Reduplication and Distributed Meaning in Indian Languages

Aadithya N Kopparapu &
Aparna Sankar
EFLU, Hyderabad
aadikopparapu@gmail.com
aparnasiyer17@gmail.com

Abstract

Reduplication is a widespread and semantically rich phenomenon in Indian languages, often giving rise to meanings such as distribution, iteration, and intensification. While individual studies have documented these effects in specific languages, there remains a lack of a unified, cross-linguistic semantic framework that can account for the interaction between reduplication and grammatical category. This paper proposes a new compositional model—the R-D model—to analyse reduplication in Indian languages through the interaction of three components: R (Reduplicator), a morphosyntactic trigger that licenses reduplication; I (Iteration), a semantic operation that introduces multiplicity of events or referents; and D (Distribution), a semantic operation that spreads these instances across space or time.

Crucially, the availability of I and D is constrained by the ontological features of the reduplicated base, which are defined in terms of the binary features [±time] and [±space]. Nouns ([+time, +space]) allow both iteration and distribution, verbs ([+time, –space]) allow iteration only, prepositions ([–time, +space]) allow distribution only, and modifiers ([–time, –space]) block both, resulting in mere intensification. Data was elicited directly from native speakers across Indo-Aryan, Dravidian, Munda, and Tibeto-Burman languages using translation tasks, context-based elicitation, and interpretation checks.

The R-D model accounts for observed patterns of reduplication across categories and languages while explaining the absence or blocking of certain meanings. It also predicts inter-language variation based on the distribution of ontological features and category-specific constraints. This framework not only captures the semantic diversity of reduplication in Indian languages but also offers a formal, compositional account of how grammatical category and semantic interpretation interact. The study contributes to the typology of reduplication and opens new directions for modelling semantic operators in morphology.

Keywords: Reduplication, Distribution, Iteration, Intensification, R-D model

1. Introduction

Reduplication is a widely attested linguistic process across the world's languages, serving a range of functions from morphological derivation to pragmatic emphasis. In South Asian languages, reduplication is not only pervasive but also semantically rich, yielding effects such as intensification, distribution, superordination etc. These effects are often sensitive to the syntactic category of the reduplicated element: when nouns are reduplicated, the resulting interpretation typically involves referential distribution; reduplication of verbs often leads to iterative or extended duration readings; and adjectives or adverbs yield intensification etc. Despite this well-documented diversity, there has been little work offering a unified semantic account of how reduplication systematically interacts with grammatical category and meaning in Indian languages.

This paper seeks to fill that gap by proposing a new compositional model of reduplication-induced meaning: the R-D model, which stands for Reduplicator-Distributor. In this account, reduplication is treated not as a single semantic operation but as a triad of operators that interact with linguistic units based on their syntactic category.

This model is inspired by previous proposals such as Balusu (2006), who analysed distributive readings of Telugu numeral reduplication using a spatial/temporal operator. While Balusu's work insightfully captures the distributive meaning of reduplicated numerals, it does not extend to other categories or other semantic effects such as iteration. The R-D model expands on this idea by proposing a general framework that can account for a broader range of categories, languages, and meanings.

The empirical focus of the paper is on reduplication in Indian languages across major language families, including Indo-Aryan (e.g., Hindi, Marathi), Dravidian (e.g., Telugu, Malayalam), Munda (e.g., Ho, Santali), and Tibeto-Burman (e.g., Ao, Ladakhi). These languages exhibit a remarkable variety of reduplication patterns, both in terms of form and meaning, yet they all show systematic interactions between category and interpretation. Our data shows that across languages, the same semantic operators are active—but their availability is determined by the ontological status of the reduplicated item. This makes a strong case for a compositional semantic model grounded in category-sensitive operator licensing.

In section 2, we begin by laying out the aims, objectives, and research questions of this study. In section 3, we survey the existing literature on reduplication in typology and South Asian languages, as well as prior formal semantic accounts. Section 4 introduces the R-D model in detail, including its formal structure and theoretical motivations. Section 5 presents the cross-linguistic data and applies the model to a range of reduplication constructions across languages and categories. Finally, section 6 offers conclusions and directions for further research.

