15 Oct. 2026

CONSTRUCTING THEORIES OF CONSTRUCTIONS: SENTENCES, WORDS, MORPHEMES AND MORPHS[1]

K P Mohanan and Tara Mohanan
mohanan.kp@gmail.com *tara.mohanan@gmail.com*

CONTENTS

1 INTRODUCTION

2 SYNTACTIC CONSTRUCTIONS: SOME EXAMPLES
- *2.1 Constructions Within and Across Languages*
- *2.2 V + it vs. V + that + it*
- *2.3 Relative Clause vs. Cleft*
- *2.4 What we have done so far*

3 MORPHOLOGICAL CONSTRUCTIONS
- *3.1 The Causative Construction*
- *3.2 The Passive Construction*
- *3.3 Other Examples of Constructions*

4 PHONOLOGY AND SEMANTICS

5 CONCLUDING REMARKS

1 Introduction

This article is an attempt to answer a question that we think is important for the communities of Dravidian Linguistics, and Indian Linguistics in general, to seriously engage with:

> *How do we bring work on Dravidian and South Asian languages in general to the attention of the international community of linguists, such that work on these languages will occupy a prominent position in world linguistics?*

In our own journey in linguistics, we (KPM and TM) have worked on theories of language structure, using evidence from Malayalam and Hindi. Our response to the above question would naturally focus on theories of language structure — phonology, morphology, syntax, and semantics: how do patterns of structure in the South Asian languages contribute to the pool of *world linguistics*.

[1] This article has benefitted significantly from comments from Wee Lian-Hee, Praveen Singh, and Vigneshwar Ramakrishnan.

Within that terrain, we will narrow our focus to what can be called constructions (e.g., passive construction, cleft construction, causative construction, applicative construction, complex predicate construction, and so on) which illuminate what needs to be investigated in the structure of sentences, words, and morphemes along the dimensions of syntax and semantics.

There are two suggestions that we would like to invite research scholars in Dravidian and South Asian linguistics to consider.

Our first suggestion is to adopt in inquiry the norms of constructing and evaluating scientific theories, with the aim to publish the work in prominent, widely circulated journals.

There are articles in the *Journal of Dravidian Linguistics* and the *Journal of Indian Linguistics* that present descriptions of South Asian Languages, without advancing knowledge claims, let alone presenting evidence and arguments in support of the claims. There are also articles that use the descriptive data as evidence in support of their claims, at the level of either grammars or linguistic theory. We would like to encourage the second category of research, exemplified by the articles in the *Journal of South Asian Linguistics*.

Our second suggestion is to pay special attention to ***constructions*** that are characteristic of Dravidian languages, and expand the terrain to languages that come under 'India as a Linguistic Area' (Masica 1975), covering Indian languages. This would bring interesting problems in these languages to the attention of linguists outside South Asia, and encourage them to take an interest in them. These recommendations can be extended to include all South Asian languages in general.

The spirit in which we are making these suggestions in this article is analogous to what experienced painters might do to help novice painters develop their capacity to paint well. Different painters have different strands of expertise, so each 'master painter' may provide advice or guidance on what (s)he is best at, and expect the learners to seek advice and guidance from other elders in the community of painters. We see our role as experienced linguists guiding and advising novice linguists, drawing on the analogy of a community of experienced and novice painters.

2 Syntactic Constructions: Some Examples

2.1 Constructions Within and Across Languges.

Consider the English sentences in (1), the corresponding Hindi sentences in (2), and the Malayalam sentences in (3):

(1) a. Hari killed a mosquito.
    b. A mosquito was killed by Hari.
    c. It was a mosquito that Hari killed.
    d. What was it that Hari killed?

(2) a. hari-ne ek macchar-ko maar-diyaa.
Hari-ERG one mosquito-ACC kill-PAST
Hari killed a mosquito.

b. hari-se ek macchar maaraa-gayaa.
Hari-INST one mosquito-NOM kill-PASS-PAST
A mosquito was killed by Hari.

c. ek macchar thaa jis-ko hari-ne maar-diyaa.
one mosquito-ACC be-PAST who-ACC Hari-ERG kill-PAST-IT
It was a mosquito that Hari killed.

d. kyaa thaa jis-ko hari-ne maar diyaa.
what be-PAST who-ACC Hari-ERG kill-PAST-IT
What was it that Hari killed?

