Malayalam Morphosyntax: Inflectional Features and their Acquisition

Submitted in partial fulfillment of the requirements

of the degree of

Doctor of Philosophy

by

Gayathri G

Roll No. 134083003

Supervisor:

Prof. Vaijayanthi M Sarma

Department of Humanities & Social Sciences

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

2019

For Priya …

Thesis Approval

This thesis entitled "Malayalam Morphosyntax: Inflectional Features and their Acquisition" by Gayathri G is approved for the degree of Doctor of Philosophy.

Examiners

2.12.2019

Supervisor

Chairman

Date: 14-11-2019

Place: Mumbai

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Gayathri G

Date: 14-11-2019

Roll No. 134083003

Abstract

Malayalam, which belongs to the South-Dravidian language family, is an agglutinative language with rich inflectional morphology. The aim of the thesis is to analyse the grammar and acquisition of Malayalam verbal inflections (tense, aspect and mood) and nominal inflections (case, number, and gender). Within the larger discussion of inflectional morphology and its acquisition, particular attention is paid to two complex morphological processes, a) the past tense formation of verbs and b) case assignment of subjects and objects.

In particular, the thesis will show the following: a) that the past tense marker selection is determined by different grammatical principles in underived and derived stems; specifically, *phonotactics* in the former and the lexical feature of *transitivity* in the latter; b) that the dative nominals of a class of predicates (variously labelled *experiencer* or *dative subject* or *psych predicates*) are in fact subjects using an array of empirical tests involving binding, control, accusative marking, and predicate alternation; and c) that inflections for number and object case rest on lexical features of the noun (stem) and the allomorphy is governed by these featural requirements. In looking at the developing grammar in the two subjects, the thesis will show that Malayalam inflectional grammar has quite direct consequences for the acquisition of inflectional morphology. Specifically, acquisition proceeds unobstructed when the mode of selection is phonological and offers more challenges when the mode of selection is morphological, i.e., when the selection depends on the learning of the lexical or grammatical features of the noun and verb stems.

Thus, using the interplay between acquisition and the grammatical description, we establish that in addition to the established factors that guide acquisition, *mode of selection* of an inflection plays a key role in determining the relative ease/difficulty in the acquisition of inflectional morphology. This follows quite neatly from the fact that children are phonologically competent even before much language is produced and that this module-competence could facilitate the acquisition of morphology. The thesis will argue that this is indeed the case.

Contents

List of Figures

List of Tables

List of Abbreviations

1	First Person	F	Feminine
2	Second Person	FNC	Functional Category
3	Third Person	FUT	Future
ABS	Absolutive	GEN	Genitive
ACC	Accusative	GF	Grammatical Function
ADJ	Adjective	HON	Honorific
ADV	Adverbial	HORT	Hortative
AGR	Agreement	HUM	Humanising Suffix
AgrP	Agreement Phrase	I	Inflection
ATOM	Agreement / Tense Omission Model	IMP	Imperative
AUX	Auxiliary	IMPF	Imperfect
C	Complementiser	INC	Incompletive Aspect
CAUS	Causative	INF	Infinitive
CGN	Case Gender Number	INS	Instrumental
COM	Completive Aspect	INTR	Intransitive
COMP	Complementiser	IP	Inflectional Phrase
COND	Conditional	IPFV	Imperfective
CORD	Coordinative Suffix	IV	Status Suffix for Intransitive Verb
CP	Complementiser Phrase	LAD	Language Acquisition Device
DAT	Dative	LEX	Lexeme
DER	Derivation	LF	Logical Form
DP	Determiner Phrase	LOC	Locative
EMPH	Emphatic	M	Masculine
EPP	Extended Projection Principle	MLU	Mean Length of Utterance
ERG	Ergative	MOD	Modal

N	Neuter
NEG	Negative
Neg	Negative Head
NegP	Negative Phrase
NH	Non-Human
NM	Non-Human Male
NMLZ	Nominaliser / Nominalisation
NOM	Nominative
NP	Noun Phrase
OI	Optional Infinitive
OPT	Optative
PASS	Passive
PET	Positron Emission Tomography
PFV	Perfective
PL	Plural
POSS	Possessive
PRF	Perfect
PRS	Present
PST	Past
PTCP	Participle
Q	Question Particle
REF	Reflexive
RI	Root Infinitive
SG	Singular
SOC	Sociative
SOV	Subject Object Verb
Spec	Specifier
SSP	Sonority Sequencing Principle
T	Tense
TAM	Tense Aspect Mood
TP	Tense Phrase
TR	Transitive
UG	Universal Grammar
V	Verb
VOC	Vocative
VP	Verb Phrase

Chapter 1

Introduction

1.1 Children and Language

It is evident even from lay observation that a child's knowledge of language increases with age in a uniform and rapid fashion. Hence it is only natural to ask what constitutes *knowledge* of a language. Knowledge of a language is identified with having a mentally represented grammar, which constitutes the native speakers' language 'competence' (Radford, 1997). Thus, to know a language is to possess such a grammar, that allows the speakers to generate legitimate sentences. The competence or the grammatical knowledge of native speakers of a language includes a tacit knowledge of the speech sounds of their language and how to combine them to make meaningful words. It also includes a knowledge of how to combine words to form meaningful sentences, how to resolve structural ambiguity in the sentences, and how to interpret them compositionally. Native speakers can also identify unacceptable combinations of phonemes, morphemes, and sentences in their language (Radford, 1997). Children grow up to become fully competent native speakers who have such a mentally represented grammar. Their development gives us glimpses into the innate language capacity and the constraints at work. Equally, they also allow us to understand aspects of the input that are critical to the development.

Despite being exposed to only a finite number of sentences, children apparently possess the capacity to understand and produce a potentially infinite number of sentences. What children have access to is the input they receive or *positive evidence* which comprises acceptable sentences in the target language, and only a subset of them. Yet, children's speech contains errors not seen in the input, for example, the regularised past tense forms in English such as *breaked** and *singed** or the regularised plurals such as *sheeps**. Although children are not ex-

plicitly taught which sentences or productions are correct and which are not, they eventually attain adult-like linguistic competence. Besides, children are seen not to make certain kinds of errors which they may logically be expected to make if they generalised solely from the received input. One often cited example is the *wanna*-contraction in English. Although children hear utterances such as *Who do you wanna invite?* and *Who do you wanna see?* (both involving object questions), they never generalise this to include constructions like **Who do you wanna come?* (subject question), though this could be a reasonable generalisation of the input (Guasti, 2002). These observations centre around the issue of the limited or impoverished input received and is known as the **poverty of the stimulus argument** (Chomsky, 1986; Guasti, 2002).

Different theories have been put forward to explain how a first language is acquired. These include a) behaviourist approaches (Skinner, 1957) which associate active learning through imitation, reinforcement, or association, b) constructivist/cognitivist approaches (Piaget, 1971; Goldberg, 1995) which link cognitive milestones with language growth where structures are assumed to be learned in a piecemeal fashion, c) social-interactionist approaches (Vygotsky, 1978) which emphasise the role of social interaction in language learning where children construct the ambient language through socially mediated interaction with linguistically knowledgeable adults, d) functional or usage-based approaches which assume that children's acquisition of language is driven by their desire to use language to perform communicative functions such as requesting something and making sense of their input (Tomasello, 2003), and e) nativist approaches (Chomsky, 1972) which assume that certain aspects of children's knowledge of language is present from the birth itself or is innate. Chomsky (1972) argues that the rapidity and uniformity in children's language acquisition point towards a species-specific, biologically endowed innate language faculty. This hypothesis is popularly known as the **innateness hypothesis**. The idea of an innate and highly organised language faculty is engaged with addressing the poverty of stimulus problem and aims to offer an account of the developmental stages in acquisition cross-linguistically. The innateness hypothesis suggests that infants are born ready to learn language, unless they have some impairment that hampers such learning (Guasti, 2002).

It is observed that languages vary quite a lot from each other and that the differences between them appear to be large and without limit. However, children are able to learn any language/languages to which they are exposed in their environment and are not limited in this ability by their race or family circumstances. Immigrant children learn the language of the country they grow up in without any difficulty. Despite the differences in the input that they are

exposed to, children attain the same level of competence in a short period of time. At about 10-12 months, they reach the one-word stage, between 20-24 months their utterances constitute two-word structures, and by 2-3 years, they have acquired a wide range of constructions involving considerable grammatical complexity (Guasti, 2002). All these facts point towards a species-specific neuro-physiological mechanism which enables the children to acquire their ambient language at such a fast pace within a short period of time without explicit teaching.

In order to explain this universal ability of humans to learn language (irrespective of their geographical location, socio-economic or educational status), Chomsky (1972) proposed a species-specific, innate language faculty which is termed as the Language Acquisition Device (LAD) or Universal Grammar (UG). According to Chomsky (1986, p. 3), UG is a "characterisation of the genetically determined language faculty. One may think of this as a 'language acquisition device', an innate component of the human mind that yields a particular language through interaction with presented experience, a device that converts experience into a system of knowledge attained: knowledge of one or another language".

This mentally represented grammar has different components: phonological component, lexicon, syntactic component, and the logical form component (Radford et al., 2009; Chomsky, 1986). The syntactic component generates legitimate structures which the phonological component converts into phonetic representations. Independently, the logical form component converts the structures generated by the syntactic component into interpretable semantic representations.

UG is a set of principles and constraints which defines the range of possible variations as well as invariant properties of languages and is assumed to aid language development. UG consists of two types of constraints: principles and parameters. Principles pertain to those properties of languages which are universal in nature, accounting for the similarities between human languages. For example, in any language, the interpretation of anaphors and pronouns is governed by common principles. Parameters determine those properties which are specific to an individual language. For example, the determination of the relative order between heads and their complements or whether subject-drop is permitted differs from language to language. In English, the verb precedes its object (e.g. *Ram loves Sita*) while in Malayalam, in the equivalent sentence, the verb follows the object (e.g. *ɾaman ʂiiṯaye ʂneehiykkuṉṉu* [ram sita love]). The child sets the parameters as appropriate to the language(s) being learned (Radford, 1997; Guasti, 2002).

Thus, according to the innateness hypothesis, children's innate knowledge of language, which is the mental grammar, consists of: a) knowledge of syntactic categories and rules for combining such categories into phrases and sentences, b) principles of language which are universal such as structure dependence, binding principles pertaining to the distribution of anaphors and pronouns etc., and c) parameters that vary across languages such as the head-direction parameter, null-subject parameter etc. While the universal principles which are invariant across languages are part of the child's language acquisition device and therefore do not have to be learned by the child, the language-specific parameters have to be learned by the child and have to be set according to the correct value (Radford, 1997; Ambridge & Lieven, 2011). Children are observed to set the parametric values appropriately according to their target language from very early ages onwards (Guasti, 2002). For example, while English acquiring kids produce utterances in which heads precede their complements, heads follow their complements in Japanese kids' utterances.

1.2 Inflections in Universal Grammar

Inflections are those morphemes that encode different kinds of grammatical information such as tense, aspect, number, gender, case and so on. They, typically, neither add lexical meaning to their stems nor change the grammatical category of their stems. For example, in the sentence *Mary danced*, the *-ed* affix places the activity as something that happened prior to the time of speaking (Fromkin, Rodman, & Hyams, 2011).

Haspelmath and Sims (2010, p. 18) define inflection/inflectional morphology as "the relationship between word-forms of a lexeme". A lexeme is an abstract concept of a word whereas word-forms are concrete words, composed of lexical meaning and grammatical functions, and can be phonetically realised. For example, WALK is a verb lexeme in English which represents the core-meaning shared by word-forms such as *walk*, *walk-s*, *walk-ed*, and *walk-ing* (Haspelmath & Sims, 2010). In the above examples, the inflection *-s* signals 3rd person singular and present tense, and *-ed* and *-ing* signal the past tense and the progressive aspect, respectively. The word-form *walk* is assumed to have a null-suffix (*walk-ø*) that signals present tense and 1st/2nd person singular or plural, or 3rd person plural depending on the subject. As Haspelmath and Sims (2010, p. 156) explain, "word-forms belonging to the same lexeme are paradigmatically related because they form a set of contrasting instantiations of the lexeme".

For example, the English word *bag* is identified as denoting the singular number because of the existence of the contrasting plural word-form *bags*. Similarly, the inflected forms *walkø*, *walks*, and *walked* form a paradigm of the tense category in English.

Inflectional morphemes signal a variety of grammatical information on different lexical categories. The typical inflectional features expressed on verbs are tense, aspect, mood, agreement, and number. Case, gender, and number are the common inflectional features expressed on nouns and adjectives. Inflectional features expressed on a particular lexical category can be combined as shown in Tables 1.1 and 1.2. Table 1.1 shows the tense forms for two aspects (infectum and perfectum) for two moods (indicative and subjunctive) in Latin and Table 1.2 shows the case marking details in Latin for singular and plural numbers.

Table 1.1 Tense, aspect, and mood marking in Latin.

	Indicative		**Subjunctive**	
	Infectum	**Perfectum**	**Infectum**	**Perfectum**
Present	canta-t	canta-v-it	cant-e-t	canta-v-eri-t
Past	canta-ba-t	canta-v-era-t	canta-re-t	canta-v-isse-t
Future	canta-bi-t	canta-v-eri-t		

(Haspelmath & Sims, 2010, p. 84)

Table 1.2 Case and number marking in Latin.

	Singular	**Plural**
Nominative	insula	insula
Accusative	insulam	insulās
Genitive	insulae	insulārum
Dative	insulae	insulīs
Ablative	insulā	insulīs

(Haspelmath & Sims, 2010, p. 83)

Tense markers signal the temporal location of an action/event (past, present, and future). Aspect signals the internal temporal constituency of an event as in whether an action/event is completed (perfective), non-completed (imperfective), ongoing (progressive) etc. The inflectional category of mood expresses a wide range of inflectional values such as imperative (com-

mand/request), subjunctive (non-realised) events, indicative (events viewed as objective facts), conditional (condition) etc. (Haspelmath & Sims, 2010).

Case expresses the syntactic (subject, object, indirect object etc.) or semantic (agent, patient, experiencer etc.) role of a noun/pronoun in a sentence. Nominative, accusative, dative, and genitive are the major cases found across languages. Typically, the nominative denotes the subject, accusative the direct object, dative the indirect object, and genitive *possession*. Number signals the quantity (singular, plural etc.). Gender assignments are determined by biological sex-differences, animate-inanimate distinctions, or are arbitrary (grammatical gender) (Haspelmath & Sims, 2010).

Inflections differ in their relative positioning within a word. While inflections that follow the base are called suffixes (e.g. the plural suffix *-s* in the English word *cat-s*), inflections that precede the base are called prefixes (e.g. *no-* in the Nahuatl word ***no****-cal* 'my house'). Infixes are those inflections that occur inside the base (e.g. ***-um-*** in the Tagalog word *s-um-ulat* 'write'). Finally, there are certain inflections that occur on both sides of the base. These are called circumfixes (e.g. ***ge- -en*** in the German word *ge-fahr-en* 'driven') (Haspelmath & Sims, 2010).

Languages vary in their propensities when it comes to inflectional assignment. Some languages are predominantly prefixing (e.g. Navajo, Hunde etc.) where the affixes precede their stems whereas some languages are predominantly suffixing where affixes follow their stems (e.g. Dravidian languages, West-Greenlandic languages etc.). Another major typological difference is between analytic languages and synthetic languages. The former class has little inflectional morphology with a lower morpheme-to-word ratio (e.g. Mandarin Chinese, English etc.) while the latter has rich inflectional morphology with a higher morpheme-to-word ratio. The latter class includes agglutinative languages with highly regular inflections showing a one-to-one correspondence between morpheme and meaning (e.g. Malayalam, Turkish etc.) and fusional languages where a single morpheme signals multiple grammatical features (e.g. Spanish, Hebrew etc.) (Haspelmath & Sims, 2010; Dryer & Haspelmath, 2013). Such distinctions employed by their target languages are seen to have an effect on the children acquiring those systems. For example, the relatively late acquisition of the sparse inflectional morphology of English is attributed to the sparse morphology of English itself (Hyams, 2008). Children born into languages with rich inflectional systems tend to acquire their inflectional morphology relatively earlier when compared to their counterparts born into languages that have sparse inflectional

morphology. Especially, agglutinative systems are seen to facilitate inflectional development since inflections belonging to such systems are typically syllabic (phonologically transparent) and can be easily separated from the base (morphologically transparent) owing to their typology (Dressler, 2012).

Inflections are of particular interest and significance in linguistics because of their positioning at the interfaces between phonology, morphology, and syntax. While inflections are functional morphemes that add grammatical information to the lexical roots, and thus within the purview of the morphological domain, their phonetic realisation is governed by the phonotactic constraints of a language. Consequently, their acquisition and production will also be constrained by the phonological rules of the target language. Similarly, the assignment of inflectional markers like case, tense, agreement etc. is governed by the syntactic component which influences the acquisition of such morphemes (Penke, 2012). In short, the patterns of inflectional development in a particular language yield a better understanding of the grammatical processes at play in that language.

1.3 Dravidian Languages

The Dravidian language family is the fifth largest language family (Subrahmanyam, 2006). Dravidian languages are spoken primarily in the four South-Indian states Andhra Pradesh, Karnataka, Tamil Nadu, and Kerala in addition to the pockets in Maharashtra, Madhya Pradesh, Orissa, Bihar, as well as Pakistan. The languages that belong to the North Dravidian sub-group are Kurux, Malto, and Brahui. The Central Dravidian sub-group comprises the languages Kolami, Naikri, Naiki, Parji, Ollari, and Gadaba. While the South-Central family consists of Telugu, Gondi (several dialects), Konda, Kui, Kuvi, Pengo, and Manda, the South-Dravidian family includes Irula, Kurumba, Kodagu, Toda, Kota, Badaga, Kannada, Koraga, Tulu, Tamil, and Malayalam (Krishnamurti, 2003).

As per the *Census of India* (2011), the Dravidian languages have 237,840,116 speakers. Figure 1.1 shows the languages belonging to the four main sub-groups of the Dravidian language family based on their geographical distribution. The broken lines in the figure represent uncertain relationships. The four main sub-groups are North Dravidian, Central Dravidian, South-Central Dravidian, and South Dravidian (Krishnamurti, 2003).

Tamil, Telugu, Kannada, and Malayalam are the four major literary languages in the fam-

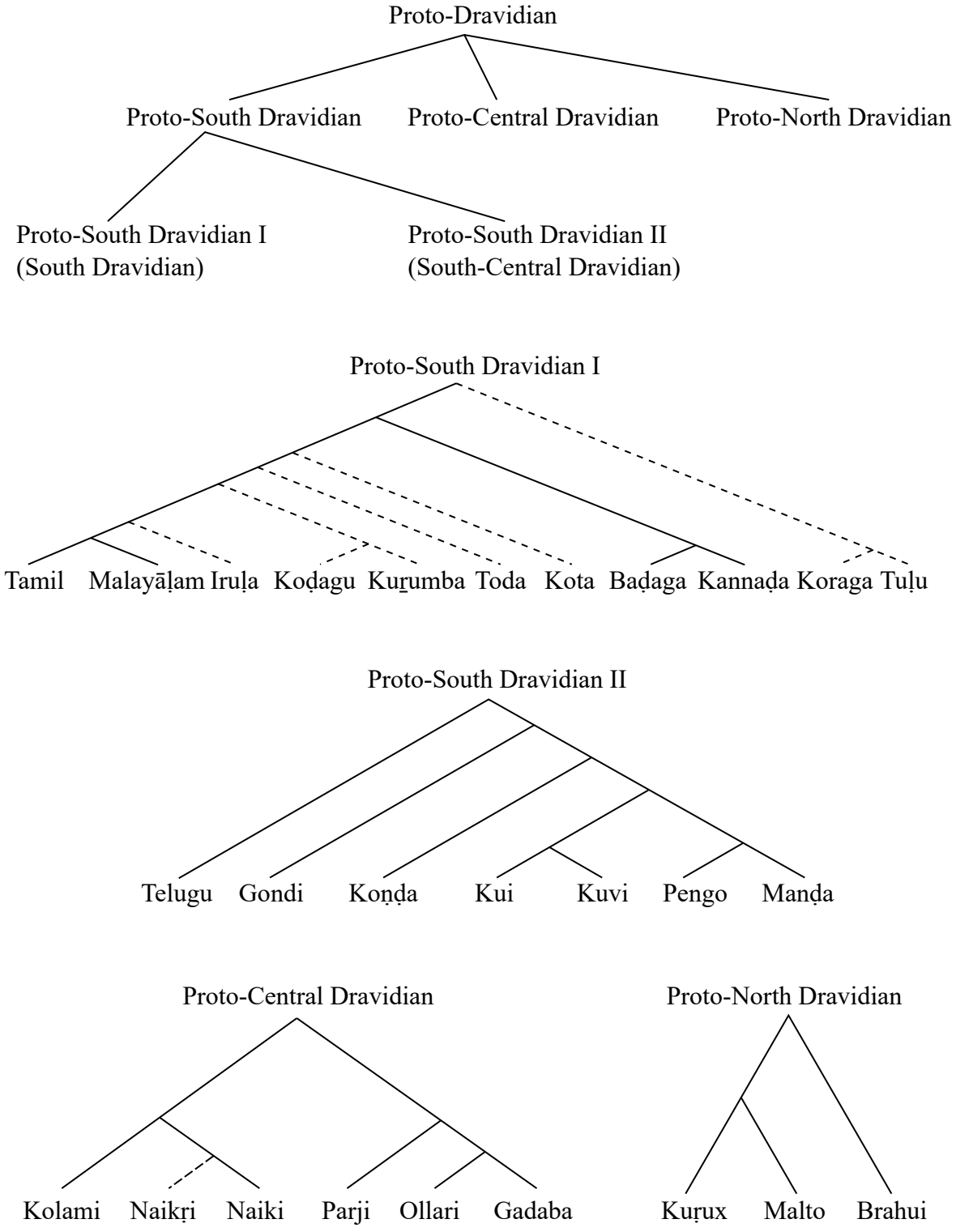


Figure 1.1 Family tree of Dravidian languages. (Krishnamurti, 2003, p. 21)

ily and are also mapped to the four states (state languages). The ancient texts *Tolkappiyam* (5 BC), *Kavirajamarga*, (9 AD), translation of the *Mahabharata* (11 AD), and *Ramacharitham* (12 AD) are the first literary texts in Tamil, Kannada, Telugu, and Malayalam, respectively. Among

the Dravidian languages, Tamil exhibits diglossia. Dravidian languages are predominantly suffixing, agglutinative languages that have post-positions and an SOV word-order (Krishnamurti, 2003). Table 1.3 shows the typological and sociolinguistic features of the Dravidian language family.