2. Research goals and methodology

2.1 Aims and Objectives

The primary aim of this paper is to unify the explanation for the reduplicative meanings of iteration, intensification and, distribution by positing a set of operators named R-D.

The objectives of this study are:

- To formulate a semantic model (R-D) in which reduplicative meanings involve a set of operators and their interactions with ontological features of the lexical entries.
- To demonstrate how the distributive meaning of reduplication is based on the grammatical category and ontological properties (e.g., temporal vs spatial) of the reduplicated element and .
- To compare and contrast data from multiple Indian languages, spanning diverse language families and show that seemingly different semantic effects of reduplication across languages can be accounted for using the same operator-based framework.

By pursuing these goals, the study seeks to contribute not only to the semantic analysis of reduplication, but also to broader typological and areal discussions on how morphology interacts with meaning across languages.

2.2 Research Questions

The central questions guiding this study are as follows:

1. What are the basic semantic operations activated by reduplication in Indian languages?
2. How do these operations differ according to the grammatical category (e.g., noun, verb, adjective) of the reduplicated item?
3. Can a model for a unified explanation be proposed and if yes, can the model account for the full range of semantic interpretations associated with reduplication across different languages and categories?
4. What exactly is shared by languages in India as a Linguistic Area for this phenomenon?

These questions are designed to evaluate both the empirical coverage and the theoretical adequacy of the model.

2.3 Research Methodology

The empirical foundation of this study is built on originally elicited data from native speakers of several Indian languages, representing the Indo-Aryan, Dravidian, Munda, and Tibeto-Burman families. The goal of the data collection process was to systematically test how reduplication behaves across different syntactic categories—namely nominals, verbs, modifiers, and adpositions—and to determine which semantic effects (distribution, iteration, intensification) are licensed in each case.

Data was elicited through direct interaction with native speakers, using a semi-structured elicitation protocol designed to control for both form and context. The elicitation process involved the following steps:

1. Translation Tasks: Speakers were presented with sentences in English (or Hindi in some cases) and asked to translate them into their native language. These source sentences included contexts designed to test distributive, iterative, and emphatic meanings. For example, speakers were asked to translate sentences like "Every child went to their own house" or "He walked and walked until he got tired."
2. Contextual Prompting: In addition to translations, speakers were given specific contexts or situational descriptions and asked to form sentences that would naturally fit those contexts in their language. This allowed the observation of how reduplication emerges (or fails to emerge) spontaneously in discourse-relevant settings.
3. Category-wise Testing: These tasks were systematically repeated across four grammatical categories: nominals, verbs, modifiers, and adpositions. For each category, it was noted whether reduplication naturally occurred or whether the language lacked reduplicative constructions in that category. For example, if a language did not allow reduplication of adpositions, this was explicitly recorded.

After eliciting the sentences, a second round of inquiry was conducted to determine the interpretative range of the reduplicated forms. Speakers were provided with a set of possible meanings or interpretations commonly associated with reduplication (e.g., "Does this sentence mean the action happened again and again?", "Does this mean each person did it separately?", "Does this just mean 'very'?"). They were then asked whether each interpretation was available or acceptable for the sentence they had produced. This helped distinguish between what is grammatically acceptable and what is semantically accessible to native speakers, allowing for a more precise mapping of form to meaning.

The elicitation strategy thus combined structured translation with open-ended production and interpretation validation, ensuring that the data was both naturalistic and theoretically targeted. This approach also made it possible to track inter-speaker and inter-language variation, especially in cases where reduplication yielded ambiguous or non-compositional interpretations.

The resulting dataset forms the empirical basis for evaluating the R-D model developed in this study, allowing us to analyse which semantic operators are active or blocked in specific syntactic and ontological environments.