(3) a. hari oru kotuwine konnu.
Hari-NOM one mosquito-ACC kill-PAST
Hari killed that mosquito.

b. hariyaal oru kotu kollappeTTu.
Hari-INST one mosquito-NOM kill-PASS-PAST
A mosquito was killed by Hari.

c. hari oru kotuwineyaaN∂ konnat∂.
Hari-NOM one mosquito-ACC-BE kill-PAST-IT
It was a mosquito that Hari killed.

d. hari enthineyaaN∂ konnat∂?
Hari-NOM what-ACC-BE kill-PAST-IT
What was it that Hari killed?

The sentences (1a), (2a) and (3a) are examples of the ***declarative active construction***. Those in (1b), (2b), and (3b) are instances of the ***passive construction***. Those in (1c), (2c), and (3c) are examples of the ***cleft construction***. And those in (1d), (2d), and (3d) are examples of the ***interrogative cleft construction***.

In the grammar of a language, we are interested in the relation between related constructions, for instance, active and passive, cleft and non-cleft, declarative and interrogative, and so on. In the analysis of constructions across languages, we are interested in homology and diversity of structure in biology, as in the case of the tripartite structure of the leg in humans, cows, and birds. We may call this unity in diversity, or invariance and variability.

2.2 V + it ***vs.*** *V + that + it*

Consider the Malayalam sentences in (4):

(4) a. *hari* ***paaTiyat∂*** *amma* *keeTTu.*
Hari-NOM sing-PAST-IT mother-NOM hear-PAST
Mother heard Hari singing.

b. *hari* ***paaTiyenn∂*** *amma* *keeTTu.*
Hari-NOM sing-PAST-COMP mother-NOM hear-PAST
Mother heard that Hari sang.

c. *hari* ***paaTiyennatə*** *amma* *keeTTu.*
Hari-NOM sing-PAST-COMP-IT mother-NOM hear-PAST
Mother heard (the fact) that Hari sang.

The difference between (4a), (4b), and (4c) lies in the forms in (5):

(**5**) a. *paaTiy**atə***
paaTi + at**ə**
sing-PAST-IT

b. *paaTiy**ennə***
paaTi + enn**ə**
sing-PAST-that

c. *paaTiy**ennatə***
paaTi + enn**ə** +at**ə**
sing-PAST-that-IT

The semantic contrast between the forms in (5) is located in the contrast between three morpho-syntactic ***constructions***:

(**6**) a. V-TENSE-NOMR b. V-TENSE-COMPR c. V-TENSE-COMPR-NOMR
[where NOMR: Nominaliser; COMPR: Complementiser]

Notice that the nominaliser *atə* 'it' in (5a) refers to an action, while that in (5c) it refers to a proposition.

A proposition is the semantic content of a declarative sentence, i.e., what the sentence asserts to be true. Thus, the sentences, "Joan and Mary are sisters," "Joan is Mary's sister," and "Mary is Joan's sister," are three distinct sentences that assert the same proposition. Conversely, "John touched the key," can assert two distinct propositions, depending on whether 'key' denotes something that opens a lock, or what you press on a piano to produce a note.

In contrast to (5b), (5c) has the additional meaning of 'factivity'; that is, the proposition expressed by the embedded clause is taken to be a fact. These contrasts are central to the critical evaluation of the truth of the propositions asserted by the sentences in (4).

The challenge that these constructions pose for theories of language structure can be stated as:

(**7**) a. How do we represent the meanings of V, *ennə,* and *atə*?
b. How do we represent the meanings of V+atə, V+ennə, and V+ennatə?
c. How do we derive (7b) from (7a)?[2]

2.3 Relative Clause vs. Cleft

Compare the sentence in (8) with the corresponding constructions in (9a) and (9b):

(**8**) *raaman* *meeriye* *karayippiccu.*
Raman-NOM Mary-ACC cry-CAUSE-PAST
Raman made Mary cry.

(**9**) a. *meeriye* *karayippicc**awan*** *raamanaaNə*
Mary-ACC cry-CAUSE-PAST-**he-**NOM Raman-NOM-is
The one who made Mary cry is Raman.