Table 1.3 Typological and sociolinguistic features of Dravidian languages.

Speakers	237,840,116
Spoken in	Southern-India, Maharashtra, Madhya Pradesh, Orissa, Bihar, Pakistan
Sub-Groups	North Dravidian, Central Dravidian, South-Central Dravidian, and South Dravidian
Proto language	Proto-Dravidian
Ancient literary texts	*Tolkappiyam* (5 BC, Tamil), *Kavirajamarga* (9 AD, Kannada), translation of the *Mahabharata* (11 AD, Telugu), *Ramacharitham* (12 AD, Malayalam)
Diglossic Status	Tamil
Typology	agglutinative, SOV, post-positions, predominantly suffixing

(Krishnamurti, 2003)

1.3.1 Grammatical Features of Dravidian Languages

Most Dravidian languages have a 10 vowel system with five short and five long ones similar to Proto-Dravidian. Table 1.4 shows the vowel inventory in Proto-Dravidian. The consonant system of Proto-Dravidian is characterised by the presence of retroflex consonants comprising stops, nasals, laterals, and approximants. The retroflex approximant is the most distinguishing phoneme in the entire phonological inventory. Proto-Dravidian has six stops with a rare three-way contrast between the dental, retroflex, and alveolar, which is retained in the cognate languages Malayalam, Irula, Kota, Toda, and Kurumba. Proto-Dravidian lacks voiced and aspirated stops (Subrahmanyam, 2006). Table 1.5 shows the consonant system of Proto-Dravidian.

Proto-Dravidian has an agglutinative morphology. The major identifiable categories in Dravidian languages are nouns, pronouns, numerals, adjectives, adverbs, postpositions, parti-

Table 1.4 Vowels of Proto-Dravidian.

	Front		**Central**		**Back**	
	Short	**Long**	**Short**	**Long**	**Short**	**Long**
High	i	ī			u	ū
Mid	e	ē			o	ō
Low			a	ā		

(Subrahmanyam, 2006, p. 786)

Table 1.5 Consonants of Proto-Dravidian.

	Labial	**Dental**	**Alveolar**	**Retroflex**	**Palatal**	**Velar**
Stop	p	t̪	t	ʈ	c	k
Nasal	m	n̪	n	ɳ	ɲ	ŋ
Lateral			l	ɭ		
Trill			r			
Approximant				ɻ		
Semivowel	ʋ				y	

(Subrahmanyam, 2006, p. 787)

cles, and interjections. Nominals comprise nouns, pronouns, and numerals since all of them are inflected for case. Nouns in Proto-Dravidian predominantly inflect for case and number. Pronouns in Proto-Dravidian are characterised by the distinctions between exclusivity vs inclusivity in 1PL and degrees of proximity in 3SG and 3PL. These distinctions are retained in cognate languages like Malayalam, Tamil, and Telugu while Modern Kannada has lost this distinction. For example, in Malayalam, the 1PL pronoun *ɲaŋŋaɭ* signals exclusion of the hearer whereas the *n̪ammaɭ* signals inclusion of the hearer. Also, the Malayalam 3SG pronouns *avan* and *ivan* both meaning 'he' refer to someone who is a remote entity and a proximal entity, respectively. Another feature is the presence of the reflexive pronoun *t̪aan* referring to the third person subject (Subrahmanyam, 2006).

Finite verbs in Dravidian languages typically inflect for tense and person (agreement). Most languages have a three-tense system while certain languages like Kannada only make a distinction between the past and non-past. While Modern Malayalam has lost its person suffixes,

Toda and Kodagu have no person suffixes in the third person. Pro-drop is a common feature of the Dravidian languages owing to their rich inflectional system (Subrahmanyam, 2006). However, in Malayalam, instead of pro-drop, there is topic drop where both subjects and objects can be elided if they can be retrieved from the context.

The characteristic syntactic constructions in Dravidian languages are a) equational sentences, b) dative subject constructions, and c) complex sentences having more than one participle. Most Dravidian languages have equational sentences without the copula, as shown in the Tamil sentence (1). However, Malayalam and many of the central and northern languages have introduced the copula to the equative constructions under the influence of Indo-Aryan languages as shown in the Malayalam sentence (2) (Subrahmanyam, 2006).

(1) en peyar ʂiiʈa
my name sita
'My name is Sita.'

(2) ente peerɨ ʂiiʈa eɳɳɨ aaɳɨ
my name sita COMP be.PRS
'My name is Sita.'

Another common construction that occurs in the language family is the dative subject construction. Although the subject is typically assigned the nominative case in these languages, certain predicates lexically assign the dative case to the nominals occurring in the subject position as shown in the Telugu example (3).

(3) maadhuri-ki anil-miida preemai kaligin-dii
madhuri.3SG.M-DAT Anil.3SG.M-on love.3SG.NM occurred-3SG.M
'Madhuri fell in love with Anil.'

(Subbarao & Bhaskararao., 2004, p. 162)

While all Dravidian languages show the DAT-NOM pattern, Tamil and Malayalam show deviation from this pattern by exhibiting a DAT-ACC pattern as shown in the Malayalam sentence (4).

(4) ʂiiʈay-kkɨ kiɭiy-e ʋeeɳam
sita-DAT bird-ACC want
'Sita wants the bird.'

Additionally, Dravidian languages also have constructions involving more than one participle, typically signalling the sequentiality of events as shown in the Kannada sentence (5).

(5)

ʋanaɟa	mane-ge	hoog-i	snaana	maaɖ-i	baʈʈe	badalaayis-i
vanaja	house-DAT	go-PST	bath	do-PST	clothes	change-PST
uuʈa	maaɖ-id-aɭu					
food	do-PST-3F.SG					

'Vanaja went home, took a bath, changed her clothes and ate food.'

(Subrahmanyam, 2006, p. 789)

1.4 Malayalam and its Linguistic Features

Malayalam is one of the four major Dravidian languages. It is a statutory provincial language spoken in the state of Kerala and in the union territories Lakshadweep and Puducherry, belonging to the Indian Union, by 34,800,000 speakers as their first language (Eberhard et al., 2019). Table 1.6 presents the characteristic features of Malayalam.

Table 1.6 Features of Malayalam.

Speakers	34,800,000
Spoken in	Kerala, Lakshadweep, Puducherry
Word-order	SOV
Typology	head-final, agglutinative
Case-marking	7 cases
Gender	masculine, feminine, and neuter
Number	singular and plural
Tense	past, present, and future
Phonemes	37 consonants, 11 vowels, 4 diphthongs

(Eberhard et al., 2019)

Malayalam evolved as a separate language from the west-coast dialect of Tamil around the 9th century (Krishnamurti, 2003). Some of the major differences between Malayalam and Tamil lie in a) the presence of the off-glide in words, b) nasal spread, c) marking agreement on verbs, d) pro-drop vs topic-drop, e) accusative marking on objects, and f) having morphological

vs periphrastic causatives. However, one particular similarity that distinguishes Malayalam and Tamil from other Dravidian languages and most Indo-Aryan languages is the dative-accusative pattern existing in these languages whereas the others exhibit a dative-nominative pattern.

Malayalam lost the off-glide in words like *vila* 'price', whereas in Tamil the corresponding words have the forms with the off-glide as in *vilay* 'price' (Rajaraja Varma, 1896). Table 1.7 shows the difference between Tamil and Malayalam in the distribution of the off-glide. Another major phonological difference between the two languages is in the nasalisation of the stop consonant in adjacent nasal - non-nasal consonant pairs. In Tamil, there is no nasal spread in words containing such adjacent pairs as can be seen in the example, *maa**ŋk**aay* 'mango'. However, in Malayalam, the nasal spreads to the stop as can be seen in the corresponding equivalent, *maaŋŋa* 'mango', (Rajaraja Varma, 1896). Further examples of equivalent words in the two languages, exhibiting such difference in the nasalisation of the adjacent non-nasal consonant, are given in Table 1.8.

Table 1.7 Difference between Tamil and Malayalam in the distribution of the off-glide.

Tamil	**Malayalam**	**Gloss**
t̪alay	t̪ala	'head'
maɻay	maɻa	'rain'
ilay	ila	'leaf'
kaʋit̪ay	kaʋit̪a	'poem'
ʋaaɻay	ʋaaɻa	'plantain'

Table 1.8 Difference between Tamil and Malayalam in the nasalisation of stop consonants.

Tamil	**Malayalam**	**Gloss**
manɟaɭ	maŋŋaɭ	'turmeric'
nanri	n̪an̪n̪i	'gratitude'
erinɟ**ɨ**	eriŋŋ**ɨ**	[throw.PTCP]
ʋan̪t̪**i**	ʋan̪n̪**i**	[come.PTCP]
t̪uɻaŋgi	t̪uɻaŋŋi	[begin.PTCP]

Malayalam which originally had agreement inflections on the verbs lost them over the course of time. Table 1.9 shows the differences in agreement between Tamil and Malayalam

verb forms. Rajaraja Varma (1896) proposes that the loss of agreement inflections might have resulted from the Brahminical ascendancy in Kerala around the 14th century. He speculates that the Brahmins who spoke the Indo-Aryan Sanskrit and Prakrits must have had difficulties with the Tamil agreement system because in Sanskrit the verb agrees with the subject only in number while in Tamil the verb agrees with the subject both in gender and number. This seems to be possible since Brahmins had a tremendous influence over the societal structure of Kerala during that period. This is most visible in the poetry tradition where the major poets of 14th - 17th centuries resorted to the style called *manipravala* which is a version of Malayalam with a heavy admixture of Sanskrit. During that period, there existed a visible demarcation between two lines of poetry: a) the Sanskritised *manipravala* poetry which was mostly the creation of Brahmin poets and b) *paaṯṯu* which follows the Tamil-Dravidian tradition (Rajaraja Varma, 1896). Agreement forms on verbs are visible in the first Malayalam literary text *Ramacharitham* (12 AD) and the later text *Adhyatma Ramayanam Kilippaaṯṯu*, assumed to be written around the second half of the 16th century or the beginning of the 17th century by Ezhuthachan (Prabodhachandran Nayar, 1985). Both these texts belong to the Tamil-Dravidian *paaṯṯu* tradition (Rajaraja Varma, 1896). Also, the popularity of *manipravala* led to the heavy influx of Sanskritised words into Malayalam vocabulary. Compared to Tamil, Malayalam has inordinately borrowed words from Sanskrit.

Table 1.9 Differences in agreement between Tamil and Malayalam verb forms.

Tamil	**Malayalam**	**Gloss**
avan vaṉṯ **aan**	avan vaṉṉu	he came
ava̱ vaṉṯ **aa̱**	ava̱ vaṉṉu	she came
avar vaṉṯ **aar**	avar vaṉṉu	they came
ṉii vaṉṯ **aay**	ṉii vaṉṉu	you (SG) came
ṉiiŋga̱ vaṉṯ **iirka̱**	ṉiŋŋa̱ vaṉṉu	you (PL) came
naan vaṉṯ **een**	ɲaan vaṉṉu	I came
naaŋga̱ vaṉṯ **oom**	ɲaŋŋa̱ vaṉṉu	we came

(Rajaraja Varma, 1896)

Associated with the presence vs lack of agreement in Tamil and Malayalam, respectively is the kind of null-subjects each language permits. Tamil permits pro-drop because the subject can be understood from the phi-feature marking on the verb. Since Malayalam lost its agreement

inflections on the verb, from a pro-drop language it became a topic drop language where both the arguments (subjects and objects) can be elided if they can be understood from the context.

Another major difference between Tamil and Malayalam is in the assignment of the accusative marker on direct objects. While Tamil optionally drops the accusative marker on objects (6) (Sarma, 2014), Malayalam overtly marks the accusative only on [+ANIMATE] objects (7) and obligatorily leaves the [−ANIMATE] unmarked (8).

(6) ena-kku ellaam-**ø** ʈeriy-um
I-DAT everything know-3SG
'I know everything.'

(Sarma, 2014, p. 115)

(7) ɲaan ʂiiʈay-kkɨ kiɭiy-e koʈu-ʈʈu
I.NOM sita-DAT bird-ACC give-PST
'I gave Sita a bird.'

(8) ɲaan ʂiiʈay-kkɨ miʈʈaayi-**ø** koʈu-ʈʈu
I.NOM sita-DAT candy give-PST
'I gave Sita a candy.'

Malayalam has morphological causatives as shown in (9) and (10). For example, a verb can be made causative by the addition of a causative suffix to a transitive verb stem. For example, *kaɻiykk-* 'eat' becomes *kaɻi**pp**iykk* 'make eat' by the addition of *-pp*. Whereas Modern Tamil has periphrastic causatives which are formed by combining the causative verbs, *ceyya* 'do', *paɳɳa* 'make', or *vaikka* 'put', with the infinitival verb form (Nizar, 2010). The sentences (11) and (12) show the causative formation in Modern Tamil.

(9) kumaar ʋa-ɲɲu
kumar.NOM come-PST
'Kumar came.'

(10) ɾaaɻa kumaari-ne ʋaru-ʈʈ-i
raja.NOM kumar-ACC come-CAUS-PST
'Raja made Kumar come.'

(11) kumaar ʋa-ɲʈ-aan
kumar.NOM come-PST-3SG.M
'Kumar came.'

(Lehman, 1989, as cited in Nizar, 2010, p. 27)

(12) ɾaaɹa kumaar-ai ʋara ʋai-t̪t̪-aan
raja.NOM kumar-ACC come make-PST-3SG.M
'Raja made Kumar come.'
(Lehman, 1989, as cited in Nizar, 2010, p. 27)

Like the other Dravidian languages, Malayalam is an SOV, head-final language and has postpositions. It has 37 consonants, 11 vowels, and 4 diphthongs in its phonological inventory. As mentioned earlier in the previous section, it has inclusive/exclusive pronouns as well (Eberhard et al., 2019). Malayalam is an agglutinative language with rich inflectional morphology. It has a three-way tense marking system. Verbs are inflected to signal the present, past, and future. Malayalam has three genders (masculine/feminine/neuter), seven cases (nominative, accusative, dative, sociative, genitive, locative, and instrumental), and two numbers (singular and plural). Malayalam inflections are phonologically salient, morphologically transparent, exhibit biuniqueness and are productive in general. These are factors that aid the inflectional development in children (Dressler, 2012). Thus, early learners of Malayalam are already in a system that is designed to facilitate their inflectional development.

1.5 Organisation of the Thesis

Nouns and verbs are the two lexical categories present in every human language. They are also the two major categories on which languages signal grammatical information using inflectional markers. As mentioned in the previous section, Malayalam is an agglutinative language with a rich inflectional system. Thus, the acquisition of Malayalam inflections is in principle interesting because of its typological properties since children born into agglutinative systems are observed to acquire the inflections relatively earlier compared to their counterparts who are born into languages with sparse inflectional systems, as mentioned earlier in Section 1.2. In this thesis we look at the nominal and verbal inflections in Malayalam, especially how these systems are acquired. The thesis will pay particular attention to case, gender, and number (CGN) features in nouns and tense, aspect, and mood (TAM) inflections in verbs. There are two important questions in the description of Malayalam inflectional grammar that have received a lot of consideration (Asher, 1969; Prabodhachandran Nayar, 1972; K. P. Mohanan & Mohanan, 1990; Jayaseelan, 2004) but, in my opinion, have been incompletely accounted for: a) the selection of the past tense marker by verbs, and b) the non-canonical case-marking on nouns in subject

positions.

The previous analyses of Malayalam past tense formation centre around classifying Malayalam verbs into morphological classes on the basis of the number of past tense allomorphs that are identified in the individual analyses. Kunjan Pillai (1965) proposes 16 classes in Malayalam to account for the differing surface forms based on the surface structure of the verb stems. Wickremasinghe and Menon (1927) propose eight classes, Sekhar and Glazov (1961, as cited in Asher & Kumari, 1997) twelve, Asher (1969) four, Prabodhachandran Nayar (1972) four, and Valentine (1976, as cited in Asher & Kumari, 1997) two. All these accounts attempt to predict the phonetic realisation of the marker depending on the nature of the stem and divide them into classes depending on the past tense marker they choose, without explaining why a particular marker is assigned to a particular class of stems in the first place, i.e, before the realisation of the surface forms. Therefore, in these accounts, the assignment of the marker remains arbitrary and thus, they are not generative accounts. This is problematic from a language acquisition point of view. A child will not know what the past tense marker may be unless he/she hears it in use. This runs counter to the predictive power of grammatical rules and their use in developing grammars. In this thesis, we will attempt to provide a unifying account of Malayalam past tense formation by looking at why a particular marker occurs with a particular class. Instead of grouping verb stems into different classes based on the past tense marker they select, we focus on the morpho-phonological constraints which determine the selection and instantiation of a specific past tense allomorph.

K. P. Mohanan and Mohanan (1990) attempt to provide a unified account of the dative nominals occurring in the subject and non-subject positions by subsuming them under the semantic roles *goal* and *possession*. However, later accounts (Nizar, 2010) have shown that K. P. Mohanan and Mohanan's (1990) proposal does not provide a complete account of the behaviour of Malayalam dative subjects. Jayaseelan (2004) argues that dative NPs are obliques and that there is always a nominative NP (which is either overt or covert) in constructions involving dative NPs. However, this account does not take into account the resulting valency change if there were covert nominative NPs in dative subject constructions. Further, none of the binding or control facts run counter to the treatment of the dative nominal as a subject. Nizar (2010) demonstrates that the behaviour of experiencer predicates in different syntactic contexts underscores the subject properties of dative nominals. However, she does not extend her analysis to modals which assign a lexical dative case in the subject position.

We argue that the dative nominals occurring in the subject positions in Malayalam sentences are indeed subjects by analysing their behaviour in various syntactic contexts. We use Ura's (2000) Agr-less Checking theory to account for the split-subject properties of dative subjects and to show how the dative nominals enter into various syntactic relations such as EPP, phi-feature checking etc., with T. The overt dative case assignment corresponds to the different theta roles selected by individual predicates and instantiates inherent case selection since the case is selected exceptionlessly.

Since these two inflectional assignments (past tense formation and dative subject marking) are grammatically complex, their acquisition is of particular interest and can provide insights into the acquisition of inflections in general. We employ the acquisition data to a) substantiate the analysis, and b) demonstrate the various effects predicted by such an analysis on language development itself, i.e., on the observable acquisition patterns of past tense forms and dative subjects. We will also show how the existing theoretical models fare in terms of the Malayalam acquisition data and finally, attempt to explain the observed patterns based on the conclusions offered by the analyses.

The thesis is organised as follows. Chapter 2 discusses the main theoretical models that try to explain the cross-linguistically observed patterns in inflectional acquisition such as the dual-route (Pinker & Prince, 1994) vs single route models (Rumelhart & McClelland, 1986), small clause (Radford, 1990) vs full competence hypothesis (Poeppel & Wexler, 1993), truncation model (Rizzi, 1993), Agreement/Tense Omission Model (Schütze & Wexler, 1996), and the prosodic constraints model (Demuth, 1996, 2001). The various factors that affect the acquisition of inflectional markers such as phonological salience, morphological and semantic transparency, frequency, etc. (Dressler, 2012) and the stem-parameter theory (Hyams, 2008), which addresses the cross-linguistic variation in inflectional development as a result of the parameter setting in children's grammar rather than affix-learning determined by various factors are also discussed in this chapter.