The theoretical framework presented in this study is based on the ontological features of the syntactic categories and how reduplication effects different categories. The analysis of the data collected is done to verify if the correlation between syntactic category and semantic effect is as predicted by the proposed theoretical framework.

3. Literature review

The following is the literature review of the papers that inspired this study.

In Abbi (1985) and Abbi (1990), Anvita Abbi shows how languages belonging to diverse language families in India, namely Indo-Aryan, Dravidian, Tibeto-Burman and Austroasiatic, show reduplicative phenomenon with shared interpretations. She claims that this similarity is due to areal influence in the South Asian Linguistic Area.

Balusu (2006) presents a detailed analysis of distributive reduplication in Telugu, arguing that reduplicated elements introduce event plurality. He utilises event semantics to explain the distributive meaning of Telugu numerals by positing a D-operator which has "spatial, temporal and participant key readings."

Expanding on this, Balusu and Jayaseelan (2013) compare Telugu data with Malayalam and Tamil, emphasizing that the phenomenon is pan-Dravidian but not uniform in form or function. They maintain that reduplicated quantifiers in these languages uniformly trigger a distributive interpretation, albeit with some morphosyntactic differences. One significant theoretical insight here is the claim that "reduplication serves as a strategy for quantificational strength in languages that otherwise lack overt distributive markers like 'each' or 'every.' ." This means that in languages which lack overt universal quantifiers, reduplication fills that role with the distributive meaning.

Together, these works draw a compelling picture: reduplication, far from being a mere stylistic or phonological quirk, is a rich grammatical strategy that various languages deploy to encode distributive semantics. Telugu and its Dravidian relatives offer particularly lucid cases, where reduplication interacts with syntactic scope, event structure, and quantificational force. These findings invite further inquiry into whether such constructions are universally available but parametrically varied in expression—or whether they reveal something deeper about how languages encode event individuation.

4. The r-d model: a compositional account of reduplication and meaning

4.1 Ontological Features of Syntactic Categories

The central innovation of the R-D model is the incorporation of ontological features that reflect the referential nature of lexical items. The interpretation of reduplication is

constrained by the ontological profile of the base, which can be captured via two binary features: [±time] and [±space]. These features are grounded in the referential domains that different syntactic categories typically invoke:

1. Verbs are expressions of eventualities, which refer to actions, events, states and, processes which unfold in time rather than being located in space. Therefore, they are marked as [+time, –space].
2. Prepositions encode spatial relationships without eventive reference, and are marked [–time, +space].
3. Nominals refer to entities that exist either in space or time or both space and time; thus, they are [+time, +space].
4. Modifiers such as adjectives, numerals and adverbs are non-referential in this ontological sense, lacking direct spatiotemporal anchoring, and are assigned [–time, –space].

These features underpin the hypothesis advanced in this study, which seeks to unify the effects of intensification, distribution, and iteration through an operator set (R-D) comprising a Reduplicator (R), an Iterator (I), and a Distributor (D). The operator set interacts with lexical items, and the interpretation is conditioned by their syntactic category and its interaction with the R-D set especially the D-operator.

4.2 The R-D Operator Set as an Areal Feature of India as a Linguistic Area

The main proposition of the study is that the R-D Operator set is a grouping of 3 distinct operators:

1. The Reduplicator (R) – The operator responsible for reduplication of the Phonological Form.
2. The Iterator (I) – The operator which iterates the instances of the lexical item.
3. The Distributor (D) – The operator which distributes the iterations of the lexical item through time or space depending on the ontological features of the lexical item.

The D-operator is the only one from the R-D set that is sensitive to the ontological features. It can distribute a lexical item only through the dimension of the feature(s) marked [+].

This means that the D-operator can distribute the instances of a lexical item L through time if and only if L has [+time] feature. This constraint allows derivation of the meanings based on syntactic categories as required. A sample example of how this works is depicted below:

Consider a word W with the features [-time, +space] and predicate function word(x). When R-D applies to this word, the following happens:

1. The R-operator reduplicates the word in its phonological form and gives the output W~W.
2. The I-operator iterates the instances of the predicate function of W and returns multiple predicate functions as a set {word(x), word(x), word(x) …}.