[2] The term *derive* in (7c) means "to identify the logical consequence of a proposition." Such derivations in logic are analogous to mathematical calculations: we can derive 16 from (3x5) + 1.

b. *meeriye* *karayippicc**atə*** *raamanaaNə*
Mary-ACC cry-CAUSE-PAST-**it** Raman-NOM-is
It was Raman who made Mary cry.

The English sentence glosses given for (9a) and (9b) indicate that we analyse *raamanaaNə*:

in (9a): as the head of the relative clause that precedes it; and

in (9b): as the clefted unit.

The *awan* in (9a) is a pronominal head, while *atə* in (9b) is not a pronoun but an expletive *it*, as in the case of *it* in the English cleft, "It was Raman who made Mary cry."

The structural representations for (9a, b) are (10a, b) respectively:

(**10**) a. [[*meeriye* *karayippi**cc***] ***awan***] [*raamanaaNə*]
Mary-ACC cry-CAUSE-PAST-**he-**NOM Raman-NOM-is
The one who made Mary cry was Raman.

b. [[*meeriye*] [*karayippi**cc***] ***atə***] [*raamanaaNə*]
Mary-ACC cry-CAUSE-PAST-**IT** Raman-NOM-is
It was Raman who made Mary cry.

What evidence and arguments support the postulation of these representations?

To answer that question, we need to contrast (9a) and (9b) with the corresponding sentences in (11a) and (11b):

(**11**) a. *raaman* *karayippiccawaL* ***meeri****yaaNə*
Raman-NOM cry-CAUSE-PAST-she-NOM Mary-**NOM**-is
The one that Raman made cry was Mary.

b. *raaman* *karayippiccatə* ***meeriye****yaaNə*
Raman-NOM cry-CAUSE-PAST Mary-**ACC**-is
It was Mary that Raman made cry.

In the relative clause construction ((9a) and (11a)), *aaNə* is a **copula**, and the units immediately preceding it — *raaman* in (9a) and *meeri* in (11a) — are dependents of the copula, which assigns the nominative case in both (9a) and (11a).

In contrast, in the cleft construction ((9b) and (11b)), *aaNə* is ~~a~~ **cleft marker**, not a copula. The unit that the cleft marker is attached to is a dependent of the main verb to which *atə* is attached. The verb, *karayippikk-* assigns NOMINATIVE case to its SUBJECT *raaman*, and ACCUSATIVE case to the OBJECT *meeri*. Our representations in (10a) and (10b) correctly predict these distinctions.

Suppose we replace the copula *aaNə* in the relative clause construction with the transitive verb, *kaNTu* 'saw'.

(**12**) a. *raaman* *karayippiccawaL* *yakSiye* *kaNTu*
Raman-NOM cry-CAUSE-PAST-she-NOM yakshi-ACC see-PAST
The one that Raman made cry saw a *yakshi* [3].

[3] *Yakshi*: A mythical female nature spirit.

b. *raaman* *karayippiccawaLe* *yakSi* *kaNTu*
Raman-NOM cry-CAUSE-PAST-she-ACC yakshi-NOM see-PAST
The *yakshi* saw the one that Raman made cry.

In the cleft construction, *aaNə* cannot be replaced by some other verb:

(**13**) a. ***raaman** *karayippicc**atə*** *yakSiye* *kaNTu*
Raman-NOM cry-CAUSE-PAST-**IT-NOM** yakshi-ACC see-PAST

b. *raaman karayippiccatine *yakSi* *kaNTu*
Raman-NOM cry-CAUSE-PAST-**IT-ACC** yakshi-NOM see-PAST

All the contrasts outlined above are correctly predicted by our representations in (10), in combination with the laws of case assignment.[4]

Our second piece of evidence comes from relative word order. The three units in (10b) can be scrambled with respect to each other.

(**14**) a. *meeriye* *karayippiccatə* *raamanaaNə*
Mary-ACC cry-CAUSE-PAST-**IT** Raman-NOM-is
It was Raman who made Mary cry.

b. *meeriye* *raamanaaNə* *karayippiccatə*
Mary-ACC Raman-NOM-is cry-CAUSE-PAST-**IT**
It was Raman who made Mary cry.

c. *raamanaaNə* *meeriye* *karayippiccatə*
Raman-NOM-is Mary-ACC cry-CAUSE-PAST-**IT**
It was Raman who made Mary cry.

d. *raamanaaNə* *karayippiccatə* *meeriye*
Raman-NOM-is cry-CAUSE-PAST-**IT** Mary-ACC
It was Raman who made Mary cry.