Chapter 3 discusses the current acquisition study, the subjects recorded, and the methods adopted for data analysis. The present study focuses primarily on the longitudinal data which comprise the spontaneous speech production of two monolingual Malayalam speaking children, a girl (1;9-2;10) and a boy (2;3 - 3;0). This chapter presents a developmental sketch of the subjects. It also discusses productivity in inflectional development and explains the criteria employed in the present study in order to determine whether an inflectional morpheme is

productively used.

Chapter 4 presents an overview of Malayalam inflections focusing on the agglutinative features, phonotactic constraints, and the selection of the affix alternants of a single inflectional feature and their impact on acquisition. It demonstrates the properties of Malayalam inflections such as phonological salience, biuniqueness, morphological transparency etc. owing to the agglutinative typology of the language. Further, it discusses the role of Malayalam phonotactic constraints such as the coda constraint, which allows only selected phonemes to occupy the coda position of the syllables, and the syllabic weight constraint, which does not permit the word-initial and word-medial syllables to have more than two morae, in determining the selection of an inflectional marker as well as the formation of different surface structures of inflectional markers.

We show that Malayalam has two kinds of inflectional selection: a) phonological selection, where inflections are assigned to signal a grammatical category in such a way that the inflectional morphemes conform to the phonotactics of the language (e.g. past tense marker selection in underived verbs), and b) morphological selection, where inflections are assigned based on the morphological features such as [±TRANSITIVE] (e.g. past tense marker selection in derived verbs) or [±ANIMATE] (e.g. accusative marker selection). Children are already seen to have an early awareness of the phonological systems of their native language as demonstrated by Jusczyk et al. (1993) and Christophe et al. (1994). They exploit information related to the distributional regularities (statistical information on the sequencing of sounds), typical word shapes, and phonotactic constraints in order to build their mental lexicon (Guasti, 2002). Thus, it is only logical to assume that phonologically determined inflectional assignment will be relatively easier to acquire than those inflectional assignments that go beyond phonology since children are seen to be equipped with the ability to imbibe their target phonotactics even before they begin their speech production.

Chapter 5 looks at the various TAM inflections present in the children's speech in detail except for the past tense morphemes. We see that all the TAM inflections except the past tense markers in the derived verbs are assigned across the board without exceptions. We also see that there are no omission errors involving any of the TAM infections and in fact, most of the markers are productively used by the children.

Chapter 6 focuses specifically on the past tense formation in Malayalam and examines its acquisition. We demonstrate that there are two past tense markers: *-i* and *Tu*, with the latter

realising as two variants *-ṯu* and *-ṉṯu* in predictable environments. In the underived stems, the syllabic weight constraint determines the past tense marker selection. Thus, in heavy stems, the marker *-i* is selected (eg. *maar-* + *-i* – *maari* 'removed') in order to resyllabify the stem-coda to the onset position in the inflected forms.

In Malayalam, verbs are either derived through sound change or affixation. In the former class, sound change is employed to effect valence change such as intransitive to transitive (e.g. *iɭak-* 'stir$_{intr}$' → *iɭa**kk**-* 'stir$_{tr}$'). Such valence increased verbs choose the past tense marker *-i* since the stems in question are phonologically heavy; consequently, the marker *-i* is selected (e.g. *iɭa**kki*** 'stirred$_{tr}$) in order to adhere to the syllabic weight constraint. In the latter class, the transitive verbs are derived through *-kk* affixation to the nominal stems (eg. *kuɻi* 'pit' + *-kk* → *kuɻiykk-* 'dig') and intransitive verbs are derived through zero-affixation of the nominal stems (e.g. *kuɻi* 'pit' + *ø* → *kuɻiy-* 'crumble'). The resultant transitive verbs select the past tense marker *-ṯu* and the intransitives, *-ṉṯu*. However, these derivation rules are not synchronically productive in Modern Malayalam.

Children's past tense marking data show that they do not have any issues in correctly assigning the past tense markers that are phonologically selected, i.e, the ones involving the underived verbs and the ones derived through sound change. While there are no omission errors involving the past tense markers, the synchronic arbitrariness of the derivation rules poses problems in the acquisition. The overgeneralisation errors show that children prefer the phonotactically driven inflectional selection and have difficulty in acquiring the past tense marking depending on the feature [±TRANSITIVE], especially since the latter is synchronically arbitrary.

Chapter 7 looks at each of the case, gender, and number inflections present in the children's transcripts, except the dative subject marking. We see that all the nominal inflections, except the accusative, gender markers, and plural markers, are phonologically selected. The children's data show that while most inflections are productively used with gender and number markers being notable exceptions, the accusative morpheme, *-e*, whose overt assignment is determined by the [±ANIMATE] feature, is omitted the most number of times with only 85% correct instances. Accusative has the lowest percentage of correct instances compared to the other markers in production except for the [+HUMAN] plural marker *-maar* (only 25% correct instances) in the boy child's transcripts. The accusative marker omission is seen both in the cross-linguistic data as well as the English-Malayalam bilingual acquisition data (Raghunathan,

Unpublished data).

Chapter 8 presents an analysis of the dative subject constructions in Malayalam and the acquisition and use of dative subjects in the children's transcripts. Malayalam dative subject predicates can be divided into three classes (Nizar, 2010). Class I predicates include certain light verbs like *ʈoonn*- 'feel', *var*- 'come', *patt*- 'be able' that are mostly N + V constructions along with the experiencer predicates such as *viɛakk*- 'be hungry', *ɖaahikk*- 'be thirsty' etc. Class II predicates include the modals *-aam* and *-aɳam*. Class III, the copula *uɳʈi* 'have' assigns the dative case to the nominals in the subject position to signal *possession*. We demonstrate that the typical subject properties of anaphor binding, control, agreement in equative constructions, nominalisation, as well as verbal properties such as tense and aspect marking, case marking and theta selection collectively confirm the subjecthood of these atypical subjects. To explain how the syntactic licensing of features work, we adopt the "split" function framework proposed by Chomsky (1995) and refined by Ura (2000). As per Ura's (2000) proposal of the splitting of GFs, the properties of control into adjuncts and binding of a (purely) subject-oriented reflexive result from phi and EPP feature checking relations with T, respectively. The agreement between a subject and its verb also involves phi-feature checking with T. We see that the dative nominals in Malayalam have the ability to have control into adjuncts and bind a purely subject-oriented reflexive. These properties demonstrate the phi-feature checking relations between the dative subject and T, which further establish the subject properties of these dative nominals.

The acquisition data show that children exhibit a high degree of competence with the use of dative subjects. Their production of dative subjects is in accord with a) the lexical properties of the particular predicates used and b) the binding and PRO-construal restrictions of Malayalam. Also, they clearly make a distinction between the dative marked subjects and indirect objects. This suggests a high degree of correlation between the adult and child grammars. The overgeneralisation errors reveal that dative marking rule targets the subject position just like the nominatives and bring clarity to the fact that both the nominative and dative NPs are treated on par as potential subject cases. The acquisition pattern of early Malayalam dative subjects show the same trend as that of early dative subjects in Tamil (Lakshmi Bai, 2004) and Telugu (Usha Rani & Sailaja, 2004). The omission errors involving the datives (both subjects and objects) are ascribed to the high formal complexity and the lack of biuniqueness of the marker itself, since the dative case is assigned across grammatical and semantic roles in Malayalam. Chapter 9 presents a summary of the results and also addresses the broader implications of the

study, the directions for future work, and open questions that still remain to be addressed.

1.6 Contributions of the Thesis

A majority of studies on early language acquisition centres on the dominant languages of the Indo-European family. Consequently, the theoretical models posited to account for inflectional development is mostly suited for Indo-European acquisition patterns (Ambridge & Lieven, 2011). This is not an ideal situation since the notion of Universal Grammar is designed to explain the properties of all the languages spoken by human beings. Mateo Pedro (2015) in his study on Q'anjob'al inflectional development notes that while many theories have been posited to explain the omission of inflectional markers, the missing pieces of information have to come from non-European languages with rich inflectional systems. It is only then that the nature of UG can be understood completely. Data from non-European languages will help in developing the theory of Universal Grammar by offering new insights into the existing theoretical framework. My thesis focuses on the inflectional development in Malayalam, which has a rich inflectional system. Acquisition studies are rare in Dravidian languages (Raghavendra & Leonard, 1989; Sarma, 1999, 2014; Usha Rani & Sailaja, 2004) and on Malayalam in particular (Girija Devi, 1972). This thesis aims to contribute to a) work on Dravidian languages in general and Malayalam in particular, b) understanding the acquisition of non-canonical subjects, and c) acquisition studies in general, by identifying *mode of selection* as an additional factor that affects the inflectional development in children.

Chapter 2

Contextualising Inflectional Development

2.1 Acquiring Inflections

One of the major tasks that confronts children is vocabulary acquisition, that is, learning the lexical and functional words of their target language and building a mental lexicon with the requisite phonological, morpho-syntactic and semantic properties. Children need to figure out the meanings of lexical and functional items in order to make sense of the whole sentential structure. This thesis looks at the acquisition of an important functional category: inflections.

Cross-linguistically, children are seen to omit inflectional morphemes in their early sentences. This phase is termed as the **telegraphic stage** (Brown, 1973). Another significant feature of language development that has been observed is children's use of non-finite forms in matrix clauses rather than finite forms. This is called the **root infinitive (RI) stage** (Rizzi, 1993) and is observed in some languages like English, French and Dutch etc. but not in Italian or Spanish.

As explained earlier in Chapter 1, inflections are situated at the interfaces between phonology, morphology, and syntax. Thus, it is only natural that different models have been proposed to account for the cross-linguistic patterns in inflectional acquisition addressing different levels of grammar which are shown in Figure 2.1[1]. In this chapter, we will discuss the major genera-

[1]We are not looking at the lexeme-learning models such as syntactic bootstrapping (Gleitman, 1990), and semantic bootstrapping (Pinker, 1984). However, we will be discussing phonological bootstrapping (Gleitman & Wanner., 1982) in order to explain the premises of the prosodic-constraints model (Demuth, 1996, 2001) which is a phonological model. Also, we only discuss the models that fit the current Malayalam data in this chapter. For example, the Variational Learning Model (Yang, 2002) which assumes that children have access to a set of UG grammars in the initial stage which are then fixed according to the input data does not explain the omission errors in the current data since the input has inflected forms in plenty which is enough to set the parameter right. Also, we do not discuss the Imperative Analogue Hypothesis (Salustri & Hyams, 2003) since the current Malayalam data

tive models on inflectional acquisition such as **the prosodic constraints model** (Demuth, 1996, 2001), which is a phonological model, the **dual-route** (Pinker & Prince, 1994) and **single-route** (Rumelhart & McClelland, 1986) models which are morphological models (the latter is the only non-generative model discussed since the major debate in morphological acquisition is between the dual-route and the single-route models), and several syntactic models such as **the small-clause hypothesis** (Radford, 1990), **full-competence hypothesis** (Poeppel & Wexler, 1993), **the truncation model** (Rizzi, 1993) and **the Agreement/Tense Omission Model** (ATOM) (Schütze & Wexler, 1996) which are also full-competence models.

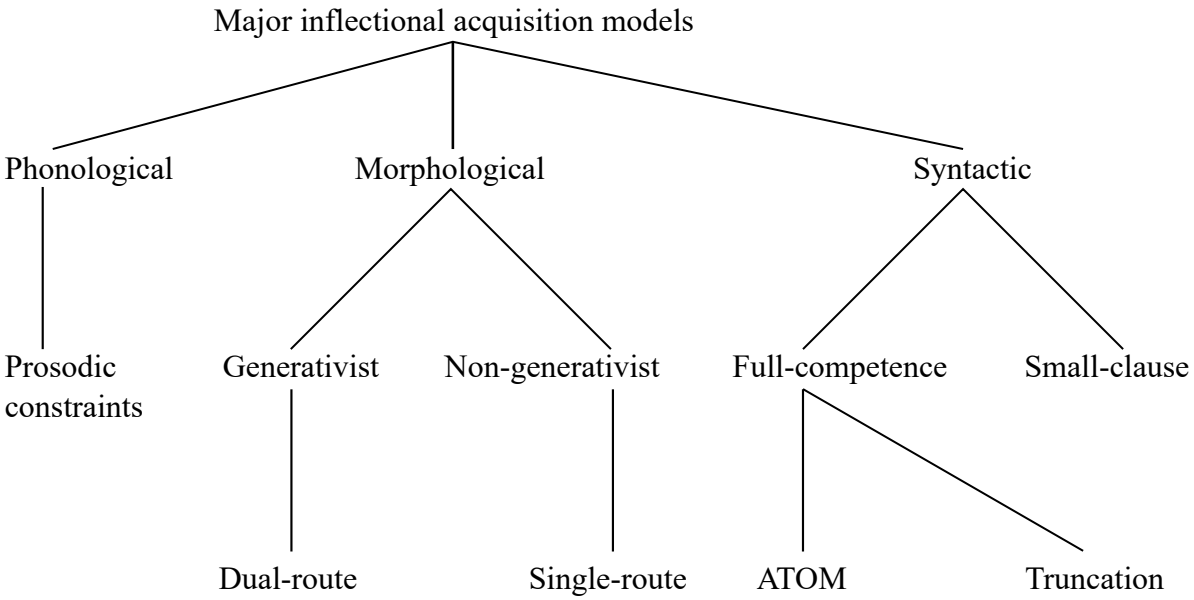


Figure 2.1 Major inflectional acquisition models.

Further, the factors that determine the relative ease or difficulty in affix-learning will be discussed in this chapter. Along with the factors that affect the early inflectional development, we will also discuss the stem-parameter theory (Hyams, 2008) that analyses the lead-lag patterns observed in crosslinguistic inflectional acquisition as a result of the difference in parameter setting among different languages rather than the impact of various factors on affix-learning. Following the discussion of factors that affect affix learning and the stem-parameter approach, different data gathering techniques adopted in studies on inflectional acquisition such as experimental elicitation or natural production, selective looking, head turn preference procedure etc. will be discussed in the following sections.

neither has an RI stage nor does it have an imperative analogue stage (or any bare form stage for that matter).

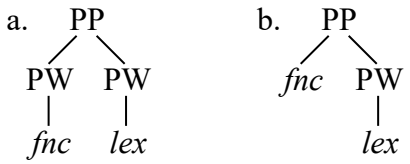
2.2 Major Theoretical Models on Inflectional Acquisition

2.2.1 Phonological Theories on Inflectional Development

Jusczyk et al. (1993) in their experimental study involving 9-month old American and Dutch infants have shown that in their first year, infants prefer to listen to words which observe the phonotactic constraints and prosodic patterns of their 'native' language. The experiments show that American infants prefer to listen longer to English words and Dutch infants to Dutch words that adhere to the phonetic, phonotactic, and prosodic organisation of their respective native languages (Jusczyk et al., 1993). This implies that children's tacit understanding of their ambient language's phonology precedes their production of word forms. Christophe and Dupoux (1996) and Christophe et al. (1997) based on their earlier experimental study (Christophe et al., 1994) have proposed that infants are probably utilising prosodic information to extract the word boundaries and internal structure of native word-forms. This helps the infants in building a prelexical representation consisting of native phonemes and syllables. This process has been called the **prosodic or phonological bootstrapping** in lexical acquisition. Once infants have established a prelexical representation of native word forms, they exploit the information about the distributional regularities (or statistical information on the sequencing of sounds), typical word shapes, and phonotactic constraints (Guasti, 2002).

Phonological models on inflectional acquisition address how early inflectional production is closely related to the acquisition of phonotactic constraints of the ambient language and how those constraints govern the production of inflections. These models attempt to account for the omission errors in early inflectional production by relating them to the interaction of the phonotactics of the child's target language. The **prosodic constraints model** (Demuth, 1996, 2001) proposes that children's words may be constrained at different levels of prosodic structure such as phonological words (PW) and phonological phrases (PP) as shown in (13).

(13) *Prosodic Words*

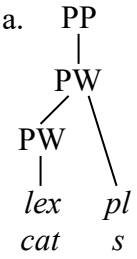


(Demuth, 2001, p. 16)

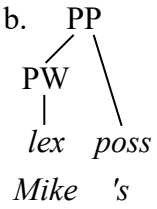
Demuth (2001) argues that the nature of these prosodic constraints may vary individually as well as cross-linguistically. Once these constraints are determined for each level of prosodic

structure, it is possible to predict which phonological units will be included or omitted in production given a new target utterance. According to this model, it is expected that the functional morphemes which are prosodified under the phonological word (PW) level will be acquired earlier than those that are prosodified at a higher level such as a phonological phrase (PP). The cases in point are the English plural marker *-s* and the possessive marker *-'s* as shown in (14).

(14) a.



b.



(Demuth, 2001, p. 24)

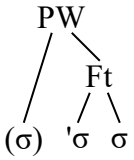
Brown (1973) reports that children exhibit an earlier acquisition of the plural morpheme *-s* compared to the possessive morpheme *-'s*. Thus, phonology may play a crucial role in determining the relative ease/difficulty in the acquisition of inflectional morphemes in addition to the semantic or syntactic constraints that are at play in children's inflectional development (Demuth, 2011).

Cross-linguistic studies have shown that children tend to produce prosodically licensed grammatical forms relatively earlier than those that are not. This accounts for some of the variability in the production of functional categories in children's early speech (Demuth, 2011). Song et al. (2009) demonstrates that children are more likely to produce the 3SG marker *-s* when it constitutes a simple coda than belonging to a consonant cluster, as in example *sees* vs *hits*. Also, the morpheme *-s* is produced more often when it occurs utterance finally than medially since utterance-medial syllables are typically shorter in duration offering only a shorter span for children to produce these segments.

Demuth (2001) using Spanish acquisition data demonstrates that children's early awareness of different levels of prosodic structure is reflected in the shape of the subject's early multimorphemic utterances. At 1;8 years of age, the Spanish child, Sophia's speech production that ranges from monomorphic to multimorphic words are prosodically constrained, where according to Demuth (2001, p. 13) "the upper bound on prosodic structure permitted is a Foot optionally preceded by an unfooted syllable" as shown in (15). Examples of Sophia's utterances are shown in Table 2.1. She argues that this result is not surprising since children are already seen to have an early awareness of the phonological systems of their native language as demonstrated by

Jusczyk et al. (1993), Christophe et al. (1994), and so on.

(15)



(Demuth, 2001, p. 13)

Table 2.1 Examples of Sophia's prosodically constrained Spanish utterances.

Child	**Adult Target**	**Gloss**
[kaléra]	/èskaléra/	'stairs'
[eméka]	/la muɲéka/	'the doll'
[nonáda]	/no ai náda/	'there isn't anything'

(Demuth, 2001)

Mateo Pedro (2015) in his study on Q'anjob'al inflectional acquisition has shown that both the initial sound and the syllable structure of a given verb condition the inflectional morphemes produced by the children. The inflectional morphemes that occured with verb stems that are vowel initial and of CVC syllable structure were seen to be acquired early by the Q'anjob'al speaking subjects. Not only were the CVC words more prominent in their speech production, but when they targeted other syllable structures, they were modified to the preferred CVC structure as shown in the examples below.

(16) tohi (CVC-V) (Xhuw, 1;11)
ch-ø-toj-i (C-CVC-V) [target form]
INC-ABS3-go-IV
'She goes.'

(Mateo Pedro, 2015, p. 118)

(17) lawi (CVC-V) (Xhuw, 2;1)
x-ø-laj-w-i (C-CVC-C-V) [target form]
COM-ABS3-finish-DER-IV
'It finished.'

(Mateo Pedro, 2015, p. 118)

Malayalam is a language that has a predominant CV/CVV syllable in addition to several phonotactic constraints which will be discussed in detail in Chapter 4. Malayalam also has

light stress on the first syllable. Whether these phonotactic constraints condition the inflectional production in children will be analysed in Chapters 5 to 8.

2.2.2 Morphological Models on Inflectional Acquisition: Rule vs Analogy or Dual-Route vs. Single-Route Models

A major debate in the acquisition and processing of morphological structures, of inflected forms in particular, concerns the dual-route versus the single-route models. While the phonological models generally account for the omission errors in early inflectional production, these models attempt to account for the overgeneralisation errors in children's speech, as in for example, forms like **breaked*, **singed* etc., produced by early English learners. The dual-route model (Pinker, 1991; Prasada & Pinker, 1993; Pinker & Prince, 1994; Marcus, 1995) argues that regular inflectional marking is carried out by a default affixation rule, and thus, regular inflected forms are not stored separately in the lexicon. On the other hand, irregular forms are stored in the lexicon in associated memory. For example, the past tense of the verb *walk* will be formed by the application of the default rule, i.e. verb + *-ed*, if it is not blocked by the existence of a corresponding irregular form in the lexicon (Pinker, 1999).