3. The D-operator then distributes the multiple instances generated by the I-operator through the spatial dimension because of the [+space] feature and returns a meaning of distribution through space. A meaning of distribution through time is blocked because W has [-time] feature.

The second main proposition of this study is that the R-D operator set is a shared feature in the Indian Linguistic Area. The operators themselves might be universal to all languages of the world. In India(and broadly South Asia), these operators are grouped together giving rise to the varied meanings of reduplication across syntactic categories which are uniform across languages in the region.

The model's application on real data taken from Indian languages will be explored in section 5.

5. Findings and analysis

Reduplication and its associated meanings of intensification, iteration and distribution are prevalent in the Indian subcontinent's various languages. However, there are nuances in the spread and depth of these features. The crosslinguistic data from different language families is analysed and these nuances are explored category-wise.

5.1 Nominals

Nominals are entities usually marked with [+time, +space] features. When R-D applies to them, the resultant meaning is distribution through the entities. In Balusu (2006), this reading is referred to as "the participant key reading." This study posits that the participant key reading is in fact the D-operator distributing through both time and space simultaneously. The entities within the universe of discourse would hence be the points of distribution as they are [+time, +space].

Common nouns, anaphors and pronouns are affected in the same way and give rise to distributive meanings.

Malayalam (Dravidian):
ellaːʋarum aʋaɾ-aʋaɾ-uɖe kaseːrajil iɾikʲkʲuka
everyone they~they-GEN chair-LOC sit
'Everyone sit in their own chairs'

Telugu (Dravidian):
san̪d̪u-san̪d̪u vet̪akaɳɖi
street~street search
'Search every street'

Hindi (Indo-Aryan):
gʰəɾ- gʰəɾ kiː kəhaːnɪ hẽ
house-house of story is
'(It is) the story of every house'

Marathi (Indo-Aryan):
ʈumt͡ʃa:-ʈumt͡ʃa: kurt͡ʃivar basa:
your~your chair:LOC sit
'Sit in your own chairs'

Ho (Munda):
hora-hora te-lel hud͡ʒui-me
route~route PROG-see come-IMP
'Look at each and every route and come'

Bodo (Tibeto-Burman):
lama-lama tʰɯŋd͡ʒɯŋ tʰaŋ
road~road straight go
'Keep taking every straight road (do not turn)'

Ao (Tibeto-Burman):
pi:-pi: kidaŋi oaŋ
you~you house go
'Go to your own houses'

5.2 Verbs

Verbs have [+time, -space] features and are hence distributed through time by the D-operator. This results in the reduplicated verb perceived as being of an extended duration. An interesting point is that a reduplicated verb cannot be the main verb of the sentence, i.e. reduplication cannot happen to verbs in the matrix clause.

Indian languages differ in how reduplicated verbs are formed. The Indo-Aryan languages utilise *converb constructions* while the Dravidian languages utilise regular verb forms themselves. This leads to an interesting consequence: reduplicated verb forms in Dravidian languages can be used in their non-reduplicated versions while those in Indo-Aryan languages cannot be.

Malayalam (Dravidian):
ɳaɖaɳɳɯ-ɳaɖaɳɳɯ kʃi:ɳit͡ʃu
walk.PTCP~walk.PTCP became tired
'I got tired from walking.'

Marathi (Indo-Aryan):
lihuɳ-lihuɳ ʈʰak le
write.PTCP~write.PTCP tired of
'(I) got tired of writing.'