As correctly predicted by our representations together with the clause boundedness of scrambling, the word order freedom we see in (14) is not permitted in the relative clause construction in (15):

(**15**) a. *meeriye* *karayippiccawan* *raamanaaNə*
Mary-ACC cry-CAUSE-PAST-he-NOM Raman-NOM-is
The one who made Mary cry was Raman.

b. *raamanaaNə* *meeriye* *karayippiccawan*
Raman-NOM-is Mary-ACC cry-CAUSE-PAST-**he-**NOM
The one who made Mary cry was Raman.

c. * *raamanaaNə* *karayippiccawan* *meeriye*
Raman-NOM-is cry-CAUSE-PAST-**he-**NOM Mary-ACC

d. **karayippiccawan* *raamanaaNə* *meeriye*
cry-CAUSE-PAST-**he-**NOM Raman-NOM-is Mary-ACC

[4] Examples of the laws of case assignment include:
"In Malayalam, subjects and inanimate objects are assigned nominative case, while animate objects are assigned accusative case."

2.4 What we have done so far

In sections 2.2 and 2.3, we demonstrated the difference in behaviour between the relative clause construction and the cleft construction in Malayalam. In doing so, we provided two examples of how to construct theories of constructions.

An important requirement for the construction and evaluation of scientific theories is that they ***predict*** and ***explain*** asymmetries that call for explanations, not merely provide a body of data.[5]

To do this, we need to organise and present data such that:

A. they contain pairs of sentences, with only one difference between the members of the pair (only one variable changed in one member of the pair); and

B. the difference corresponds to a difference in the behaviour (in acceptability or in meaning) of the two sentences.

Section 2.2 demonstrated a difference in the meanings of three constructions, which differ only in the **presence vs. absence** of the nominaliser *atə*, or the complementizer *ennə*:

paaTiyatə (V + PAST + *IT*)

paaTiyennə (V + PAST + that)

paaTiyennatə (V + PAST + *that* + *it*).

These contrasts constitute the equivalents of:

a) what is called a ***minimal pair*** in phonology: a minimal difference in form that corresponds to a difference in meaning;

b) the principle of changing only ***one variable at a time*** in experimental science; and

c) ***ceteris paribus*** 'all else being equal' in other domains.

For our purposes, the asymmetries are those of behaviour under experimental conditions.

There are two important aspects of theory construction that we have taken pains to demonstrate above:

One of them is the need to not only to provide data, but also to provide them in a way that makes it clear to the evaluator what the theory needs to and sets out to ***predict and explain***.

In section 2.2, for instance, we pointed to what we need to predict and explain as the contrasting behaviour of the minimal pairs in (4), where the **presence**

[5] It is important to pay attention to the technical meanings of the terms 'predict' and 'explain'. In common English, 'predict' means 'foretell the future', as in the case of predictions in astrology and weather forecasts. In scientific inquiry, it means 'deduce the logical consequences' of the statements of a theory. In ordinary English, 'explain' means 'elucidate', as in "The teacher explains difficult concepts very well." In scientific inquiry, it means 'provide reasons for something that is puzzling.' To prove that an alleged prediction of a scientific theory is indeed a prediction, we have to provide a derivation of the prediction from the statements of the theory. Such derivations are like derivations in mathematics: e.g., deriving 5 from (5x2) – (15-10): (5x2) – (15-10) = 10 – (15-10) = 10 – 5 = 5.

vs. absence of the nominaliser *atə* and the complementizer *ennə* (all else being equal) corresponds to important contrasts in meaning.

Section 2.3. demonstrated how we defend our theoretical proposals by providing relevant evidence and rigorous arguments ***in support of our proposals*** and ***against alternatives*** in the literature, or the potential alternatives that readers might think of. This is particularly needed in the context of debates — for instance, between K P Mohanan (1982) and P Madhavan (1987) on whether or not Malayalam clefts are mono-clausal or biclausal. Mohanan (1982) assumes that clefts are monoclausal:

[[hari] [aTicca tə] [raamaneyanə]]
Hari-NOM beat- PAST-IT Raman-ACC-BE one bottle-
It was Raman that Hari beat.