In contrast, the single-route model (Rumelhart & McClelland, 1986; Daugherty & Seidenberg, 1994; Elman et al., 1996) assumes that inflected forms are non-compositional and are stored in an associative memory network like the one posited by the dual-route model for irregular forms. According to this model, generalisations are formed based on phonological analogies. The child will form schemata of particular word-forms that are phonologically similar as he/she encounters an increasing number of inflected forms. The strength of each schema will depend on the input frequency and the number of members in a particular schema and so on. For example, the past tense for the verb *walk* will be produced by searching the memory for the past tense form *walked*. If *walked* is not available, the child is expected to generate a past tense form in analogy with all the phonologically similar forms. Even then, if a form is not available, the child may just say *walk* without inflecting it (Ambridge & Lieven, 2011).

Jaeger et al. (1996) in their PET emission study show that separate areas in the brain get activated during the tasks involving regular and irregular past tense forms in English. During the tasks involving regular past tense forms and nonce words, the dorsolateral prefrontal cortex, which is associated with online formulation of intentional and novel behaviours, has been acti-

vated. This part of the brain has not been activated for irregular past tenses. However, during the tasks involving the irregular past tenses, the middle temporal gyrus, which is associated with auditory memory traces, has been activated, which has not been the case with regular or nonce forms (Jaeger et al., 1996). Marslen-Wilson and Tyler (1997) in their study involving two aphasic patients, who have difficulties in the comprehension and production of inflected forms in English, have shown that while the subjects have difficulty with regular past tense forms, they do not exhibit this deficit for irregular words demonstrating the latter class' storage as separate content words. Evidence from German past-participle formation using the affix *-t* offers further evidence for the dual-route model (Penke, 2012). In German, if the verb stem already ends in a coronal stop [t], the marker *-t* is added after schwa epenthesis as in example *gehustet*. Going by the dual-route model, the child should produce forms like **gehustt* until it has figured out the schwa epenthesis rule. Whereas according to the single-route model, the child should either produce *gehustet* through analogy or simply *gehust* if it cannot retrieve the epenthised form. However, the form **gehustt* should never be found. The acquisition data from German attests precisely this kind of form lending strength to the proposals of the dual-route model (Penke, 2012).

Malayalam does not have a regular/irregular distinction when it comes to inflectional morphology, unlike English. Nevertheless, past tense and plural marking in Malayalam is assigned depending on the different classes of stems. There are default rules and exceptions for each class of stems. The overgeneralisation errors in inflectional marking in the longitudinal data have to be examined to determine the predictive strength of these proposals with respect to the Malayalam data.

2.2.3 Syntactic Theories on Inflectional Development

The Small Clause Hypothesis

English speaking children are seen to produce sentences that appear devoid of functional elements. They typically involve the omission of grammatical morphemes such as tense markers, auxiliaries such as *have* and *be*, and the copula *be* as can be seen in the examples given below.

(18) Crommer wear glasses. (Eve, 2;0)

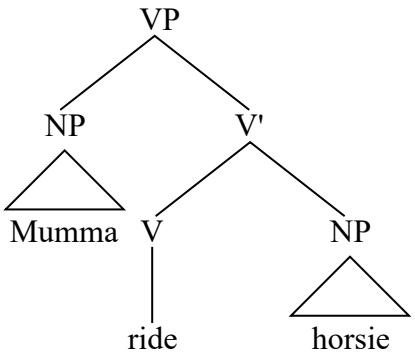
(19) Eve gone. (Eve, 1;6)

(20) That my briefcase. (Eve, 1;9)

(Guasti, 2002, p. 106)

All these sentences lack the overt expression of the functional features expressed by the I node. In order to account for sentences such as these and others that lack the expected grammatical markers, Radford (1990) proposes the **small clause hypothesis**. According to him, children's early sentences are VPs (21) and lack the higher functional projections unlike the adult sentences. This hypothesis also accounts for the use of RIs in finite contexts. Children cannot assign or check the finite inflections or perform other covert or overt checking because they lack the required syntactic projections. He argues that although functional categories are part of UG, their availability is subject to maturation.

(21)



(Guasti, 2002, p. 107)

However, this is more of an English-centric account. Cross-linguistic studies have shown that children do not lack functional categories in their early speech productions. In languages with richer functional morphology than English, children are observed to distinguish between finite and non-finite verb forms. This is also observed with respect to their use of negation in French and German, and verb movement in V2 languages such as Dutch and German. Children's production of *wh*-questions also demonstrates that children's mental syntactic representations should include CPs (Guasti, 2002). The production of RIs also varies cross-linguistically. Compared to children acquiring English, Italian and Catalan acquiring children produce very few RIs. The small clause hypothesis cannot explain the varying rates of RIs in different languages (Ambridge & Lieven, 2011) or account for the variation in the appearance of functional categories. Radford does suggest that it might be a transient stage. Nonetheless, evidence suggests that this cannot be the reason for the 'deviant' forms in children's productions (Guasti, 2002).

The Full-Competence Hypothesis

Early learners of non-V2 languages like French are seen to distinguish between the non-finite and finite forms in negative sentences which contain a NegP located between IP and VP, just like in the adult constructions. In languages like French, while the finite verb precedes the negation (22), the infinitive follows it (23). Children's utterances also adhere to the same pattern ((24)-(25)) (Guasti, 2002).

(22) Marie ne **mange** **pas** (Vfin Neg)
Marie NEG eats not
'Marie does not eat.'
(Guasti, 2002, p. 110)

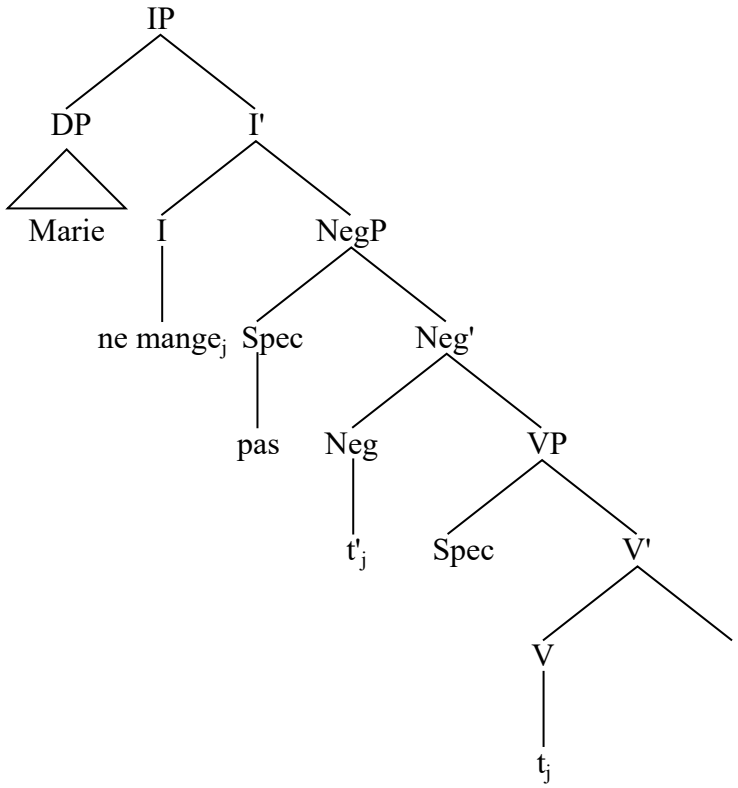
(23) pour ne **pas** **manger** (Vfin Neg)
in order to NEG not eat.INF
'no meaning.'
(Guasti, 2002, p. 110)

(24) pas manger la poupée (Nathalie, 1;9)
not eat.INF the doll
'The doll does not eat.'
(Guasti, 2002, p. 110)

(25) elle roule pas (Grégoire, 1;11)
it rolls not
'It does not roll.'
(Guasti, 2002, p. 110)

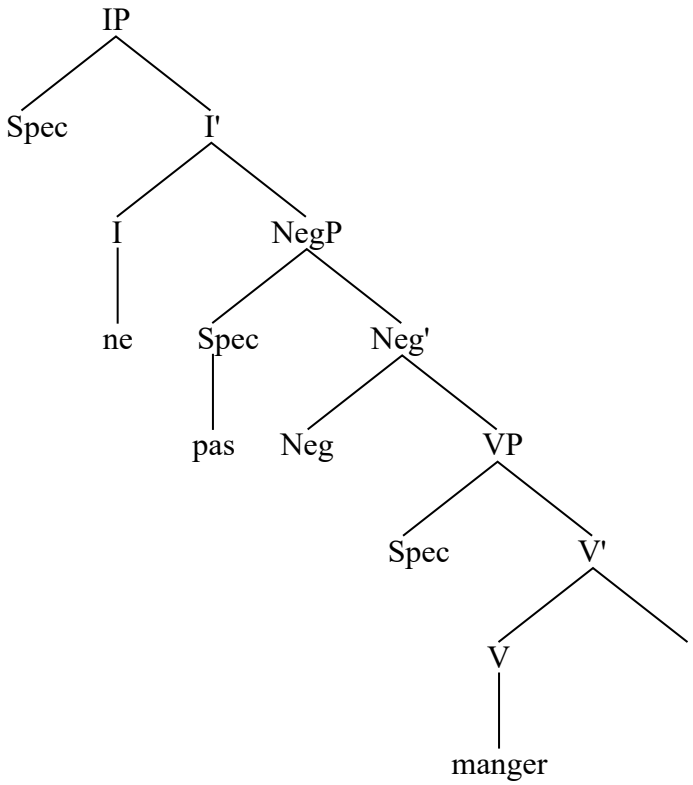
As mentioned before, negative sentences comprise an IP and a VP with a NegP located between them, the specifier of which hosts the negation *pas* in French. The surface structure of finite negative sentences (where the finite verb precedes negation) shows that the verb has raised to I (26), which in turn demonstrates that children's utterances are at least IPs. On the other hand, the non-finite verb forms remain in the VP itself (27) (Guasti, 2002). Therefore, the surface structure of the finite sentence (22) is (26) and the non-finite (23) is (27).

(26)



(Guasti, 2002, p. 114)

(27)



(Guasti, 2002, p. 114)

Likewise, in V2 languages, the finite verb occurs in the second position (28) while the infinitives remain in the clause final position, i.e., VP (29). The fact that the verb occurs in the second position in the finie contexts reveal that the verb has raised to a higher node. Since

German is a head-final, SOV language, which takes its VP complements to the left, in order to arrive at the correct surface structure (where the finite verb occupies the second position in a sentence), the finite verb has to move to C, and not I. Early German learners are observed to differentiate between the non-finite and finite forms as they occur in adult grammar as shown in (30) and (31) respectively. The fact that children produce sentences that adhere to the differences in positioning between finite and non-finite verbs demonstrates that children's finite verbs are raised to C, which further implies that the finite sentences of early V2 language learners are CPs and not VPs. The surface structure of the finite utterance produced by the German child Andreas (30) is shown in (32).

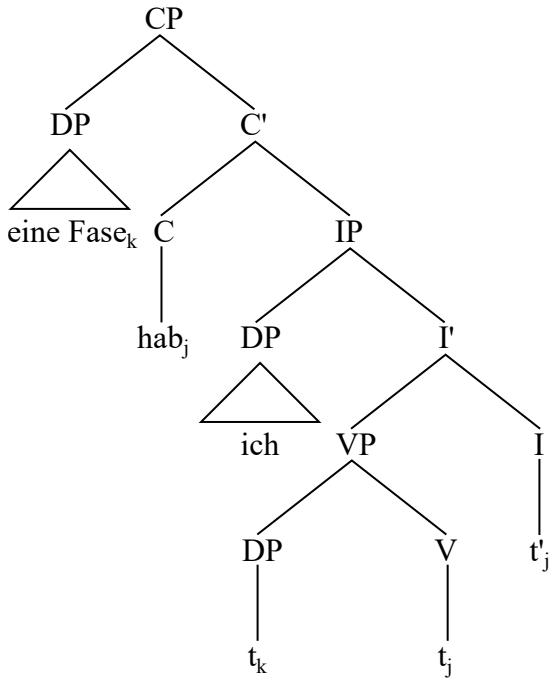
(28) Johann **kaufte** ein Buch
Johann bought a book
'Johann bought a book.'
(Guasti, 2002, p. 114)

(29) Simone wird das **lesen**
Simone will that read.INF
'Simone will read that.'
(Guasti, 2002, p. 111)

(30) eine fase hab ich (Andreas, 2;1)
a vase have I
'I have a vase.'
(Guasti, 2002, p. 112)

(31) hij op kussens slapen (Hein, 2;6)
he on cushions sleep.INF
'He sleep on cushions.'
(Guasti, 2002, p. 112)

(32)

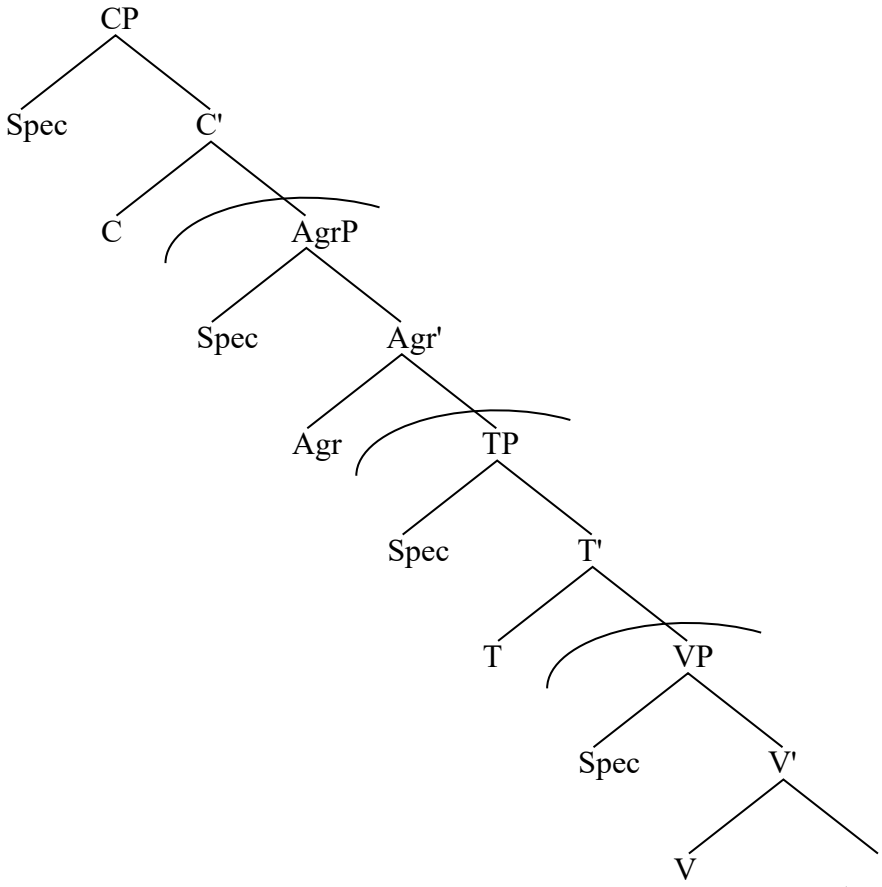


(Guasti, 2002, p. 118)

Using evidence from V2 languages which clearly show that children can distinguish between non-finite and finite forms, Poeppel and Wexler (1993) propose the **full-competence** hypothesis which states that children have access to functional categories from early utterances onwards. The **truncation** model (Rizzi, 1993) and the **Agreement/Tense Omission Model** (ATOM) (Schütze & Wexler, 1996), which will be discussed in the following sections, are two of the prominent models that try to account for the omission errors and RIs under the full-competence framework.

Truncation Model The truncation model (Rizzi, 1993) states that children can project and have access to all the functional categories up to CP but they do not know that it is obligatory to do so, unlike adults. Hence, children optionally truncate functional projections. The truncation mechanism can target any functional projection as shown in (33).

(33)



(Guasti, 2002, p. 142)

Once a functional projection is truncated, every dominant projection above the targeted projection is also truncated. Only the projections dominated by the truncation site remain. The fact that truncation is optional explains how children can produce RIs and omissions of markers along with correctly inflected forms. For example, in a language like Malayalam which is inflected for tense and not agreement, the model predicts that the truncation below TP would result in the omission errors of tense inflections as in (34) where the subject appears in the nominative case which is the default case in Malayalam.

(34) *avan paaʈʈi paaʈ_
he song sing.*0PST
avan paaʈʈi paaʈ-i [correct form]
he song sing-PST

According to this model, RIs are truncated below TP. Therefore, they are either VPs or they may contain some functional layer above VP which explains their restricted distribution. Thus, the model correctly predicts the absence of RIs in constructions involving auxiliaries since auxiliaries are licensed by T. Similarly, RIs cannot occur with clitic and weak pronoun

subjects since they are licensed by Agr. Further, it also explains the absence of non-subject clause in the first position in RIs since the licensing of non-subject constituents requires a CP layer which is absent in RI clauses, which are structures truncated below CP (Guasti, 2002). The model also predicts the absence of RIs in *wh*-questions since the latter involve movement to C. However, Hamann (2002) has found that RIs occur in early English *wh*-questions which cannot be explained by the truncation model. Nonetheless, the presence of RIs in early English *wh*-questions is a disputed phenomenon since this pattern has not been attested in any other languages so far (Guasti, 2002).

Rizzi (1993) proposes that the complete acquisition of the obligatoriness of the CP projection in all the clauses is subject to maturation. However, like the small-clause hypothesis, it cannot account for the varying rates of RIs in different languages. Also, it cannot explain language-specific patterns of tense/agreement omission or non-omission, for example, the omission of subjects with RIs as compared to finite verbs where the subject is present (Ambridge & Lieven, 2011, p. 145).

The Agreement/Tense Omission Model (ATOM) This model works under the assumption that when a grammatical feature is underspecified (or fails to be specified), it does not surface in a syntactic representation and, consequently, the syntactic processes associated with it (e.g. feature checking) do not take place. The ATOM (Schütze & Wexler, 1996) proposes that children's omission of tense or agreement morphemes is the result of the interactions between several constraints present in the child and adult grammar such as the *tense constraint* (35), *minimise violations* constraint (36), and the *uniqueness constraint* (37) during the stage of development termed as the **optional-infinitive** stage (Wexler, 1994), which is another term for the root infinitive stage. Of the three constraints, the first two hold in both adult and child grammars while the third (uniqueness constraint) only holds in child grammar. Overriding the uniqueness constraint is subject to maturation.

(35) *Tense Constraint*

A main clause must include a specification of tense.

(Guasti, 2002, p. 133)

(36) *Minimise Violations*

Given two interpretations, choose the one that violates as few grammatical constraints as possible. If two representations violate the same number of constraints, then either

one may be chosen.

(Guasti, 2002, p. 135)

(37) *Uniqueness Constraint*

A subject can check the uninterpretable feature of either T or Agr, but not both.

(Guasti, 2002, p. 134)

According to the model, when the child omits tense or agreement markers, the child is adhering to the uniqueness constraint. Although this omission violates the tense constraint, the minimise violations constraint allows the violation of one constraint and permits the production of instances of omission. At the same time, the child can correctly mark both tense and agreement adhering to the tense constraint. Both omission of marking and correct marking involve the violation of one constraint (either (35) or (37)) but is permitted by the minimise violations constraint in child grammar. This explains the simultaneous occurrences of the correct instances of inflectional assignments and omission errors in children's productions during the same developmental period. In the adult grammar, the uniqueness constraint does not hold and the utterances will be correctly marked for both tense and agreement and do not violate (35) and (36).

In addition to explaining the difference between child and adult productions, the ATOM also offers explanation to the pattern of errors characteristic of children's utterances in the optional-infinitive / root infinitive stage. Table 2.2 shows the patterns of inflectional assignment in child language.

Table 2.2 Error patterns predicted under the ATOM.

Inflectional Combination	**Result**	**Example**
[+T, −AGR]	−NOM; default case	**Her play* **Her played*
[−T, +AGR]	+NOM; no tense marking	**She play*

(Schütze & Wexler, 1996, p. 678)

In the first case, where the child correctly marks the tense but fails to mark the agreement, the nominative case is unassigned since it is the Agr that checks the nominative case. As a consequence, the default accusative case surfaces instead. In the second instance where the child correctly marks the agreement, but fails to mark the tense, the present tense morpheme *-s* and past tense morpheme *-ed* cannot be assigned because tense is not specified. However, the

subject is assigned the nominative case because it is checked by Agr. Similar to the production of correctly inflected forms and omission of inflections during the same developmental period, children also produce RIs and inflected utterances during the same period. The child can correctly mark both tense and agreement adhering to the tense constraint, consequently violating the uniqueness constraint. Since the violation of one constraint (either (35) or (37)) is permitted by the minimise violations constraint in child grammar, the model explains the co-occurrence of RIs and correct instances of inflected forms in children's speech during the optional / root infinitive stage. According to this model, the disappearance of uniqueness constraint from the child grammar is subject to maturation.

