to use either general expressions as in examples 5a, 6a, and 7a or specific verbs as in 5b, 6b, and 7b. In other words, the polysyym associated with mukkalu cēyu ‘make/do into pieces’ are more productive in Telugu and it occurs in varied contexts as evident in a Telugu corpus. It was used in expressions used to describe actions of stealing, making arrangements for some event, filing a complaint, demanding, cooking, disturbing arrangement of things, assassinating and so on. Thus, it appears that one of the semantic functions of cēyu ‘make/do’ verb within C&B domain is to integrate the action and is effect on the object and help agents focus on the result component of action. In the context of a discussion on cognitive relativism Tai [13] commented that ‘result is a semantic prime in Chinese verb semantics and that the action-result schema has played a much more important role in Chinese than in English. Our observation that result precedes action in C&B verbalizations highlights the importance of result sub-event that requires further investigation.

VIII. CLUSTER ANALYSIS

We have used the cluster analysis technique described in Majid et al [8] with 43 clips as rows and 20 verbs as columns to generate a dendrogram (see Fig. 2). The clusters in the figure capture the distribution of main verb forms used across the entire stimulus set. Video clips that are most similar to one another are described by the same verb and hence have the shortest leaves or lines linking them. Notice that on the left side top part of the dendrogram we have videos 2 and 13 and 23 clustering together with the verb tempu, a break-verb involving an instrument. At the bottom of the dendrogram, we have the event, cutting a rope with scissors (video 24) is clustering with chopping a cloth with hand (video 34), both in turn are connected to the cluster above them with videos 35 and 38 that deal with breaking a piece of wool into one or more pieces. It is use of complex expression mukkalugā cēyu / tempu ‘make / snap / chop into pieces’ that is determining the similarity across these events. The verb teggoṭṭu was used to describe an action in which the agent has less control over the result compared to the events involving a knife or other bladed instruments that are described using cut-type verbs such as kōyu. The outermost leaf of the dendrogram has the verb pagalagottu ‘break’ that applies only to rigid objects such as a
plate or a pot (in videos 39 and 40) and not to any other object in any other video. Similarly, videos 1 and 36 depicting events of tearing were described by the verb *cimpu* or *mukkalu cēyu* ‘make/do into pieces’ by all the participants. The fact that cutting and chopping events occupying both the upper and the lower middle portion of the dendrogram demonstrates the polysemous nature of the complex verb forms (as evident in Figure-1). The resultant variation in the number and types of verbal descriptions is evident in the size of the leaves in the Telugu dendogram that is considerably larger compared to the leaves in the English dendrogram published in Majid et al [8].

IX. OBSERVATIONS ON THE EXTENSIONS OF THE C&B VERBS

- The most general verb form associated with the C&B event verbalization is *cēyu* ‘make/ do’ as a complex expressions that occurred with both cutting and breaking actions. With respect to non-canonical situations represented in video clip numbers. 13, 21, 23, 37, 43, 50, and 54, there was a tendency to use *cēyu* ‘make/do’ form along with *kāyu, virucu* etc. resulting in considerable variation.
- The verb *kāyu* ‘cut’ is more general, typically used to describe actions on objects such as carrot, orange, rope and fish that do not require much force or a heavy instrument. However, it was also extended for describing some breaking events.
- Break verbs such as *naruku, virucu, viraggotu, teggotu*, *pagalagotu* were used exclusively for describing events involving hand or a heavy instrument and the action is accomplished with force on long rigid objects. They hardly ever replaced the cut-verbs.
- Tear-verbs had the least range of extension in that they were applied only to describe separation in the material integrity of 2D objects such as cloth. However, when the event is non-canonical, for example a cloth being cut with a stick or a hammer, other types of cut and break verbs were used in addition to general purpose make/do expressions.
- Since hair is always cut with a scissors, this event elicited the verb *kattirincu* that is specific to the scissors.
- When the result was not a complete destruction, but a mere separation, verbs such as *vidādyu /vidagotu* were used.

X. FINAL COMMENTS

On the whole, our results are in agreement with those reported by earlier researchers with respect to the fact that categorization of the semantic domain of C&B events is determined by factors such as (1) predictability of the locus of separation; highly predictable with cut-type verbs compared to break-type verbs (2) Separation brought about by a sharp bladed instrument are distinguished from other ways of separation (3) unlike Tamil and Hindi results reported by Narasimhan [7], Telugu speakers made a distinction between smashing and snapping by using different verbs to describe the events in which rigid objects such as pot and plate (using *pagalagottu* as opposed to long thin objects such as rope, carrot or stick were involved (using *tempu / teggotu*) (4) there is a need for an in-depth investigation to understand the semantics of a variety of complex verb forms noted in this study that went simultaneously with both cut-type and break-type expressions. Finally, the variation in the verbalization of the 43 agentive actions of cutting and breaking could be attributed to participants’ bodily experience of having engaged in those actions in addition to their knowledge of the action verbs. What the participants seem to be apprehending while watching the video-clips is the particular situations embedded in the scenes that are in turn influenced by the culture in which they are living. Conclusive statements about cutting, breaking and tearing verbs in Telugu awaits further research that pays greater attention to issues connected to type of events and a variety of objects represented in the videos, dialects associated with the language under investigation, and number and age range of participants.

**Abbreviations**

| acc | : | accusative case |
| non | : | non-finite (or participle) form |
| past | : | past tense |
| PNG | : | Person, Number, Gender |
| SOV | : | subject, object, verb |

**ACKNOWLEDGMENT**

The authors would like to thank the Department of Science and Technology (DST), Government of India for supporting a larger investigation of which this study is only a part. Thanks are due to the researchers at the Max Planck Institute for Psycholinguistics, Netherlands for sharing the video-clips that made this study possible. The authors are grateful to all the participants who took part in this study and to S. Bapi Raju and Jigar Patel, fellow members on the DST funded project for helping us generate the dendrogram. An earlier version of this paper was presented at the 40th All India Conference of Dravidian Linguists held at the University of Hyderabad during June 18-20, 2012.
REFERENCES


Vasanta Duggirala received Master of Science (Speech and Hearing) and a Doctoral degree in Audiology from University of Mysore, India, and University of Memphis, USA respectively. She is currently a professor at the Department of Linguistics, Osmania University, Hyderabad, India. She has been teaching, conducting research, and publishing on topics related to communication disorders, special education, and different branches of applied linguistics including cognitive linguistics. She has co-edited two books: Language Development and Language Disorders: Perspectives from Indian Languages published by Bahri Publications, New Delhi (1994), Practice and Research in Literacy published by Sage, New Delhi (2002) and served as a guest editor for the special issue on ‘clinical linguistics’ of the Indian Journal of Applied Linguistics, Vol. 36, (2010). Since 2009, she has been involved in an interdisciplinary national level project on multilingualism in India funded by the Department of Science and Technology (DST), Government of India under their cognitive science initiative during the XI plan.

Y. Viswanatha Naidu has Bachelors degree in Sanskrit, Masters degree in Linguistics from Osmania University, and M.Phil degree in Computational Linguistics from the International Institute of Information Technology, Hyderabad, India. He is currently a senior research fellow on the DST funded research project on multilingualism in India under their Cognitive Science Research Initiative. His articles in the field of linguistics have appeared in journals such as Indian Linguistics, International Journal of Mind, Brain and Cognition and Lecture Notes in Computer Science. His research interests are cognitive and computational linguistics.