Here, aaNə is not a copula, even though we have glossed it as a form of the verb BE for morpho-phonological purposes: as far s syntax is concerned, it is a cleft marker, and hence the above structure is monoclausal.

In contrast, Madhavan assumes that it is biclausal:

[[hari] —— [aTicca tə]]$_{NP}$ [[raamane]$_{NP}$ aanə]]$_{VP}$
Hari-NOM beat- PAST-IT Raman-ACC BE
It was Raman that Hari beat.

In this analysis, [[hari] —— [aTicc]atə]] is a nominal clause, with *atə* as a nominaliser, and [[raamane]$_{NP}$ aanə]] is the VP that goes with the subject NP. Internal to the clause inside the subject is a gap/trace, coindexed with the pronominal head atə.

If Malayalam clefts are monoclausal, as Mohanan argues, the standard Chomskian analysis of move-alpha would not extend to them, even in terms of the latest Chomskian theories. If, on the other hand, Malayalam clefts are bi-clausal, as Madhavan argues, Chomskian theory does not require a revision. Which of these analyses we should accept in syntactic theory is for the readers to decide.

3. Morphological Constructions

3.1 The Causative Construction

Consider the relation between the intransitive verb *ooT-um*, the transitivised *ooT-ikk-um*, and the doubly transitivised *ooT-ipp-ikk-um* in the sentences (16a-c) below:

(**16**) a. kuTTi ooTum
child-NOM run-FUT
The child will run.

b. amma kuTTiye ooTikkum
mother-NOM child-ACC run-CAUSE-FUT
The mother will make the child run.

c. acchan ammayekkonTə kuTTiye ooTippikkum
father-NOM mother-INST child-ACC run-CAUSE-CAUSE-FUT
The father will get the mother to make the child run.

We will use the terms ***morph*** and ***morpheme*** to denote:
minimal morpho-phonological units, and
minimal morpho-syntax-semantics units respectively.

To illustrate the correspondence/association between the morphic and morphemic levels with the verb forms in (16), we will use vertical lines to signal the association as in autosegmental phonology:

(**17**)	cited form gloss	***a***. *ooTum* *will run*	***b***. *ooTikkum* *will cause to run*	***c***. *ooTippikkum* *will cause to make run*
	morphs	/ooT/ /um/	/ooT/ /ikk/ /um/	/ooT/ /ipp/ /ikk/ /um/
	morphemes	OOT FUT	OOT CS FUT	OOT CS CS FUT

The verb form in (17a) has the future tense marker,and no causative morph. In (17b), it has one causative morph; and in (17c), it has two causative morphs. [In the morphemic representations in (17), CS stands for CAUSE.]

Note that in (16c), the first causative morph is realised as /ipp/, while the second is realised as /ikk/. Why? The answer lies in what is called the Obligatory Contour Principle (OCP), proposed originally for tonal phonology in Leben (1973), and Goldsmith (1976), extended in McCarthy (1986) to vowels and consonants in Arabic roots, and extended further in Mohanan (1994) to case morphs in Hindi.

OCP as it appears in the works mentioned above can be generalised to capture the phenomenon that the constraint describes, namely, the tendency for languages to prohibit identical linguistic units in adjacent positions. The universal constraint of GOCP (Generalised OCP) can be stated as:

(**18**) **GOCP**: Identical adjacent units are prohibited.

Which units, and along which dimensions, submit to this prohibition is specified in individual grammars. If the causative morph /ikk/ appears twice in (16c), it would violate the GOCP of morphs, and hence, it is replaced by the allomorph /ipp/.

As stated in the introduction, it is important for theories of language structure to adopt the norms of scientific inquiry. One such norm for theory construction is the need to engage with alternative explanations. We must ask:

Why should we explain the morphophonemic patterns of causatives described above in terms of GOCP?
Why not use either (19a) or (19b)?
(**19**) a. ikk → ipp / — ikk b. ipp → ikk / ipp —

The answer is this: (19a) would not be able to rule out the possibility of three causative morphemes with two occurrences of /ipp/, and (19b) would not be able to rule out two occurrences of /ikk/.