Hindi (Indo-Aryan):
t͡ʃəlʈe:-t͡ʃəlʈe: mɛ:ne: ʊska: ha:ʈ pəkəɽa:
walk.PROG~walk.PROG I his hand held
'I held his hand while walking'

Ao (Tibeto-Burman):
ken aŋa-aŋa bilem-oko
song listen.PTCP~listen.PTCP got-tired
'(I) got tired of listening to the song'

Bodo (Tibeto-Burman):
aŋ gaɾi sɯlai-sɯlai fɯinai
I vehicle change~change to come
'I had to change vehicles many times to come'

Ho (Munda):
ɖuʔub-ɖuʔub-ʈe hud͡ʒuʔu-me
sit.PTCP~sit.PTCP come-IMP
'Sit and come' (implies sitting many times in different places)

Santali (Munda):
ɖuɽub-ɖuɽub-ʈiŋ laːŋga-enaiʔin
sit.PTCP~sit.PTCP bored-of
'Bored of sitting for a long time'

Telugu goes a little further and distinguishes the verbal reduplicated meanings based on the aspect of the verb. For progressive verbs, reduplication denotes only an extended duration of the verb, while for perfective verbs a second meaning of "multiple instances" is also available.

naɖit͡ʃi-naɖit͡ʃi alisipojaːnu
walk.PERF~walk.PERF got tired
(I) got tired of walking

naɖusʈu-naɖusʈu alisipojaːnu
walk.PROG~walk.PROG got tired
(I) got tired while walking

5.3 Adpositions:

In Indian languages, adpositions mostly follow the noun and are hence postpositions. They have the [-time, +space] features and are hence distributed through space by the D-operator.

Malayalam (Dravidian):
kuʈʈikaɭ aɖuʈʈu-aɖuʈʈu iɾuɳɳu
children near~near sat
'The children sat side by side'

Banjara (Indo-Aryan):
maːr paːt͡ʃ-paːt͡ʃ aːd͡ʒoː
me behind~behind come
'Come behind me' (keep following me)

Hindi (Indo-Aryan):
ʊske piːt͡ʃeː-piːt͡ʃeː jaːo
his behind~behind go
'Go behind him' (follow him)

Santali (Munda):
sur-sur-laŋ duɽub aː
near~near-both sit will
'Both will sit side-by-side'

Ho (Munda):
abu buru kuɽi-kuɽi-ṯe senuwa
we jungle side~side-from will go
'We will go through the side of the jungle' (to keep taking one particular side)

Based on the dataset and speaker sample available for this study, no instances of reduplication involving adpositions were attested in the Tibeto-Burman languages. One interesting finding was that some native speakers don't distribute the postposition and take an iterative meaning instead. In Telugu, for example, distributive reading of postpositions is absent for some native speakers. They instead perceive the reduplicated forms in an iterative way.

peŋkulanu pai-paina peʈʈu
Tiles up~up put
'Put the tiles on top of each other' (distributive reading)

peŋkulanu pai-paina peʈʈu
Tiles up~up put
'Cover (something else) with the tiles superficially' (pragmatic meaning derived from iterative reading)

5.4 Modifiers

Modifiers like adjectives, adverbs and numerals have [-time, -space] features. When R-D applies to them, the D-operator is unable to distribute the iterations generated by the I-operator. The study posits that this iteration without distribution is the reason for intensified meaning. The reason for this assumption of iteration is based on the meaning of the reduplicated form of the word for "again" in Indian languages.

Malayalam (Dravidian):
ɲaːn avanoːɖɯ ʋiːɳɖum-ʋiːɳɖum paraɲɲu
I him again~again told
'I told him again and again'

Marathi (Indo-Aryan):
mi puɳhaː-puɳhaː ʋit͡ʃaːɾle
I again~again asked
'I asked again and again'
As can be seen in the above sentences, the intensification of "again" is iteration of the number of instances of "again". This can be extended to all other modifiers.

There is a difference in how adjectives and adverbs intensify as opposed to numerals. Numerals are quantifiable and hence intensification of numerals adds to the number of references.

Telugu (Dravidian):
pillalu muːɖu-muːɖu paɳɖlu konnaːɾu
kids three-three fruits bought
'The kids bought three fruits each'

Hindi (Indo-Aryan):
laɽkõ-ko pãːt͡ʃ-pãːt͡ʃ kɪʈaːbẽ mɪliː
boys five-five books got
'The boys got five books each'

For adjectives and adverbs, the intensification is qualitative and hence doesn't add to the number of references.