The ATOM correctly predicts the absence of RIs in constructions involving a) non-subject constituents in V2 languages, b) auxiliaries, and c) pro-drop languages. As mentioned earlier, RIs are non-finite constructions and thus, they remain in VP. Consequently, they cannot move to C, unlike the finite verbs, and license a non-subject constituent in Spec CP. Since auxiliaries are licensed by T, they cannot be licensed in constructions where T is underspecified. In pro-drop languages, Agr does not have an uninterpretable feature since it licenses null-subjects. As a result, children only have to raise the lexical DPs as far as SpecTP, in order to check the uninterpretable feature of T. Thus, children's utterances will satisfy both the tense constraint and the uniqueness constraint whereas in an RI clause, where there is no specification for tense, the utterance will violate the tense constraint. Therefore, under minimise violations constraint, OI clauses will not be produced since the child will be producing constructions that have comparatively fewer violations (Guasti, 2002).

Although the ATOM captures the difference between child and adult constructions and the varying combinations of inflectional assignment, there are some properties for which the model fails to provide an account. The model does not predict forms like **Her plays* because in addition to the tense, the morpheme *-s* also marks agreement in English. It has been pointed out that while the model predicts that forms like **Her plays* should not be produced, researchers have found that children do produce such forms with agreement between non-nominative subjects and finite verbs (Ambridge & Lieven, 2011, p. 148). (Guasti, 2002, p. 139) notes that it might be difficult for ATOM to explain the incompatibility of RI clauses with clitics and weak pronouns. According to the model, in an RI construction where Agr is specified and tense is not, clitics and weak subject pronouns are expected to be found since they are licensed by Agr. However, such constructions are never cross-linguistically attested. Guasti (2002, p. 139) also points out that the

model does not account for the lack of infinitives in *wh*-questions in cross-linguistic data since underspecification of tense does not prevent RIs from occurring in *wh*-questions. However, as mentioned earlier while discussing the truncation model, the presence of RIs in early English *wh*-questions is a disputed phenomenon since this pattern has not been cross-linguistically attested (Guasti, 2002).

Table 2.3 presents a comparative summary of the different syntactic models discussed above. Malayalam is a language that lacks subject-verb agreement, with nominative as the default case. However, it is marked for tense and has a rich nominative inflectional system. Since there is no agreement, the children's errors in tense marking and case-marking serve as the primary indicators with which to evaluate the predictions of the syntactic models with respect to the acquisition of Malayalam inflections. Malayalam is also a language where inflectional morphology is driven by its phonology and phonotactic constraints as can be seen Chapter 4. The phonology strongly impacts acquisition as we shall see in the subsequent chapters. We will discuss the observable patterns in the verb and nominal acquisition data and evaluate the data vis-à-vis the theories in Chapter 9.

Table 2.3 Comparative summary of the major syntactic models of inflectional acquisition.

	Small-clause	**Full-Competence**	
		Truncation	**ATOM**
Lack of functional categories in early grammar	✓[2]	×	×
RIs with non-subject constituents in V2 languages	×	×	×
RIs with auxiliary constructions	×	×	×
RIs in pro-drop languages	×	×	×
RIs in *wh*-questions	×	×	✓[3]
RIs with clitic and weak pronoun subjects	×	×	✓[2]
Maturational account	✓	✓	✓

[2]problematic
[3]disputed

2.3 Affix-Learning vs Parameter Setting

Dressler (2010, 2012) and Penke (2012) discuss several factors that influence the segmentation and the acquisition of inflected morphemes. We provide a brief summary of these factors in Section 2.3.1. and will use these criteria in evaluating the acquisition of inflectional morphemes in Malayalam in subsequent chapters. The various factors that influence early affix learning are a) frequency, b) phonological salience, c) relative positioning of the marker, d) morphological transparency, e) biuniqueness, f) semantic or conceptual transparency, and g) language typology (Dressler, 2010, 2012; Penke, 2012). On the other hand, Hyams (2008) proposes that the relative ease or difficulty observed in the acquisition of inflections across languages is a result of parameter-setting differences in child grammar. We will also discuss Hyams' (2008) stem parameter theory in detail in Section 2.3.2.

2.3.1 Factors that Influence the Acquisition of Inflectional Morphemes

Frequency of the affixal form is an important factor which is understood in two ways: a) the *type frequency* of an inflectional marker, i.e. the number of stems to which an inflectional marker applies and b) the *token frequency* of an inflectional marker which denotes the number of times an inflectional affix occurs in the child's input. The higher the type and token frequencies, the faster and easier their acquisition (Penke, 2012).

Phonological salience of an inflectional morpheme also affects the rate of acquisition. Phonological salience is the perceivability of an inflectional morpheme in the input. Hence, syllabic or multisyllabic markers should be easier to identify than markers that consist of a single consonant or even just a VC (Penke, 2012). Associated with the concept of salience is the relative positioning of the inflectional marker with respect to the base. Markers occuring at the periphery of a word are seen to be identified earlier than markers that occur word-internally (Dressler, 2012). Stephany and Voeikova (2009) attribute the early emergence of case markers in agglutinating languages, where they constitute separate markers and are expressed after the plural markers, to the periphery effect.

Morphological transparency plays an important role in aiding the segmentation process. Morphologically transparent inflectional markers are those which do not alter the phonological form of a stem to which they are affixed. It is easier to identify such markers than those that alter the lexical base. For example, forms such as *walk-ed* can be easily segmented into the

stem, *walk*, and the past tense morpheme *-ed*, whereas irregular forms such as *brought* and *took* make it difficult to isolate the lexical base from the affix making segmentation impossible (Penke, 2012). Dressler (2010, p. 112) demonstrates that while the English plural form *brother-s* is morphologically transparent, the form *brethr-en* lacks morphological transparency since it is difficult to detect the base. In the latter case, the affix itself is identifiable although it alters the base.

An inflectional marker is said to exhibit *biuniqueness* if there is a one-to-one correspondence between the meaning it expresses and its form. Such markers are expected to be acquired earlier than others that express different and multiple grammatical features as in the case of fusional affixes which will be discussed later while explaining the influence of typology in affix-learning. An example is the *-s* suffix in English which functions as the possessive and plural marker on nouns but as both the 3rd person and singular number marker on verbs. This affix is observed to be acquired later (Brown, 1973). The ambiguity between the form and its meaning results in a surprisingly late acquisition given the fact that English morphology is so sparse (Penke, 2012).

Another crucial factor is the *semantic* or *conceptual transparency*. Inflectional affixes that encode semantically or conceptually more basic or transparent notions such as number, tense, aspect or person are observed to be acquired earlier than affixes that are exponents of formal features such as case, conjugational classes of verbs, or grammatical gender in languages like Hindi (e.g. *kitaab* (feminine) vs *pustak* (masculine)) (Penke, 2012).

The morphological properties of languages vary when it comes to inflectional assignment. Some languages are predominantly prefixing (e.g. Navajo, Hunde etc.) and the affixes precede their stems, whereas some languages are predominantly suffixing where affixes follow their stems (e.g. Dravidian languages, West-Greenlandic languages etc.). Another major typological difference is between analytic languages and synthetic languages. The former class has little inflectional morphology with a lower morpheme-to-word ratio (e.g. Mandarin Chinese, English etc.) while the latter has rich inflectional morphology with a higher morpheme-to-word ratio. The latter class includes agglutinative languages with highly regular inflections showing one-to-one correspondence between morpheme and meaning (e.g. Malayalam, Turkish etc.) and fusional languages where a single morpheme signals multiple grammatical features (e.g. Spanish, Hebrew etc.). For example, in Turkish, which is an agglutinative language, the form *ev-ler-e* corresponds to [house-PL-DAT] where each morpheme exhibits biuniqueness and can be easily

segmented. Whereas in Russian, a fusional language, the equivalent of the above discussed Turkish form is *dom-am* [house-DAT.PL] where the suffix *-am* signals both dative and plural and thus less transparent (Dressler, 2010, p. 112). Such distinctions employed by typologically different languages affect the course of acquisition of those systems. For example, children born in environments with richly-inflected languages tend to acquire their inflectional morphology relatively early in comparison to their counterparts hearing languages with sparse inflectional morphology (Dressler, 2012). The relatively late acquisition of the inflectional morphology of English is attributed to the sparseness of the morphology of English (Hyams, 2008).

2.3.2 Parameter Setting and the Acquisition of Inflections

In order to differentiate between the inflectional systems that are typologically different, Hyams (2008) makes use of the core and peripheral distinctions in grammatical systems. Hyams (2008) proposes that the grammatical status of inflectional systems varies across languages such that inflection can either be considered part of the core grammar, which is the case in languages like Italian, Russian, Turkish etc. with rich inflectional systems, or it can be considered a peripheral property of the grammar of a language, which is the case in languages like English with sparse inflectional systems. The former system will be acquired relatively easily compared to the latter.

According to Hyams (2008), cross-linguistic variation in the rate of acquisition of verbal inflections is determined by what she terms as the "stem parameter". The stem parameter is stated as "a verbal stem does or does not constitute a well formed word" (Hyams, 2008, p. 197). Children's omission errors are seen as the result of the setting of the stem parameter. For example, the uninflected forms in early English occur because such forms are well-formed at the morphological level since bare stems constitute well-formed words in English. On the contrary, in languages like Italian, children rarely produce uninflected verbs because a bare stem is ill-formed in such languages which in turn results in ungrammaticality. In the former case, since the verbal stem constitutes a well-formed word and the parameter has been set accordingly, in order to satisfy the well-formedness condition, the child does not have to learn any inflections. As a result, the acquisition of English inflections is delayed. While in the latter, the stem parameter setting expedites the inflectional acquisition process.

Hyams (2008) also employs cross-linguistic data to support her proposal. The overgeneralisation of overt affixes in lieu of zero-affixation observed in rich inflectional languages like Russian and Serbo-Croatian is assumed to be the result of the child setting the stem parameter

to the positive value thereby replacing all zero affixes with overt ones. Also, the fact that the acquisition of tensed forms and other morphologically complex structures precedes the mastery of bare imperatives in Turkish is taken as the result of the parameter setting which fixes bare stems as ill-formed. Hyams (2008) argues that the generally observed cross-linguistic phenomena such as overgeneralisation and undergeneralisation are better explained by the parameter setting effects rather than by the effect of affix-learning principles.

As mentioned earlier, Malayalam is an agglutinative language. Most of its inflections are highly productive (except for very few markers like the sociative marker *-ooṭi* which occurs only with a few verbs of saying, and the second locative *-aṯṯi* which occurs with a restricted set of stems), phonologically salient, morphologically and semantically transparent, and exhibit biuniqueness (a notable exception is the marker *-aan* which signals both the imperative and the infinitive (eg. pook-aan [go-IMP / go-INF])). Thus, Malayalam, with its rich inflectional system having features that aid inflectional acquisition and typological advantage, presents an interesting case. With most of the factors that enhance the inflectional development in place, Malayalam acquisition presents an opportunity to look beyond the established factors and figure out the additional constraints that determine inflectional development if there are any. We will demonstrate in the following chapters that in Malayalam, *mode of selection* of a marker, i.e, whether a marker is selected based on phonological or morphological/semantic features, indeed has an effect on the acquisition of inflections.

2.4 Different Approaches to the Development of Inflections

Acquisition studies focus on children's structures, their similarities or differences compared to adult structures, and the kind of errors they make in order to understand the nature of their linguistic development. In order to do so, researchers adopt different methodologies. The *naturalistic approach* and the *experimental approach* are the two main methodologies that are often adopted. Figure 2.2 shows different approaches to the development of inflections and the kind of studies they employ.

The naturalistic approach deals with naturalistic data. The studies utilising naturalistic data do so mainly by longitudinally recording children's spontaneous speech. Advantages of this approach is that a single corpus can test a range of hypotheses and the role of the input can be examined as well. Also, the naturalistic approach approximates natural learning without

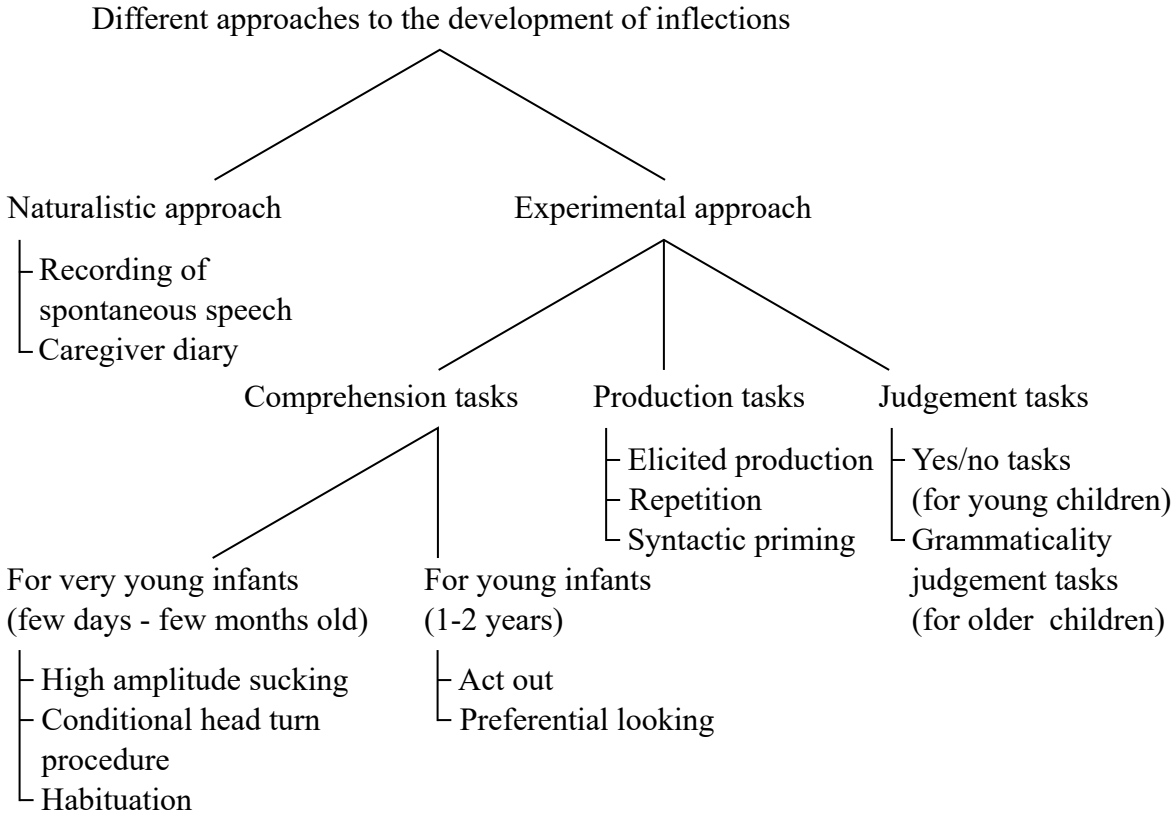


Figure 2.2 Different approaches to the development of inflections.

the artefact of experimentation. However, these studies are time consuming and the sampling is often thin. Another variant of this approach is the caregiver diary study where the parent keeps detailed notes on the child's linguistic progress. This is often combined with longitudinal recording of spontaneous speech (Ambridge & Lieven, 2011). Children's spontaneous speech can be recorded cross-sectionally as well.

In the experimental approach, researchers employ specially designed tasks to test the child's linguistic competence. In acquisition experiments, tasks are typically designed to analyse children's comprehension, production, and repetition/imitation skills. Some of the comprehension study tasks include **act-out** tasks where the child can be asked to enact a scene using a nonce word and **preferential looking** tasks where the child usually hears a sentence and watches two scenes, one of which is matching with the sentence. With very young infants, methods like **high-amplitude sucking** (e.g. few days old infants), which tests whether infants can distinguish between sounds, identify phonotactic and prosodic patterns employed by languages etc., **conditioned head-turn preference procedure** (e.g. infants aged 4 months), where audio is produced through two different loudspeakers and children have to fix their gaze on one one of the loudspeakers, and **habituation**, where new categories are introduced to see if it causes recovery of

looking times implying the formation of two categories after the child gets habituated with the repeated presentation of similar stimuli, are generally used (Ambridge & Lieven, 2011).

In the production testing tasks, which are usually employed with older children, the child is often asked to describe a scene using a novel or nonce verb. This type of an experiment, often known as **elicited production**, not only has a high degree of control over the target variables but can also be used for collecting data on structures that are infrequent in spontaneous speech. Nonetheless, this is seen to be relatively difficult for children mainly because of memory issues. In the **repetition** tasks, the children are asked to repeat the experimenter's utterance in order to analyse the deviations from the target utterance. The target utterances can be ungrammatical and the child's resulting substitutions or corrections are taken as indicators of grammatical competence. Another production testing task is **syntactic priming** where the experimenter describes a scene by employing a particular syntactic structure. Then, the child has to describe a new scene by employing the same structure which was previously presented as the prime (Ambridge & Lieven, 2011).

Further, there are judgement tasks that analyse children's interpretations or notion of grammaticality/ungrammaticality. These involve **yes/no judgement** tasks, where the child is shown a picture and is asked a yes/no question where the child just has to answer yes or no, and **grammaticality judgement** tasks where the child hears a sentence which is illustrated in a picture and rates the acceptability of the sentences (e.g. 'said it right' vs 'said it a bit silly'). The former and latter are suitable for young and older children, respectively (Ambridge & Lieven, 2011).

However, it has to be noted that no methodology can be foolproof since children's performance will not just reflect their linguistic knowledge but will depend on factors such as memory and attention span as well (Ambridge & Lieven, 2011). In this thesis we make use of naturalistic data. The spontaneous speech of Malayalam speaking children is analysed to study their inflectional development as a whole since a longitudinal corpus offers insights into a range of phenomena and can test different hypotheses/predictions.

Chapter 3

Subjects and Method

3.1 Longitudinal Data

This study focuses primarily on longitudinal data. Cross-sectional data is only used to either substantiate the analyses of the longitudinal data or to verify if they diverge from each other in any respect. The subjects in this longitudinal study are two monolingual Malayalam speaking children, a girl and a boy. Both children belong to middle-class families with Malayalam as the only language of communication at home and in the larger social environment. Spontaneous speech samples of the children were recorded once a fortnight as follows. The study is covered by the Indian Institute of Technology, Bombay's ethics approval and the recordings were made with the full consent of the parents. The language profiles of the children's parents are included in Appendix A.

The girl child, whom we label A, was recorded between the ages of 1;9.14 and 2;10.3, and the boy child, whom we label H, was recorded between the ages of 2;3.28 and 3;0.17. Overall, A was recorded over 26 sessions and H over 18 sessions. Each session ranges in duration from 30 to 50 minutes. All the recordings were made on a Sony IC recorder in the children's homes, capturing the children's spontaneous interactions with the researcher generally, and occasionally with their mothers and grandmothers. Additional materials like puzzles and picture books were used to encourage the children to speak and to facilitate interaction with the researcher.

All the recordings have been transcribed in ELAN (Wittenburg et al., 2006), a multimedia language annotation tool developed by the Max Planck Institute for Psycholinguistics. Six different tiers have been created to organise the data at different levels and include the following: utterance → child's utterance → words → morphemes → gloss → inflections. The utterance

tier includes utterances of both the child and the people interacting with the child from which the child's utterances have been separated to form a different tier. The children's utterances have been entered in a primary utterance tier and sub-tiers were created to separate the data at the word, morpheme, and gloss levels. A further tier encodes the inflections and is separate from the morpheme tier. Figure 3.1 presents the sample picture of the data organised at different levels. The data consists of a total of 3689 and 1072 utterances for A and H, respectively. While every instance of inflectional marking has been analysed, non-words, false starts, fillers, songs, imitations, and repetitions (except when used for emphases) have been excluded from the analysis. Examples of such exclusions are included in Appendix B.