Interestingly, while GOCP holds on morphs, it does not rule out multiple occurrences of the same morpheme. Notice that (16c) has two adjacent occurrences of the morpheme CAUSE. An outline like the account above of the

causative construction in Malayalam raises several interesting theoretical questions which we will not be able to address within the confines of a short article.

One last point before we turn to the next construction. In defending the analysis using GOCP of morphs in the causative construction in the grammar of Malayalam, we are in effect defending the constraint in (18) as part of Universal Grammar. This is how a rigorous theoretical work on a Dravidian language can contribute to theories of language structure. Mere descriptions of Malayalam would not have an impact in world linguistics.

3.2 The Passive Construction

We saw the Malayalam passive construction in (3b). More examples are given in (21b) and (22b) to further illustrate the construction:

(**20**) kuppi poTTum.
bottle-NOM break-FUT
The bottle will break.

(**21**) a. hari kuppi poTTikkum.
Hari-NOM kuppi-NOM break-CAUS-FUT
Hari will break the bottle.

b. hariyaal kuppi poTTikkappeTum
Hari-INST bottle-NOM break-CAUSE-PASS-FUT
The bottle will be broken by Hari.

(**22**) a. amma kuTTiyekkoNT∂ kuppi poTTippikkum
mother-NOM child-ACC-INST bottle-NOM break-CAUSE-CAUSE-FUT
Mother will make the child break the bottle.

b. ammayaal kuTTiyekkoNT∂ kuppi poTTippikkappeTum
mother-INST child-ACC-INST bottle-NOM break-CAUSE-CAUSE-PASS-FUT
The bottle will be made by mother to be broken by the child . .

As in the case of (19d), speakers of Malayalam find it hard to interpret (22b) even though the verb is morphologically well-formed.

There are two features that are theoretically interesting about the Malayalam passive construction.

First, a great deal of the analysis of passives (from Chomsky (1957) up to Chomsky (1981)) involves differences in word order. That analysis works for languages like English where grammatical functions are coded in terms of word order, but not for languages like Malayalam where grammatical functions are coded in terms of case, and which allow all the six word order possibilities of subject, object, and verb. What is possible in Malayalam, as in other languages, is the assignment of subjecthood to the undergoer in the passive construction, coupled with the assignment of adjuncthood to the subject of the active sentence — the analysis in Relational Grammar and Lexical Functional Grammar.[6]

[6] For a brief introduction to Relational Grammar, see the Wikipedia entry:

Second, when combined with the discussion of the causative construction in section 3.1, it raises a serious challenge to theories of morphology such as Amorphous Morphology and Whole Word Morphology.[7] These theories deny the need for the postulation of MORPH as a theoretical construct, as distinct from the concept of MORPHEME: they do not sanction the analysis of the causative-passive construction where the morphemic string:

VERB+CAUSE+CAUSE+PASS+FUT

is realised as [verb morph + ipp + ikk + appeT + um].

As far as we can tell, an alternative analysis without morphs is not feasible.

Perhaps this comment calls for a bit of exposition by way of clarifying terms:

Unit	Level of Representation
PHONE	minimal unit: PHONETIC
PHONEME	minimal unit: PHONOLOGICAL
ALLOPHONE of a phoneme	a phone associated with that phoneme
MORPH	minimal unit: MORPHO-PHONEMIC
MORPHEME	minimal unit: SYNTACTICO-SEMANTIC
ALLOMORPH of a morpheme	a morph associated with that morpheme

In traditional terminology, *allo-* refers to one of the instantions of X. For there to be an *allo-*, there have to be more than one instantions of X. For example, the English *-en* /n/ (as in *oxen*/ɒksn/) and *–es* /z/ (as in *cows*) are morphs that are allomorphs of the plural morpheme.

Notice that in the table above, we are departing from traditional use, to extend the term *allo-* to instances of morphemes that are associated with only one morph, as in the case of *re-* in *rearrange*, and *–ity* in *sanity*.

In Malayalam, *-um*, *-appeT*, *-ikk*, and *-ipp* are morphs.

-um is an instantiation (hence an allomorph) of the morpheme FUTURE.

-appeT is an allomorph of the morpheme PASSIVE in the word *koll-appeT-um* 'will be killed', but an allomorph of the verbalising morpheme in *sankaT-appeT-um* 'will be sad'.