Malayalam (Dravidian):
ammuːmma paʈʈjeː-paʈʈjeː ɳaɖaɳɳu
grandmother slowly~slowly walked
'Grandmother walked very slowly'

Telugu (Dravidian):
ʋaːɖu mella-mella-gaː maːʈlaːɖæːɖu
he slow~slow-ly speak:PAST:3P.SING.MASC
'He spoke very slowly'

Marathi (Indo-Aryan):
laːmb-laːmb rasʈaː aːheː
long~long way is
'It is a very long way'

Banjara (Indo-Aryan):
eɳa: moʈe:-moʈe: gʰar ʧa:
here big~big houses are
'There are very big houses here'

Ho (Munda):
suwe-suwe sene-me
slow~slow walk-IMP
'Walk very slowly'

Santali (Munda):
lage-lage taɽam-me
fast~fast walk-IMP
'Walk very fast'

Ladakhi (Tibeto-Burman):
kʰo gjokspa-gjokspa ʧenuk
he fast~fast go.PRES
'He is going very fast'

Bodo (Tibeto-Burman):
bejɯ kʰaileŋ-kʰaileŋ tʰabadɯŋg
He lean~lean walks
'He walks in a zig-zag way'

6. Conclusion

This study has introduced the R-D operator set positing a unified, operator-based account of how reduplication interacts with syntactic category and ontological features to yield distributive, iterative, and intensifying meanings in Indian languages. By positing three distinct operators—Reduplicator (R), Iterator (I), and Distributor (D)—and grounding their effects in binary time/space features of syntactic categories, the model successfully captures cross-linguistic patterns in Indo-Aryan, Dravidian, Munda, and Tibeto-Burman data. Empirical findings demonstrate that nouns systematically invoke spatial-temporal distribution, verbs yield extended or repeated events, adpositions distribute through space, and modifiers intensify via iteration without distribution. The R-D framework thus offers a compositional, typologically robust explanation for a wide array of reduplicative phenomena across South Asian languages.

The reliance of this study on elicited data calls for broader corpus-based validation, especially in under-documented Munda and Tibeto-Burman varieties. Future research could refine the categories (e.g. temporal entities like month), explore interactions with aspect and information-structure, and test the model's predictions in real-time processing and acquisition studies.

List of Abbreviations

R-D – Reduplicator-Distributor (Semantic Model)
R – Reduplicator (Operator for phonological reduplication)
I – Iterator (Operator for iterating instances)
D – Distributor (Operator for spatial/temporal distribution)
GEN – Genitive (case marker)
LOC – Locative (case marker)
IMP – Imperative (mood marker)
PTCP – Participle
PROG – Progressive aspect

L1 / L2 – First Language / Second Language
[+time] / [-time] – Temporal ontological feature (positive/negative)

NOM – Nominative (case marker)
1SG, 2SG, 3SG – First, second, third person singular
1PL, 2PL, 3PL – First, second, third person plural
AUX – Auxiliary verb
DSTR – Distributive
ADV – Adverb
V – Verb
NP – Noun Phrase
PP – Prepositional Phrase
SOV – Subject-Object-Verb (word order)

feature (positive/negative)
[+space] / [-space] – Spatial ontological feature (positive/negative)

References

Abbi, Anvita. 1990. Reduplication in Tibeto-Burman languages of South Asia. *Japanese Journal of Southeast Asian Studies*, 28(2), 171–181.

Abbi, Anvita. 1985. Reduplicative structures: A phenomenon of the South Asian linguistic area. *Oceanic Linguistics Special Publications*, 20, 159–171.

Balusu, Rahul. 2006. Distributive reduplication in Telugu. In *Proceedings of NELS 36*.

Balusu, Rahul and K. A. Jayaseelan. 2013. Distributive quantification by reduplication in Dravidian. In Lisa Matthewson (ed.), *Strategies of Quantification*, 60–86. Oxford: Oxford University Press.