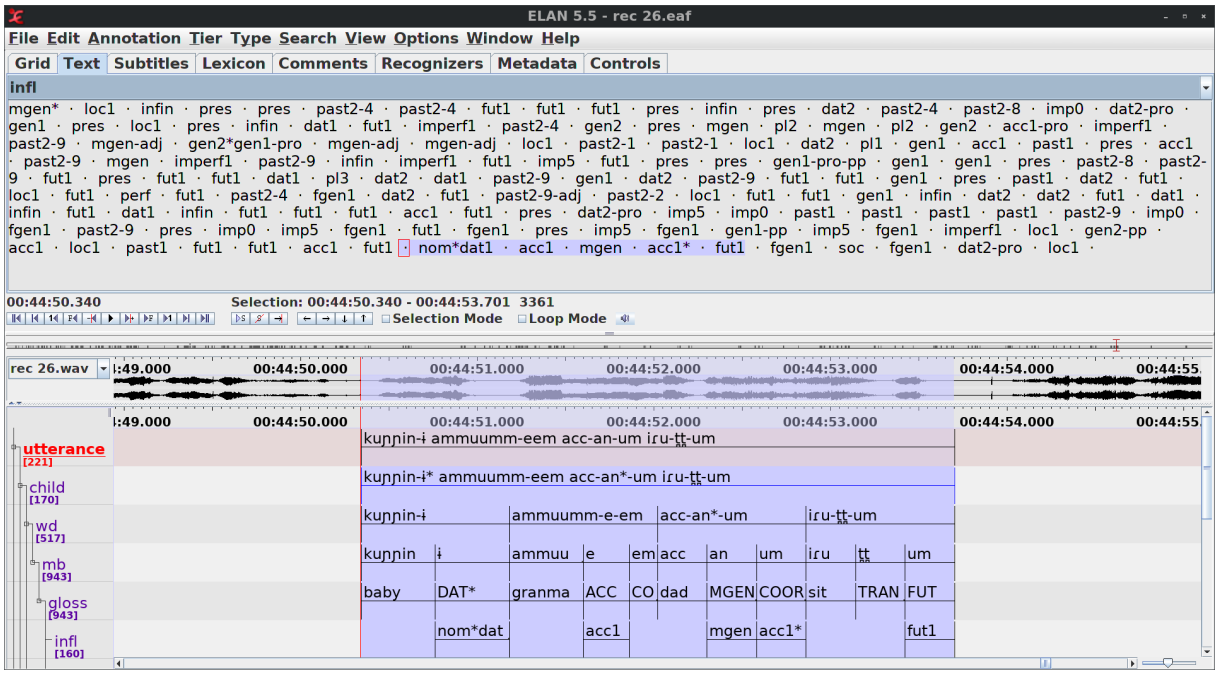


Figure 3.1 Sample picture of the ELAN file with data organised at different levels.

3.1.1 Error Analysis

Here, we briefly describe the notations used to indicate the disparities between children's and adults' utterances in this thesis. While omission errors are taken as indicators of incomplete acquisition of a particular marker, commission errors are considered instances of productive use of a particular affixation rule. The details of error analysis are discussed in the following.

Errors in inflectional production can be broadly classified into two types: a) omission errors and b) commission errors. The latter can be further subdivided into substitution errors and overgeneralisation errors. Omission errors occur when the marker is not employed in contexts

where it is obligatory. In the following example from the Malayalam data, the child fails to overtly mark the object with the accusative marker *-e*.

(38) **puli_** kaɳ-ʈ-aa? (H, 2;5.2)
leopard.*0ACC see-PST-Q
puliy-e kaɳ-ʈ-aa? [adult form]
leopard-ACC see-PST-Q
'Did you see the leopard?'

Likewise, in the Turkish and German examples below, the children have omitted the agreement markers, the first person singular and the third person singular markers, respectively, in contexts where they are obligatory.

(39) *kaakti_ [Correct form: kaakti**m**] 'get up'.

(Aksu-Koç & Ketrez, 2003, p. 37)

(40) *das komm_ nich in frage [Correct form: das kommt nich in frage] 'that is out of the question'.

(Penke, 2012)

In commission errors, when the inflectional marker is employed with stems that do not normally take that inflection, we call them substitution errors. In the German example below, the 2PL marker *-t* is incorrectly substituted with the 3PL marker *-n*.

(41) *ihr gehn [Correct form: ihr geh**t**] 'you go'

(Penke, 2012)

Similarly, in the Malayalam data the accusative marker *-e* which is only assigned to animate objects is incorrectly assigned to the subject noun.

(42) baag-ilɨ **kokkin-e*** oɳʈ-oo? (A, 2;5.16)
bag-LOC crane-*ACC have-Q
baag-il **kokkɨ** oɳʈ-oo? [adult form]
bag-LOC crane have-Q
'Is there a crane inside the bag?'

When the default rule is used with exceptional categories that do not normally obey that rule, we see commission errors of overgeneralisation. Children's productions of the past tense forms of irregular verbs show overgeneralisation of the default past tense rule of *-ed* assignment, creating *bleeded**, *singed**, **holded* etc. (Pinker & Prince, 1994) instead of the expected correct

forms *bled*, *sang* and *held* respectively. The observed u-shaped learning curve provides further evidence of such a rule application. Children begin by producing the irregulars (rote-learned), then move to a stage of producing regularised forms (default rule application), and return subsequently to the unpredictable irregulars (Pinker, 1999). Both omission and commission errors indicate that the child has not yet fully acquired the inflectional marker under consideration.

Malayalam is a topic drop language and as such permits discourse-salient arguments to be dropped. If the child drops arguments (subject, object or indirect object) that are retrievable from the discourse context and if the productions are otherwise acceptable in adult language, they are deemed grammatical outputs. When the utterances are expressly marked with an asterisk, they are deemed to be ungrammatical productions. The adult equivalents are included for the following standard reasons: a) the elided arguments are given within square brackets as can be seen in (43), b) they show the actual surface structures of word forms as children's phonological productions deviate from those of adults, and c) they provide the correct forms for comparison against children's productions with errors in them.

(43)	biykkatt̪	paappam	ṯi-n̪n̪u	(A, 1;9.2)
	biscuit	food	eat-PST	
	[ɲaan]	biṣkkatt̪	ṯi-n̪n̪u	[adult form]
	[I]	biscuit	eat-PST	
	'I ate biscuits.'			

Where commission errors occur, the utterances are marked with an asterisk '*' to indicate that the forms are incorrect, as shown in (44), and the correct forms are provided in the adult equivalents as expected forms.

(44)	iiccay-e	koll-i*		(V: 3;1.3)
	fly-ACC	kill-PST		
	[ɲaan]	iiccay-e	ko-**n̪n̪u**	[expected form]
	I	fly-ACC	kill-PST	
	'I killed the fly.'			

Omission errors are indicated using an underscore '_' in the child's utterance and are glossed with '*0' in the gloss as shown in (45).

(45) kiḷi_ kaaṇi-cci ṯay-aam (A, 2;1.1)
bird-***0ACC** show-PTCP give-MOD
kiḷiy-e kaaṇi-cci ṯar-aam [expected form]
bird-ACC show-PTCP give-MOD
'Will show (you) the bird.'

3.1.2 Morphological Development

Three different criteria are used to analyse the morphological development of the subjects with respect to inflectional morphology. They are: a) inflectional paradigms, b) Mean Length of Utterance (MLU), and c) type and token frequencies. To these, we also add productivity of the markers.

Inflectional Paradigms

Inflectional paradigms represent the development along the paradigmatic axis of the morphological structure. Their occurrence is a critical stage in the development of inflections because they show that the child has successfully figured out how the minimal phonological changes in words are related to corresponding changes in the function and meaning of the inflected forms. Thus, the emergence of inflectional paradigms indicates that the child has been able to segregate the stems and inflectional pieces and has also been able to figure out the grammatical information that is encoded in the inflectional pieces. However, the emergence of inflectional paradigms does not entail that the acquisition of inflectional morphemes is complete. Rather, it indicates that the child is in the process of acquiring the inflectional markers (Stephany & Voeikova, 2009).

In our data, inflectional paradigms are manifest in A's and H's speech in the very first recording itself. The details of inflectional paradigms during each age are given in Appendix C. Both the children's speech productions show a subsequent increase in the total number of inflectional paradigms as well as in the number of multi-member paradigms as the children progress in age. This increase in both the number and variety of paradigms indicates the productive use of both nominal and verbal inflections by the children. Tables 3.1 and 3.2 provide examples of inflectional paradigms in A's and H's speech, respectively.

Table 3.1 Examples of inflectional paradigms in A's speech.

Number of members	Age	Category	Form	Gloss
3	2;1.15	noun	*koŋŋan*	monkey.NOM
			koŋŋan-e	monkey-ACC
			koŋŋan-te	monkey-GEN
6	2;7.16	verb	*va-n̪n̪u*	come-PST
			var-um	come-FUT
			va-n̪n̪i	come-PTCP
			va-n̪n̪-iʈʈi	come-PTCP-PFV
			var-aʈʈe	come-OPT
			var-aɳ-ee	come-MOD-IMP

Table 3.2 Examples of inflectional paradigms in H's speech.

Number of members	Age	Category	Form	Gloss
3	2;4.14	noun	*naan*	I.NOM
			naan-te	I-GEN
			naa-n-um	I-DAT-CORD
6	3;0.2	verb	*pokk-oo*	go-IMP
			poo	go.IMP
			poov-un̪n̪u	go-PRS
			pooy-i	go-PST
			pook-um	go-FUT
			pook-uv-aa	go-PTCP-AUX

Mean Length of Utterance (MLU)

The Mean Length of Utterance (MLU) is the number of morphemes per utterance. MLU values serve both as a criterion to determine productivity and a benchmark to analyse the linguistic development of a child. Especially for an agglutinative language like Malayalam, MLU values are important determiners of inflectional development since morphemes are concatenated with high regularity to signal each grammatical category. The MLU count at each recording for both the children is given in Figure 3.2. Although the overall development shows an increase

in the MLU values, we observe that in some sessions the MLU plummets for A. These are noticeable at the ages of 1;10.15 (MLU=1.9), 2;0.16 (MLU=2.9), 2;1.15 (MLU=2.8), and 2;3.16 (MLU=3.7) We also observe a sudden rise at the ages of 2;3.1 (MLU=4.9), 2;5.16 (MLU=6.3), 2;6.1 (MLU=7.1), and 2;7.16 (MLU=6.7). These local variations are more often the result of the child's mood during those particular recording sessions. In the sessions with overall low MLUs, she was either unwell or sleepy. H is seen to have lower MLU values in the earlier sessions compared to A, but as the study progresses, H attains MLU values comparable to that of A as shown in Figure 3.2. A is more verbal than H in general, but the MLUs of both the children do not differ by much at the various ages.

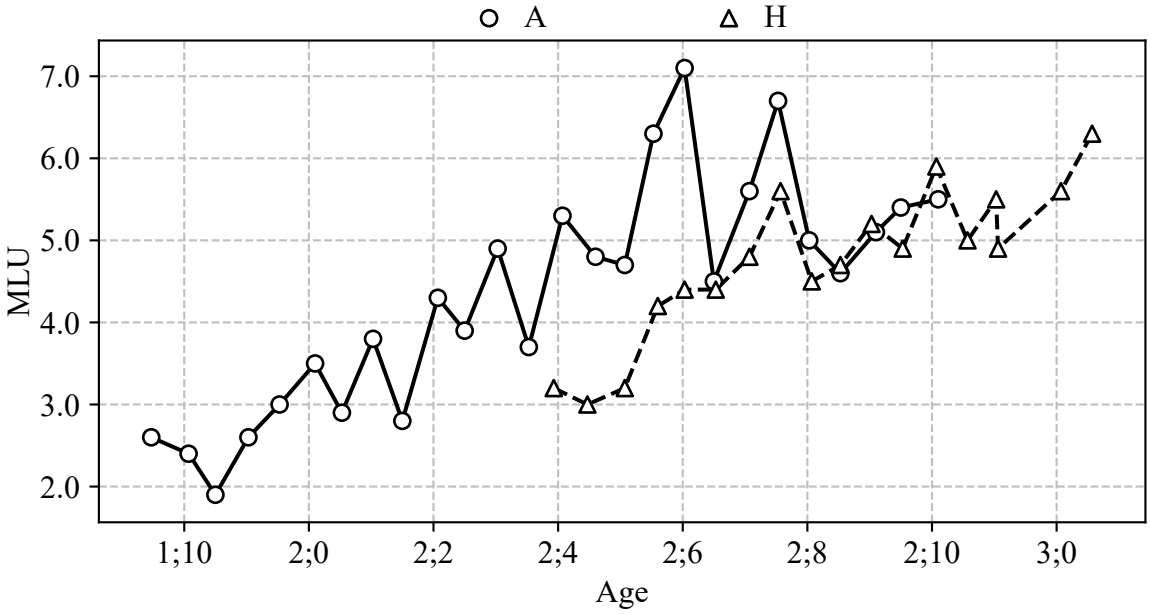


Figure 3.2 MLUs of subjects.

Type and Token Frequencies

The *type frequency* of an inflectional marker is the number of unique stems to which it applies whereas the *token frequency* indicates the number of times an inflectional affix occurs in the child's output. Both *type* and *token frequencies* are markers of productivity as well as indicators of inflectional development. The occurrence of an inflectional marker across a number of stems and the use of the marker in every obligatory context are required to determine whether a) it is productively employed and b) if it has been completely acquired.

The development of vocabulary size and the productive use of inflectional markers go hand in hand. The vocabulary size has a direct impact on the type and token frequencies of

inflectional markers. Figures 3.3 and 3.4 provide the details of A's and H's vocabulary, respectively. Not surprisingly, nouns and verbs constitute a dominant part of the vocabulary in both the children. A has on the face of it more proper nouns in her speech, primarily because she refers to herself by her name instead of using the first person pronoun. Children's pronoun production tokens comprise personal pronouns involving the distinctions inclusive vs exclusive and proximal vs remote, demonstratives, interrogatives, indefinites, possessives, and the reflexive *-ṯaan*. Table 3.3 presents the details of pronoun tokens in A's and H's speech. While there are more nouns in A's speech data, H's data has more verbs than nouns. Although H is older than A, the latter has greater number of types and tokens in her speech. This is not surprising because, in general, A is very verbal than H. Adjectives, adverbs, and postpositions constitute a very small part in both the children's speech. The numerals in children's transcripts, especially the ones signalling more than two entities, are not used correctly by the children. While the children correctly use the numeral *oru* / *oṉṉɨ* 'one' to signal a single entity and *ɾaɳʈɨ* to signal two entities, they use the numerals *muuṉṉɨ* 'three', *ṉaalɨ* 'four', and *aɲɟɨ* 'five' to signal plurality in general without adhering to the distinct count each numeral encodes. Further, while using the numerals greater than one, they omit the plural marker. Examples of such instances involving the production of numerals are given below.

(46) aɲɟɨ paicca (A, 2;4.18)
five paise
aɲɟɨ paiṯa [adult form]
five paise
'Five paise' (showing three one rupee coins)

(47) *muuṉṉɨ ceeʈʈan_ (H, 2;4.14)
three brother-*0PL
muuṉṉɨ ceeʈʈan-maar [expected form]
three brother-PL
'Three brothers' (pointing to a picture of a group of boys)

Most of the verbal and nominal inflections are in place in the children's speech from the very first recording onwards. Figures 3.5 and 3.6 demonstrate the details of A's and H's inflectional production, respectively. Tense and case inflections are the two dominant and obligatory markers in Malayalam. These two categories together constitute the majority of inflectional production in both the children. The chart shows that H has twice as many case inflections as

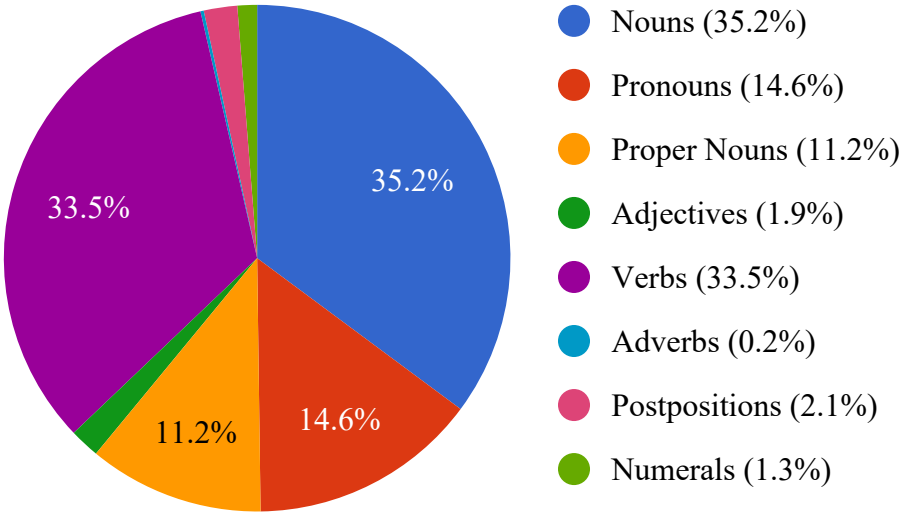


Figure 3.3 A's vocabulary count.

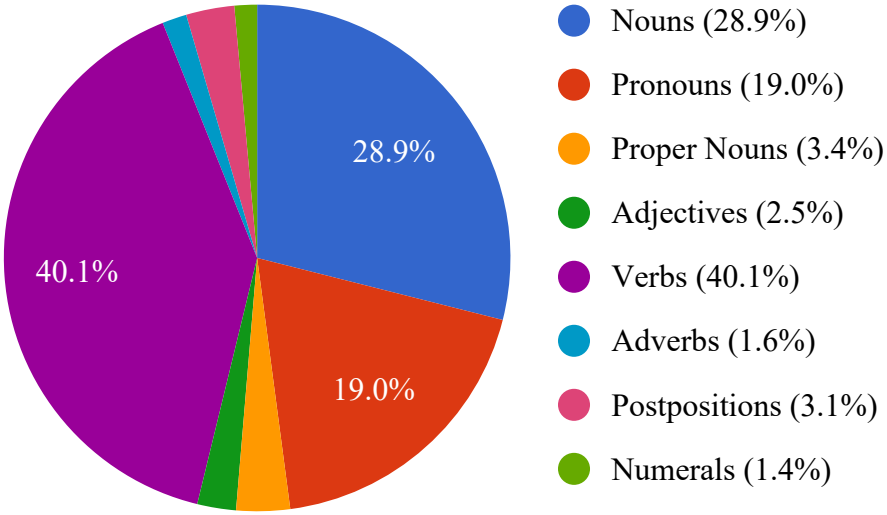


Figure 3.4 H's vocabulary count.

Table 3.3 Pronoun count in A's and H's transcripts.

Category	**A's Transcripts**	**H's Transcripts**
Personal	712	293
Demonstrative	321	64
Interrogative	158	87
Indefinite	22	6
Possessive	28	49
Reflexive	7	-

A. It is to be noted that Malayalam is a topic drop language and there is a difference between the two children in their use of topic-drop. A utilises this feature extensively in her speech and often drops objects as well if they can be recovered from the context of utterance. In contrast, H tends to have overt subjects and does not resort to topic drop as frequently as A does. This, then, is the reason for the difference in the use of case markers between the two children.

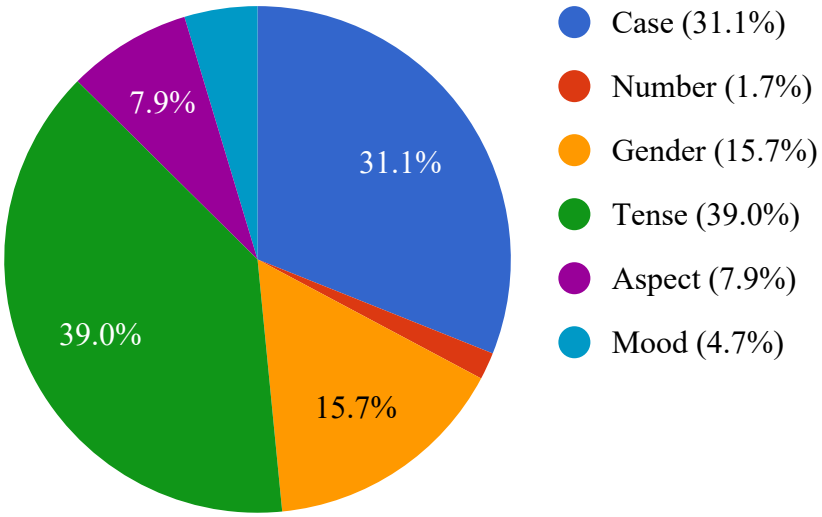


Figure 3.5 A's inflectional production.

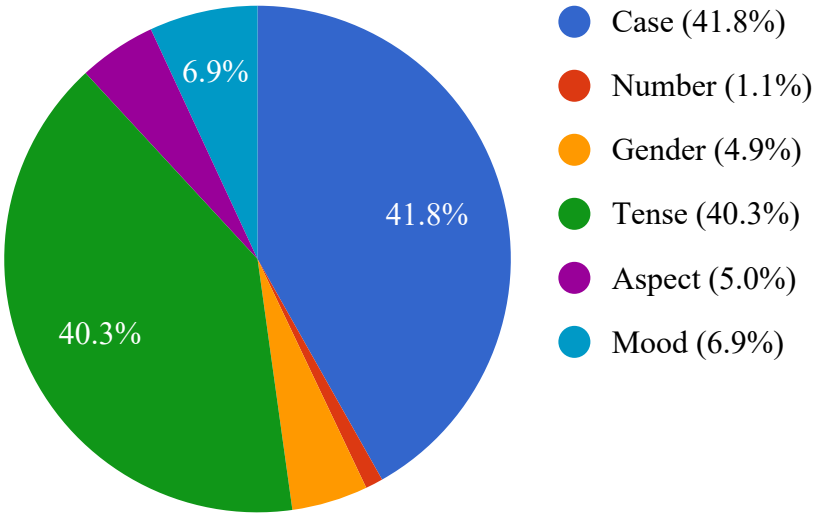


Figure 3.6 H's inflectional production.