-ikk is an allomorph of the verbalising morpheme in the word *dukkh-ikk-um* 'will grieve', but an allomorph of the morpheme CAUSE in the word *dukkh-ipp-ikkum* 'will cause to grieve'.

-ipp is an allomorph of the morpheme CAUSE in the word *ooT-ipp-ikk-um* 'will cause to X to cause Y to run'.

---

https://en.wikipedia.org/wiki/Relational_grammar.
For Lexical Functional Grammar, watch the YouTube Video, "Lexical-Functional Grammar 1: What is LFG?" at https://www.youtube.com/watch?v=EoCLhS_0cmE&t=503s

[7] For Amorphous Morphology, see "Morphology without word-internal constituents: a review of Stephen R. Anderson's A-Morphous Morphology" by Andrew Cartairs-McCarthy at https://link.springer.com/chapter/10.1007/978-94-017-3710-4_8
For Whole Word Morphology, see *Pace Pāṇini: Towards a word-based theory of morphology* by Alan Ford, Rajendra Singh, and Gita Martohardjono

We assume that the following constraints are uncontroversial.

Every sentence must have at least one word.
Every word must have at least one syllable.
Every syllable must have at least one phone.

We extend parallel constraints to morphs and phones:

Every morph must have at least one phoneme.
Every phoneme must have at least one phone.

Given these constraints, the theory we assume does not permit zero morphs, zero words, or zero phrases (null categories or traces). Hence, there is no allomorph of the singular morpheme in *mouse*, and no allomorph of the plural morpheme in *mice*.

3.3 Other Examples of Constructions

As mentioned in the introduction, the second theme of this article is the need to pay special attention to the phenomena and the constructions that are characteristic of South Asian languages. From that perspective, this article would not be complete without mentioning a few of these constructions: the Conditional Construction, the Complex Predicate Construction, the Light Verb Construction, the Noun Incorporation Construction, and the Auxiliary Construction.

The Conditional Construction is typically a structure where the main clause of the sentence depends on a condition stated in a subordinate clause. It lends itself to stating general truths, as well as possible, hypothetical, or imaginary scenarios.

The Complex Predicate Construction is one where two grammatical elements combine to function as a single predicate — whether N-V, Adj-V, or V-V — where either of the two units can contribute meanings and arguments to the mother predicate. (Butt (1995); Alsina et al (2008))

A Light Verb Construction is a particular Complex Predicate construction in which the meaning and argument structure of second unit have been 'bleached' out of existence. A detailed exploration of this construction is beyond the scope of this article, so we will simply point interested readers to the following article. (Mohanan (2017))

The Noun Incorporation Construction is one where a noun stem that is an argument the verb forms a lexical unit with the verb $[\text{N-V}]_{\text{V}}$, and the noun, although a part of the predicate, functions as the object of the verb. (Mohanan T. (1995))

The Auxiliary Construction in English has a main verb, preceded optionally by a modal verb, perfective, progressive, and passive, in that order, the maximum expansion of it being illustrated by (23):

(23) *That building might have been being built in 2012.*

The auxiliary construction in Malayalam is a single word with *that* and is preceded by a host of grammaticalized verbs or auxiliary verbs, as illustrated in (24), and ends with the main verb, where the main verb can itself be a serial verb, as in (24d).

(**24**) a. kuTTi ooTi
child-NOM run-PAST
The child ran

b. kuTTi ooTumaayirunnu
child-NOM run-FUT-BE-PAST-SIT-PAST
The child would run.

c. kuTTi ooTikkoNTirikkumaayirunneene
child-NOM RUN-PAST-KOL-PAST-SIT-FUT-BE-PAST-SIT-PAST-MODAL
It might have been the case that the child would keep running.

d. kuTTi ooTiccaaTinaTakkukayaayirikkum
child-NOM run-PAST-jump-PAST-walk-INF-BE-PAST-SIT-FUT
The child must be moving around, running and jumping.

Once again, all that we can do here is to point interested readers to Mohanan and Mohanan (2008).

4 Phonology and Semantics

The examples we have discussed so far involve the combination of syntax and semantics. That does not mean that there are no constructions that appeal to both phonology and semantics.