3.1.3 Productivity

The notion of productivity implies that the child should display the ability to construct newly inflected forms and that these forms must not be learned by rote. Productivity, then, provides

corroborative evidence for the acquisition of a relevant morphological/morpho-syntactic rule that governs the use and presence of a specific inflectional marker. Overgeneralisations are taken to be instances of the child's productivity since they are revelatory of the process of rule-learning in the child (Berko, 1958). While productivity reveals that the child has induced the relevant rule, the actual occurrence of the inflectional marker in obligatory morpho-syntactic contexts demonstrates the child's mastery and complete acquisition of the inflectional marker. Once the acquisition of the relevant rules is complete, the child produces both the regular and deviant forms correctly.

Most acquisition studies consider productivity as the primary criterion in determining whether or not an inflectional marker has been acquired. There are, however, differences across researchers in terms of the definitional criteria used to determine the productivity of inflectional markers. Dressler, Kilani-Schoch, and Klampfer (2003) define productivity as the ability to form new potential words. Bybee (2001) similarly defines productivity as the extent to which a pattern is likely to apply to new forms. Stephany and Voeikova (2009) consider type frequency as the productivity determining factor. Cazden (1968) suggests that an inflectional marker may be deemed fully acquired if it is used correctly in the obligatory contexts at least 90% of the time in three consecutive samples of data. It is to be noted that the productive use of a marker (i.e., the number of stem types to which it is added) does not entail that it has been fully acquired.

In this thesis, we determine the productivity of a marker by combining two criteria. First, the marker has to be used with three or more stem types. We subject all inflectional markers to this criterion. Second, for those markers that are frequently omitted, we also adopt Cazden's (1968) criterion of 90% occurrence in three consecutive samples. By combining the requirements, we think that greater accuracy is achieved in the determination of productivity of markers that are prone to omissions by the subjects. Nevertheless, a marker can be considered fully acquired only when the number of errors (both commission and omission) visibly decreases and the child's productions approach the adult patterns.

3.2 Cross-Sectional Data

The cross-sectional data used in this thesis, comparatively with the longitudinal data, is taken from the study, IIT Bombay and Leibniz Dream. We only look at the TAM and CGN marking details in the cross-sectional data. In addition to the cross-sectional data, we also make use of

the bilingual Malayalam-English acquisition data (Raghunathan, unpublished data) to see if the patterns observed in the current monolingual data correspond to or deviate from the bilingual data.

Chapter 4

Characteristics of Inflectional Selection in Malayalam

4.1 Introduction

The formal realisation of inflectional markers in Malayalam shows certain properties. First, agglutination; that is, a word-form signalling several grammatical functions is agglutinatively concatenated. Second, phonotactic determination; the form of each inflection is subject to phonotactic, well-formedness constraints. Third, and perhaps the most important property, is the inflectional selection by the head of the word-form. In the following sections, we will see that inflectional selection in Malayalam is either a) phonologically determined or b) morphologically determined.

The selection of an inflectional marker is driven by phonological rules and phonotactic (syllable structure) requirements (e.g. the past tense marker *-i* assigned to heavy syllabled verb stems), or by morphological features of the stem such as *animacy*, *transitivity* etc. (e.g. accusative marker *-e* assigned only to [+ANIMATE] objects). Morphologically selected markers may also require phonological accommodation subsequently. Central to our argument that *mode of selection* is important is that the inflections selected primarily by the phonological component are shown to have less formal complexity and are acquired more readily when compared to the morphologically driven counterparts. In the latter cases, children acquiring inflections in Malayalam have to look beyond the phonology to deduce the inflectional assignment rule. In Chapter 2, we discussed phonological bootstrapping and as Demuth (2001) argues, children have an early awareness of the phonological systems of their native languages and, consequently, a

phonologically determined morpheme selection will be relatively easier to acquire.

4.2 Malayalam Inflections and Agglutination

Malayalam has highly regular inflections which are agglutinatively stacked with a one-to-one correspondence between the inflectional marker and the grammatical feature that it manifests, as can be seen in (48). There are separate morphemes to signal the grammatical categories of gender, number, case, and so on.

(48) sundar-an-maar-ey-um sundar-i-maar-ey-um
beauty-M-PL-ACC-CORD beauty-F-PL-ACC-CORD
'Handsome men and beautiful women.'

Typologically, Malayalam inflections also demonstrate characteristic features such as phonological salience, morphological and semantic transparency, and high productivity, as mentioned in Chapter 2. These properties enable the acquisition of inflectional affixes, unlike a more fusional system (e.g. the English verbal affix *-s*, which signals person, number, present tense, and agreement all melded together).

4.3 Phonotactic Constraints on Inflectional Affixes

In this section, we describe and provide a brief account of the phonotactic rules that are crucial to the inflectional selection and word formation in Malayalam. This will also facilitate our discussion of the inflectional material and its acquisition[1]. K. P. Mohanan (1982, 1986) has proposed the *no-coda hypothesis* based on his observation that native Malayalam speakers parse syllable final consonants as part of the following syllable's onset in order to avoid creating a coda despite the potential violation of the sonority hierarchy of the onset sequence.

K. P. Mohanan (1982, 1986) uses the *pa*-language game to make this argument. According to K. P. Mohanan (1982), in *pa*-language, a dummy syllable *-pa-* is inserted between the actual syllables of the given word. The *pa*-language forms of Malayalam speakers are in contrast to that of native Hindi speakers, who parse consonants into the coda and have access to more complex codas and, therefore, syllables. For example, when asked to convert the word

[1] I have developed a more complete and detailed account of the syllable structure properties. This is not included in the thesis since it is somewhat tangential to the core argument being made here and will serve as a distraction to the main idea of the thesis.

baṉḍʰanam 'binding' into *pa*-language, Malayalam speakers produced *pa-ba-pa-ṉḍʰa-pa-nam* with the nasal sonorant following the inserted syllable rather than preceding it. Hindi speakers create *pa-ban-pa-ḍʰa-pa-nam*, where the nasal is syllabified into the preceding coda. K. P. Mohanan uses this parse as evidence to argue that Malayalam is a no-coda language[2]. His other arguments for analysing Malayalam as a no-coda language are a) Malayalam speakers parse the words like *bʰakṯi* 'devotion' and *campa* 'jasmine' as *bʰa.kṯi* and *ca.mpa*, respectively, as opposed to Hindi speakers who parse them as *bʰak.ṯi* and *cam.pa*, respectively, b) Malayalam speakers judge words like *amma* 'mother' and *kuṯṯi* 'child' to comprise three and four segments, respectively, while Hindi and Marathi speakers judge those words to contain four and five segments, respectively (K. P. Mohanan, 1982; T. Mohanan, 1989).

We disagree with this analysis. Malayalam (like the other Dravidian languages) poses strict restrictions on the contents of the coda. Coda slots are limited to the sonorants – nasals, liquids and glides. Nasals are further restricted in the word-internal coda and are only permitted when they form a member of a homorganic pair (**nṯ**, **mp** etc.), and stop consonants are permitted only when they are part of a geminate (Rajaraja Varma, 1896). The *pa*-form *paba-paṉḍʰa-panam* produced by the Malayalam speakers is, in fact, the form that should be expected, since it respects the homorganicity constraint; the target *pa*-word respects the phonotactic constraints. Had they produced the form *pabanpaḍʰapanam* the resulting coda-onset cluster **np** would indeed have violated Malayalam's homorganicity constraint.

Steriade (1981) notes that in Malayalam, the clusters at the margins of the words do not violate the Sonority Sequencing Principle (SSP) and therefore, there is no reason to expect such violations in medial clusters. She demonstrates with the Romanian equivalent of *pa*-language that one may not assume such language games to be exclusively of the syllable sensitive type. According to Steriade (1981, p. 28-29), "in Romanian, the *pa*-rule inserts after every vowel, regardless of whether the vowel exhausts the rime, a *-p-* followed by a copy of the vowel: *vorbesk* 'I speak' becomes [vop**o**rbepesk] with the *pV* sequence inserted in the middle of both rimes rather than after them".

In another paper, Steriade (1999) argues that the fact that English speakers are divided on parsing the word *lemon* as either *le.mon* or *lem.on* reflects different parses of the same auditory stimulus. She states that syllabic intuitions partly correlate with word-phonotactics and that it

[2]T. Mohanan (1989) argues that the Sanskrit syllable structure is overlaid on the native (Dravidian) one and creates a series of rules/constraints by which a resolution between the conflicting demands of the two grammars are resolved.

is the phonotactics that determines syllable structure and not the other way around. Steriade (2002) proposes that in Malayalam, the mode of parsing relates to the orthographic and word-edge properties of the language and not coda avoidance per se.

Additional evidence against the no-coda hypothesis comes from the phonetic study conducted by Broselow et al. (1997). The study reports that phonetic shortening of both long and short vowels in closed syllables is observed in Malayalam and, further, that there is no significant difference in the duration of consonants after long/short vowels. The fact that vowel duration is shortened when coda increases implies that the coda is part of the rime which comprises both the nucleus and the coda consonant. The findings of the study demonstrate that Malayalam has codas and that they are not resyllabified to the onsets in the surface structure as K. P. Mohanan (1982) and T. Mohanan (1989) argue in their accounts. Given these different arguments, we do not assume that Malayalam is a non-coda language, but rather one with well-articulated coda constraints, as is the norm for Dravidian languages.

The core Malayalam syllable is a V (e.g. the first syllable in *a.ri.vi* 'knowledge') which can be augmented to form the following syllables: (C)V(C) (e.g. the second syllable in *n̪i.ɻal* 'shadow'), (C)V(V) (e.g. the first syllable in *maa.ri* 'moved'), and (C)V(V)(C) (e.g. *miin* 'fish'). Three of the most important constraints that govern the syllable formation in Malayalam are i) tautosyllabic cluster reduction, ii) coda restriction, and iii) syllabic weight constraint. The tautosyllabic cluster reduction constraint dictates that consonant clusters are not permitted in the onset and coda positions. Thus, typically, only single consonants are permitted as onsets and codas.

The coda constraint dictates that only certain consonants can occupy the coda positions of a Malayalam syllable. These are the sonorant consonants **m**, **n**, **ɳ**, **l**, **ɭ**, **r**, **ɻ**, and **y**. In addition to those sonorant consonants, while forming a word, Malayalam phonotactic constraints only allow geminates and homorganic nasal-stop pairs as heterosyllabic consonants in an adjacent coda-onset pair such as **ŋk**, **ɲc**, **n̪t̪**, **mp**. The underlying voiceless stop consonants surface as voiced ones and the clusters become **ŋg**, **ɲɟ**, **n̪d̪**, and **mb**, respectively, phonetically. Thus, coda augmentation is limited to a specific set of consonants which are the sonorants listed above, the first half of a geminate, and the nasal of a homorganic nasal-stop pair. Table 4.1 provides examples of legitimate codas in Malayalam.

Coda consonants that occupy the word-final position require further analysis. Only the sonorant consonants [**m**, **n**, **ɳ**, **l**, **ɭ**, **r**, **y**] are allowed to occupy the coda position at the end of a

Table 4.1 Examples of Malayalam coda consonants.

Syllable Structure	Coda type	Example
CVVC	sonorant [n]	mii**n** ‘fish’
CVC	glide [y]	n̪e**y** ‘ghee’
CVCCV	initial segment of a geminate	ka**t̪**.t̪i ‘dagger’
CVCCV	initial segment of a geminate	ci**p**.pi ‘seashell’
CVCCV	initial segment of a geminate	ca**k**.ka ‘jackfruit’
CVCCV	initial segment of a geminate	mu**ʈ**.ʈa ‘egg’
CNCV	ŋ of ŋk	pa**ŋ**.kɨ ‘share’
CNCV	ɲ of ɲc	ʋa**ɲ**.ci ‘boat’
CNCV	n̪ of n̪t̪	t̪a**n̪**.t̪a ‘father’
CNCV	m of mp	t̪u**m**.pa ‘leucas’
CNCV	ɳ of ɳʈ	mu**ɳ**.ʈɨ ‘’dhoti’

word. For example, *maa**n*** ‘deer’ vs *maa**ʈɨ*** ‘cattle’. The stems that have other coda consonants at the end of the word are re-syllabified to the onset position by the addition of the release vowel [ɨ] which has no phonemic value in Malayalam, and which is also seen to be a Dravidian feature. The tendency to have an open syllable structure is so strong that except for the consonants [m] and [n], the rest of the sonorant consonants can alter between the coda and the onset position at the end of a word, especially in continuous speech, as in example *paal* and *paalɨ* both referring to ‘milk’. Table 4.2 shows the distribution of word-final coda consonants in Malayalam.

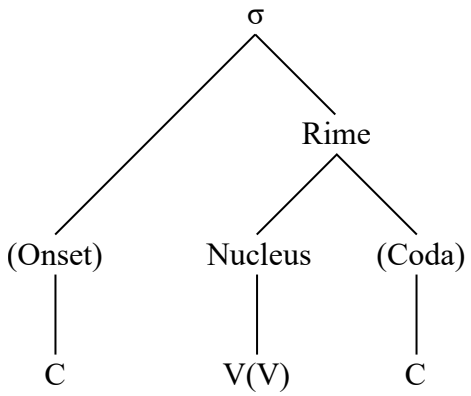
Associated with the coda constraint is the syllable weight constraint. There is a general tendency to break up heavy syllables and resyllabify the coda consonants to the onset positions. This is most visible during the selection of inflectional markers. Vowel markers are selected to resyllabify the heavy-syllabled stems. (49) lists the kinds of syllables that are permitted in Malayalam. (50) shows the structure of the maximal syllable that is permitted in Malayalam.

(49) V
(C)V
(C)V(V)
*(C)(C)V
(C)V(V)(C)

Table 4.2 Word-final coda consonants.

Word	**Permissible**	**Not-Permissible**
miin ‘fish’	miin	miinɨ*
maɳam ‘smell’	maɳam	maɳamɨ*
kuuɳ ‘mushroom’	kuuɳ / kuuɳɨ	-
paal ‘milk’	paal / paalɨ	-
ʈeeɭ ‘scorpion’	ʈeeɭ / ʈeeɭɨ	-
caar ‘gravy’	caar / caarɨ	-
ɲey ‘ghee’	ɲey/ɲeyɨ	-
ʋiiʈɨ ‘home’	ʋiiʈɨ	ʋiiʈ*
paampɨ ‘snake’	paampɨ	paamp*

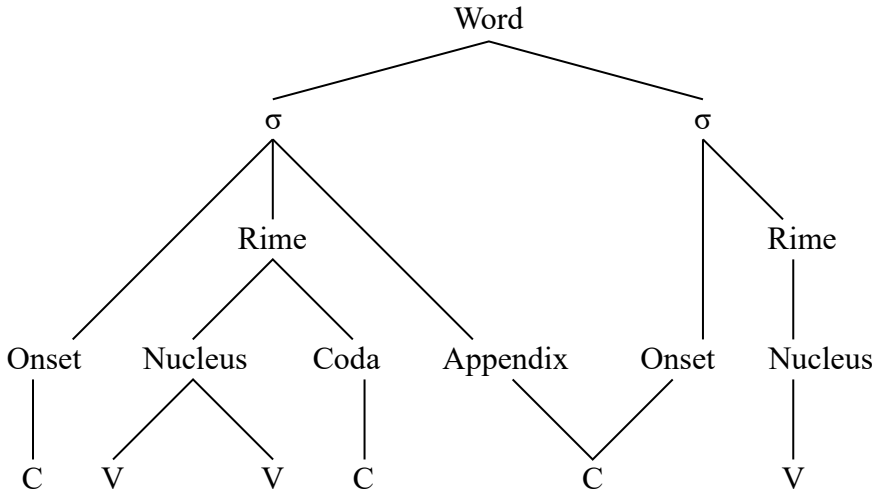
(50) Maximal Syllable



Given the syllable structure presented above, it is necessary to account for the behaviour of geminates. In Malayalam, only voiceless consonants occur as geminates in word-medial positions. The gemination of voiceless consonants in word-medial positions occurs in order to preserve their [−voice] feature and to prevent undergoing voicing. Gemination of voiceless stops occurs at morpheme boundaries as well in order to retain their [−voice] feature as can be seen in the word-forms *vaaɻ-ʈʈ-i* [praise-TR-PST], *ɲeer-ʈʈu* [thin-PST] etc. When this happens at the morpheme boundaries, the first half of the geminate is attached as an appendix directly to the syllable node as can be seen in (51). Thus, forms like *vaaɻʈ.ʈi* and *ɲeerʈ.ʈu* have the structure (51). This analysis deviates from K. P. Mohanan’s (1982, 1986) analysis where he posits the word-final codas as appendices. On the surface of it, these forms appear to be violating the homorganicity constraint. However, the fact that the first half of the geminate is not a part of the

rime since it is an appendix which is attached directly to the syllable node enables the violation of the homorganicity constraint. When the appendix comes between the coda and onset, the homorganicity constraint cannot apply since it applies only to the adjacent coda-onset pairs.

(51)



It is nonetheless necessary to distinguish native Malayalam syllabic structure from that of Sanskrit, given the large scale lexical borrowing from Sanskrit into Malayalam. Although numerous Sanskrit words exist in Malayalam, they are either phonologically accommodated or are infrequent and restricted to formal contexts. Unaccommodated, borrowed Sanskrit words are learned by children during formal learning and do not form a part of the input they hear otherwise. Our analysis also contradicts T. Mohanan's (1989) syllable structure account. T. Mohanan extends K. P. Mohanan's (1986) *no-coda* constraint by proposing that because of the heavy influx of Sanskrit vocabulary in Malayalam, Modern Malayalam has two types of syllable structures (and phonotactics) at play in its phonology. The initial strata of word-formation are governed by Sanskrit's phonotactics which permits coda consonants, while the final stratum, applies Dravidian phonotactics resyllabifying the coda consonant to the onset position even at the expense of violating the SSP (T. Mohanan, 1989).

This analysis is counterintuitive. Quite apart from the fact that it leans on K. P. Mohanan's analysis which itself is unsupported as we saw earlier, there are several reasons not to award such primacy to the Sanskrit phonotactics. Even though Malayalam has borrowed a large number of content words from Sanskrit, it is unlikely that Sanskrit phonotactics is given primacy in Malayalam morphological derivations per se. Compared to the number of content words that has been borrowed, the number of functional morphemes (affixes) that has been borrowed is vastly disproportionate and negligible. Thus, Sanskrit has not been able to alter or affect the grammar of Malayalam in any way (Rajaraja Varma, 1896). The rubric of Malayalam grammar

is still Dravidian in nature.

We find that a majority of these borrowed words are used in extremely formal contexts such as literature or in academia. They are not typically part of the core vocabulary. For example, no one uses the words *n̪iiloolpalam* 'blue lotus' or *ambaram* 'sky' in their colloquial speech unless they are trying to be pedantic. Those Sanskrit words that have transitioned into the every-day vocabulary are phonologically modified to fit the Dravidian (native Malayalam) phonotactic constraints through cluster reduction, devoicing, deaspiration and so on. Examples are *pappanaavan* 'Lord Vishnu' from the Sanskrit *padmanaabʱan* and the forms *vekɨn̪ti* and *ɕakɨn̪ti* from *vyakt̪i* 'person' and *ɕakt̪i* 'power' (the former two phonologically accommodated forms are often employed by Malayali politicians in their speeches). Further examples of such phonological accommodations are given in Table 4.3.

Table 4.3 Examples of phonological accommodation of Sanskrit based words in Malayalam.

Original Word	**Phonologically Accommodated Word**	**Gloss**
st̪ʰalam	s̪alam	'land'
sraavɨ	corakɨ	'shark'
swapnam	s̪oppanam	'dream'
kaaʃʈʰam	kaaʈʈam	'poop'
mooʂt̪ikk-	moot̪t̪ikk-	'to steal'
nyaayam	ɲaayam	'reason/justice'
anyaayam	anniyaayam	'injustice'
d̪viipɨ	d̪iipɨ	'island'
t̪ulsi	t̪uɭas̪i	'holy basil'
kʃama	ʂema	'forgiveness'

It also has to be noted that the Sanskrit borrowings were only accessible to a certain social class of Kerala society and were never a part of the common man's vocabulary. The adoption of some of these words into the general vocabulary is a relatively recent phenomenon and an outcome of the Kerala Education act of 1958 and Kerala State's Literacy Campaign of 1990 which in turn led to the achievement of high literacy rates and the consequent influx of more Sanskrit words into everyday use. Despite the increased *awareness* and *learning* of Sanskrit words, they still undergo phonological accommodation as shown in the examples in Table 4.3. Thus, it seems counterintuitive to give primacy to Sanskrit phonotactic constraints in Malayalam

word-formation processes.

The above-discussed aspects of syllable structure control the resultant surface structure of an inflectional affix, especially the consonant-initial inflections. For example, the addition of the plural markers *-maar* and *-kaḷ* violate the homorganicity constraint on the adjacent coda-onset pairs as can be seen in the underlying forms *kaḷḷa**n-m**aar* [thief-PL] and *mara**m-k**aḷ* [tree-PL], respectively. However, the corresponding surface structures of the resultant word-forms are *kaḷḷa**mm**aar* and *mara**ŋŋ**aḷ*, respectively, which preserve the homorganicity constraint through assimilation of features.