For ease of exposition, we will use tag questions in (25) to illustrate:

(**25**) a. Hari broke the bottles, didn't he?
b. It was Hari that broke the bottles, wasn't it?
c. The bottles were broken by Hari, weren't they?

A tag in a **tag construction** is composed of a pro-form for the verb and a pronoun that agrees with the subject of the stem. The unmarked intonation pattern for the tags in (25) is that of a falling intonation, unlike the rising intonation for the **yes-no question**, "Did Hari break the bottles?"

In the marked case, the tag has a rising intonation which signals a certain *degree of uncertainty on the part of the speaker*. The tag with the rising intonation forms a minimal pair with the tag with the falling intonation.

To take another example, the distinction between sub-compounds and co-compounds in Malayalam lies in their word melodies (Mohanan 1982). A sub-compound has a single word melody, while in a co-compound, each stem has its own word melody. (For an example from Urdu/Hindi, see Butt et al (2020).

Our intention here is to point to the need to investigate phono-semantic constructions, not to provide a comprehensive outline of different phono-semantic constructions.

Languages are like species in biology: we can call them linguistic species. To extend the analogy further, constructions are like organs in biology. That lends itself to the investigation of the structure of constructions, as in the case of the structure of the heart, the lungs, the arms, and so on. Just as the evolution of organs is an important pursuit in biological evolution, the

evolution of constructions needs to be an important pursuit in historical linguistics.

5 Concluding Remarks

To return to the question we raised in section 1:

> How do we bring South Asian languages to the attention of the international community of linguists, such that graduate students as well as established researchers working outside these languages would find in them something that is tempting to explore?

This article attempts to answer that question, using theories of language structure as the terrain, and using evidence primarily from Malayalam. We hope to have also provided illustrations of how these goals can be achieved. Our focus is on constructing and evaluating theories of constructions, which calls for subscribing to the norms of constructing and evaluating scientific theories, and facilitating the publication of work on these languages in international journals that are highly prestigeous and have international visibility.

References

Alsina A., Bresnan, J., and Sells, P. (eds) (2008) *Complex Predicates*. CSLI, Stanford: CA.

Butt, Miriam (1995) *The Structure of Complex Predicates in Urdu*. Stanford Ph.D. dissertation

Butt, Miriam, Jabeen Farhat, and Bögel Tina. (2020) "Ambiguity resolution via the syntax-prosody interface: The case of *kya* 'what' in Urdu/Hindi." *Prosody in Syntactic Encoding*. De Gruyter

Chomsky, Noam (1957) *Syntactic Structures*. …

Chomsky, Noam (1981) *Lectures on Government and Binding*. …

Leben, Will (1973). *Suprasegmental Phonology*. MIT Ph.D. dissertation

Goldsmith, John (1976). *Autosegmental Phonology*. MIT Ph.D. dissertation

Madhavan, P (1987) *Clefts and Pseudoclefts in Malayalam: a Study in Comparative Syntax*, Doctoral Dissertation, CIEFL, Hyderabad.

Masica, Colin P. (1976) *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press

McCarthy, John (1986). "OCP Effects: Gemination and Antigemination." Linguistic Inquiry 17, 207–264.

Mohanan, K. P. (1982) "Grammatical Relations and Clause Structure in Malayalam." In Bresnan, J. (ed.) *The Mental Representation of Grammatical Relations*. Cambridge: Mass: MIT Press, 1982.

Mohanan, T. and Mohanan K. P. (2008) "Multiple Tenses in the Malayalam Verb." In S Inkelas and K Hanson (eds) *The Nature of the Word*. The MIT Press. Cambridge: MA. USA (with K P Mohanan)

Mohanan, T. (1994) "Case OCP: A Constraint on Word Order in Hindi." M Butt, T King, and G Ramchand (eds) *Theoretical Perspectives on Word Order in South Asian Languages*. CSLI Publications. Stanford: CA. 185-216.

Mohanan, T. (1995) "Wordhood and Lexicality: Noun Incorporation in Hindi." *Natural Language and Linguistic Theory* 13.2:75-134. The Netherlands.

Mohanan, T. (2017) "Grammatical and Light Verbs." In M Everaert and H van Riemsdijk (eds) *The Companion to Syntax, Second Edition*. Chapter 49. Blackwell Publishers.