Vowel hiatus between a vowel-final stem and a vowel-initial suffix is resolved by the insertion of a glide, for example (eg: *poo-* + *-i* – *pooyi* 'went'). Consonant final stems show resyllabification of the final consonant into the onset of the suffixal vowel, for example, *paaṭ-* + *-i* – *paaṭi* 'sang'). An open syllable structure can be seen in both *poo.yi* and *paa.ṭi* which have a CVV.CV structure adhering to the syllabic weight constraint. We will see that phonotactics plays a crucial role in the selection of the past tense morphemes and case markers in Chapters 6 and 8, respectively.

4.4 Phonological vs Morphological Selection

As we said at the beginning of the chapter, inflectional selection is determined either phonologically or morphologically. On the one hand, the affixal morphemes conform to the phonological features of the stem endings and the dominant syllable structure. For example, the past tense marker *-i* and not *-ṯu* or *-ṉṯu* is chosen for heavy stems such as *ciṯar-* 'scatter' so that the coda consonant can be resyllabified to the onset position in the resultant word-forms such as *ci.ṯa.ri*, thus adhering to the syllable weight constraint. Thus, the selection of the past tense marker *-i* is driven by the phonological component. This is similar to the past tense marking in English with the alternation between t~d as in *ripped* [t] versus *ribbed* [d].

On the other hand, the selection of certain affixes is sensitive to the lexical features of the stem. For example, the allomorphy of the plural marker varies according to the features [±HUMAN] / [±ANIMATE] of the nominal stem. The [+HUMAN] nouns choose *-maar* as in the example *amma-maar* [mother-PL]. However, the marker is different for [−HUMAN, +ANIMATE] nouns. They are assigned the marker *-kaḷ* as in the example *kiḷi-kaḷ* [bird-PL]. This is different from the plural marker selection in English where the plural allomorphs are phonolog-

ically driven given the [±voice feature] of the stem and whether the stem-final consonant is an alveolar sibilant or not (eg. *kæt-s* [cat-PL], *læd-z* [lad-PL], and *bʌs-iz* [bus-PL]).

4.5 Acquisition and Mode of Affix Selection

As seen in previous chapters, Malayalam inflectional markers are highly productive, phonologically salient, morphologically and semantically transparent, and exhibit biuniqueness. Thus, children acquiring Malayalam inflections are working with a system that enables learning because the morphology is mostly transparent. Young children have been shown to demonstrate competence with the phonotactics of the language they are hearing even before they begin their actual speech production (Guasti, 2002). Therefore, the phonologically determined inflectional assignment ought to be acquired more easily than those that are determined by the lexical features of paradigmatic classes. We propose that *mode of selection* will play a crucial role in the acquisition of inflections and in later chapters we will test the validity of this proposal.

Chapter 5

The Grammar and Acquisition of Malayalam Verbal Inflections

5.1 Verbal Inflections in Malayalam

Verbs constitute the grammatical core of an utterance since a) they select their arguments and assign them case and grammatical roles b) they determine the structure of the clause, and finally, c) they are often the locus of multiple grammatical features such as tense, aspect, mood, and agreement. Verbs across many of the world's languages realise grammatical information through inflections. Hence, the acquisition of verbal inflections is an area of great interest and is fundamental to furthering our understanding of several key issues in language acquisition.

The major verbal inflectional categories in Malayalam are tense, aspect, and mood (TAM). In this section, we discuss the three main categories of tense (present, future, and past), two aspects (perfective and imperfective), and three moods (imperative, conditional, and optative) which are signalled by individual inflectional markers. These above-discussed categories are only a subset of the inflected verb forms in Malayalam. We are only looking at those inflections that are found in colloquial speech and in turn, occurs in the children's input. It must be noted that verb stems in Malayalam are bound stems in general, except in the bare-imperative forms (e.g. *paaṭi* 'sing!', *keeḷkki* 'hear!'), and always carry affixal material, even when non-finite. The template of a Malayalam verb form when the TAM affixes are concatenated together is **stem**+**tense**+(**aspect**/**mood**) as shown in the examples below.

(52) ʋa-n̪n̪u [come-PST]

(53) ʋa-n̪n̪-iʈʈi [come-PST-PFV]

(54) ʋa-n̪n̪-aal [come-PST-COND]

All the verbal inflections discussed here, except the past tense morpheme *-Tu*, are vowel-initial and therefore syllabic and phonologically transparent. Vowel hiatuses are resolved by consonant epenthesis and stem-final consonants invariably resyllabify as suffixal onsets. They apply exceptionlessly across all the verbs. This makes them highly productive and being semantically and morphologically transparent as well, eminently acquirable.

Cross-linguistically, verbs can inflect for tense, aspect, mood, voice, and agreement. In the following discussion, we provide a brief overview of these grammatical features to better understand the verb in Malayalam. Tense may be said to be "the grammaticalisation of location in time" (Comrie, 1985, p. 9). For example, the *-ed* in the sentence '*She cook-**ed** rice*' signals that the action happened in the past with reference to the time of speaking and is, therefore, the past tense marker. Verbs which encode tense and agreement (with the subject) are called *finite* verbs (e.g. *cook-s*, *cook-ed*) and are otherwise said to be *non-finite*, which include infinitives, gerunds, and present and past participles (e.g. *break-ing*, *brok-en*). Many languages separate the time continuum grammatically into the present, the past, and the future. Dravidian languages, in general, as well as others like Romance languages, Turkish, Korean etc. grammaticalise three distinct tense forms. However, not all languages employ all three tenses. Many languages only make a two-way distinction between past and non-past in which the latter covers both the present and the future times (e.g. Japanese), or between future and non-future (e.g. Greenlandic, Quechua etc.), or grammaticalise some tenses and periphrastically indicate others (Germanic languages). More rarely, there are languages that encode degrees of remoteness in the grammatical feature of tense (eg. Bantu languages of Africa, Australian aboriginal languages and native languages of the Americas) making elaborate distinctions with various cut-off points such as remote, recent, one day or a few hours etc. (Comrie, 1985, p. 87). Equally, there are languages that lack the tense grammaticalisation feature altogether and are called *tenseless* languages (e.g. Chinese, and Burmese). Agglutinating languages indicate tense distinctions most often by employing verbal inflections.

Malayalam makes a three-way tense distinction, namely the present, the past, and the future. The present tense is marked invariably by the suffix *-un̪n̪u* and the future tense by the suffix *-um*, as can be seen in (55) and (56). Both the present and the future tense markers are *phonologically* accommodated with the stem. In Table 5.1 we show examples of the present and

future tense forms with consonant and vowel-initial stems.

(55) amma pook-unnu
mother.NOM go-PRS
'Mother goes.'

(56) amma naaɭe pook-um
mother.NOM tomorrow go-FUT
'Mother will go tomorrow.'

Table 5.1 Present tense and future tense forms in Malayalam.

Gloss	Stem	Syllable Structure	Present	Future
'see'	kaaɳ-	CVVC	kaa.ɳun.nu	kaa.ɳum
'cry'	ka.ɾa-	CV.CV	ka.ɾa.yun.nu	ka.ɾa.yum
'die'	caa-	CVV	caa.kun.nu	caa.kum

Malayalam has two past tense markers *-i* and *-Tu*, and the latter is realised as two allomorphs *-ṯu* and *-ṉṯu*. In Malayalam, past tense marking is determined in one of two modes. In underived (lexically primitive) verbs, phonology and phonotactic constraints control/determine the form of the inflectional suffix. In deadjectival and denominal (derived) verbs, past tense inflection is determined by the lexical feature [±TRANSITIVE]. Thus, in the latter lexical class, affix selection is not phonologically motivated. Past tense marking is a much debated and complex topic in Malayalam inflectional grammar (Asher & Kumari, 1997) and the acquisition of past tense marking is particularly interesting owing to this complexity, as we will see in Chapter 6 in greater detail. Here we limit ourselves to this brief description of the past tense morphemes.

A second time-centred grammatical category associated with verbs is *aspect*. Comrie (1976) defines aspect as a way of viewing the internal temporal nature or constituency of a situation. *Aspect* thus conveys other temporal information of the action or event, such as its duration, or state of completion, or frequency, as it relates to the time of action. It describes whether an event occurs at a single point of time, over a continuous period of time, or at discrete points in time, while tense indicates the event's location in time. Thus, with reference to the temporal features of the event, tense can be said to encode the 'when' while aspect encodes the 'how'. A basic aspectual distinction seen in a number of languages is that between the *perfective* (which marks the completion of events) and the *imperfective* (which marks the continuity of

events). Other aspects include the *progressive* (which indicates ongoing events) and the *habitual* (which signals repetitive events). The *-ing* morpheme in the sentence '*She is cook-**ing** rice*', indicates that the action is not completed and is in progress and hence marks the progressive aspect, while the form of the verb *to be* marks the present tense. This happens in Malayalam as well, where the present tense marker *-u̪n̪n̪u* can also convey the progression of events. Likewise, the future marker *-um* can signal habitual actions as well. Cross-linguistically, the interaction between the features of tense and aspect is quite complex and inflectional affixes often signal information about both rather than achieve a unique one-to-one mapping.

In Malayalam, both inflections and verbal auxiliaries can be used to encode the verbal aspect. For example, the lexical auxiliaries[1] like *kaɭay-* 'lose', *vekk-* 'put' etc. signal the perfective aspect while the auxiliary *vaɾ-* 'come' signals the imperfective aspect (Asher & Kumari, 1997). In the following, we only look at inflectional aspectual marking. The primary aspectual distinction in Malayalam is between the perfective and the imperfective. While the perfective aspect signals the completion of the event, the imperfective signals an ongoing action that is conceived as happening continuously or repetitively along a timeline. The inflectional marker *-iʈʈɨ* encodes the perfective as can be seen in (57). The perfective also allows the temporal sequencing of events.

(57) ɾaaman siit̪ay-e kaɳʈ-**iʈʈɨ** pooy-i
rama.NOM sita-ACC see-PFV go-PST
'Rama left after seeing Sita.'

The imperfective aspect is marked by combining the marker *-uka* with the copula *aaɳɨ* 'to be'. It denotes an ongoing action that is conceived as happening continuously or repetitively along the timeline.

(58) kuʈʈi paʈʰiykk-**uka-aaɳɨ**
child.NOM study-PTCP-AUX.IPFV
'The child is studying.'

The grammatical category of *mood* and *modality* expresses the speaker's attitude to the proposition. Some of the main moods are the indicative (statement or fact), imperative (command, request, or prohibition), conditional, subjunctive (wish, possibility), interrogative (ques-

[1]Malayalam, like the other Dravidian languages, employs a variety of lexical verbs to function as auxiliaries. These include the reflexive auxiliary *koɭɭ-*, and others such as *t̪aɭɭ-* 'push', *vekk-* 'put', *iɾiykk-* 'sit', *n̪okk-* 'look', *kaaɳ-* 'see' etc. used to signal various aspects by removing their lexical meaning (Asher & Kumari, 1997).

tion) and optative (hopes, wishes, commands). The sentence '*Cook the rice!*', is a command and is said to be in the imperative mood. Whether the mood is morphologically encoded varies cross-linguistically. In Malayalam, while the indicative is not morphologically encoded, other moods are. Asher and Kumari (1997, p. 304) note that "the indicative forms of the verb are those which carry no modal overtones. The indicative is obligatorily marked for tense."

The inflectionally marked moods in Malayalam are the imperative, the conditional, and the optative. Other moods like intentional, debitive etc. are signalled by the modal *-aɳam* and permission, potential etc are signalled by the modal *-aam* (Asher & Kumari, 1997). Except for the bare stem imperatives, all the mood inflections which are discussed below apply across the board without exceptions and are hence highly productive.

The major imperative markers are *-oo*, *-ee*, and *-aan* and indicate increasing degrees of emphasis in command[2]. It has to be noted that in addition to these markers that signal the imperative mood, there are bare stem imperatives. While the vowel-final stems do not undergo any change in their surface structure during the formation of their corresponding bare imperative forms (e.g. *poo* 'go!', *iri* 'sit!' etc. from *poo-* 'go', and *iri-* 'sit', respectively), the consonant-final stems like *uuɾ-* 'remove', *paaʈ-* 'sing' etc. form their bare-imperatives *uuɾɨ* and *paaʈɨ*, respectively, with the aid of the release vowel [ɨ] in order to adhere to the phonotactic constraints (see Chapter 4) in Malayalam. Two notable exceptions are the sonorant consonant-ending verbs *vaɾ-* 'come' and *t̪aɾ-* 'give' which form their bare stem imperatives as *vaa* 'come!' and *t̪aa* 'give!', respectively, patterning with the vowel-ending stems instead of forming the imperative with the addition of the release vowel [ɨ]. The bare imperative forms can be considered as instances of zero (null) affixation.

(59) n̪ii pokk-oo/pokk-ee/pook-aan
you.SG.NOM go-IMP
'Go!'

The conditional denotes an event whose realisation is dependent upon another event. It is indicated by adding the inflection *-aal* to the verbal stem. The conditional affix is not a frequently occurring marker unlike the imperatives or the tense markers. The conditional is also marked non-inflectionally using *eŋkil*.

[2]In addition to these markers, the morphemes *-in* and *-aalum* also signal the imperative. However, these are very much formal and not present in colloquial speech (Asher & Kumari, 1997). These are not discussed in this thesis.

(60) amma koṭuṯṯ-aal kuṭṭi paal kuṭiykk-um
mother.NOM give-COND child milk drink-FUT
'The child will drink the milk if the mother gives it.'

The optative expresses hopes, wishes, commands etc. and is signalled using *-aṭṭe* in Malayalam.

(61) ɾaaman paray-aṭṭe
rama.NOM talk-OPT
'Let Rama talk.'

Many languages also have agreement morphemes that encode the properties of the subject and/or object of the verb. These typically express the features of person, number, and gender (or phi-features) of the nominal on the verb. For example, in the sentence, '*She cook-s rice*', the morpheme *-s* indicates that the subject is 3rd person, singular and thus, shows agreement with the subject for person and number (but not for gender). There are also languages where the verb does not agree with the subject and in Malayalam (as in Japanese and Chinese) there is no subject-verb agreement. In this, Malayalam differs from the other Dravidian languages (especially, Tamil) where the agreement is rich and morphologically visible.

5.2 Acquisition of Verbal Inflections

Given the information that resides in or centres around the verb, as we saw above, the acquisition of verbal inflections across the world's languages has been the focus of much research. Verbal inflectional systems are variegated, and it is both interesting and important to see how children acquire the various grammars if we are to understand how language becomes manifest in the growing child. In particular, many European languages as well as non-European languages (Japanese (Clancy, 1985), Qanjobal (Mateo Pedro, 2015), Turkish (Haznedar & Ketrez, 2016) etc.) have been well-researched and documented from a developmental perspective. However, linguistically rich India is much less documented in this regard and there are only a few studies scattered over a few languages.

In this section, we will focus on published work on the acquisition of verbal inflections in Tamil since it is a sister-language and most similar to Malayalam in grammatical terms. Often, the morphological patterns require reference to Tamil as well as their joint history to enable a clearer account of the grammar of Malayalam. In the following, we survey studies on the ac-

quisition of Tamil verb morphology by Raghavendra and Leonard (1989), Lakshmanan (2006), and Sarma (2014).

Raghavendra and Leonard (1989) have investigated the acquisition of verbal inflections in three monolingual Tamil speaking children aged 2;2, 2;6, and 2;7 over a period of four weeks each. The results show a high percentage of use of tense inflections and different aspectual markers in obligatory contexts. The few errors are, primarily, substitution errors as may be seen in (62) where the child substitutes the present tense agreement marker for the future.

(62) Adult: ni eppaDi skulu-kku po-v-e:?
you how school-DAT go-FUT-2SG
'How will you go to school?'
Child: skuTTer-le po-r-e:n (*po:ven* expected)
Scooter-LOC go-PRS-1SG
'I am going in the scooter.'

(Raghavendra & Leonard, 1989)

The authors suggest that this particular substitution error could be because Tamil permits the use of the present tense marker to indicate the future in certain contexts which might have triggered the erroneous productions in the kids, though it is not clear that such an interpretation is possible in this example. This study shows that Tamil speaking children as young as 2;2 use a variety of inflections productively. According to the authors, the typological features of Tamil enable such early development, and these include a) morphological and phonological transparency of the morphemes, b) basic SOV word order, and c) verbal inflections being suffixes rather than prefixes and thus consistent with the verb-final typology. The authors suggest that children might be sensitive to such consistencies in typology.

Lakshmanan (2006) analyses the data of two monolingual Tamil speaking children, a girl, observed between the ages 1;9.16 to 1;11.20 (five recordings), and a boy at the age of 2;0 (one recording). The children have omission errors in obligatory contexts in their use of verbal inflections even though they employ tense and agreement inflections early enough. Lakshmanan proposes an alternative analysis by correlating the omission errors with the overall word length constraints for the individual child. One of the children's word productions were maximally bisyllabic. The author suggests that the omission errors may be a consequence of this performance constraint. However, her analysis also shows that children's competence does include adult-like knowledge of inflections.

Lakshmanan notes that the children tended to use the verbal participle instead of the finite past tense forms. Since the verbal participle is homophonous with some of the past tense allomorphs in Tamil, she suggests that the children could actually be using past tense forms without the overt subject agreement. The author goes a step further to also say that the Tamil children may actually be going through a stage in which they assume that Tamil is like Malayalam where verbs inflect for tense, aspect and mood without subject agreement as in (63) where the child is observing an insect.

(63) Child: puucci po-gi (P, 1;9.26)
insect.NOM go-PRS
Adult: puucci poo-gi-di (Colloquial)
insect.NOM go-PRS-3SG [adult form]
'An insect goes.'

(Lakshmanan, 2006, p. 196)

The Tamil speaking children's grammatical competence in her opinion is not unlike that of the adults despite the fact that the underlying target inflections are difficult to recover because of the morphophonemic and assimilation processes in the adult forms (Lakshmanan, 2006).

However, Lakshmanan's suggestion that Tamil children may be going through a stage in which they assume that Tamil is an agreement-less language like Malayalam cannot be a possible explanation since children are seen to set their target parameters to the correct value from very early ages onwards (Guasti, 2002). Thus, there is no reason for the child to set the value according to Malayalam and then later revise it to Tamil since the input forms contain the obligatory agreement forms in plenty. Besides, cross-linguistic data have shown that agreement comes early in children's productions (Guasti, 2002, 1993). Moreover, Sarma (2014) shows that early Tamil shows adult-like competence in the use of agreement markers.

Sarma (2014) also supports the full-competence hypothesis by looking at the data of a Tamil speaking child between the ages 0;9 and 2;9 from the CHILDES database. In a language like Tamil with rich inflectional morphology, the verbal complex contains most of the crucial grammatical information such as agreement, tense, aspect, mood, reflexivity and voice. Thus, the verbal system of inflections is critical to language acquisition. The author shows that the child's utterances comprise samples of the various constructions in Tamil and that inflections are present from an early age onwards. The child is seen to be using different kinds of predicates together with their associated case assignment and agreement features and is capable of

distinguishing between types of predicates through the use of default (66), and real agreement (64), (65).

(64) naa ayi-v-een (V, 2;4.5)
(naan azhu-v-een) [adult form]
I.NOM cry-FUT-1SG
'I will cry.'
(Sarma, 2014, p. 113)

(65) ee-kku mammam veeN-um (V, 2;9.4)
(ena-kku mammam veeND-um) [adult form]
I-DAT food.NOM want-3SG
'I want food.'
(Sarma, 2014, p. 114)

(66) ee-kku kulur-du (V, 2;1.18)
(ena-kku kulur-u-du) [adult form]
I-DAT cold-PRS-3SG
'I am cold.'
(Sarma, 2014, p. 116)

This indicates that the grammatical competence of the child is not really different from the grammatical competence of a native adult speaker (Sarma, 2014). In Sarma's earlier study (1999) on the development of Tamil verbal inflections, she demonstrates that even in very young children (less than two years old) the agreement morphemes are in place.

From these studies, we can see that verbal inflections are in place early enough in child Tamil. One of the studies argued that the typological features enable such early acquisition. These studies lead us to expect certain patterns in the acquisition of Malayalam verbal inflections and to explain those patterns in terms of both typological and language specific features. In the rest of this chapter, we discuss each TAM inflection present in the data. This will allow us to understand the evolving grammars and the principles that govern them.

5.3 Acquisition of Malayalam TAM inflections

In the preceding discussion of verbal inflections we discover that, predominantly, the inflectional markers apply across the board. The main accommodation between the stem and the affix

is governed by phonotactics or syllable structure constraints and involve resyllabification or glide epenthesis. Only the past tense markers -*ṯu* and -*ṉṯu* are partially determined by the lexical features of the verb. Given the previous studies and this highly productive nature of affixation, we expect that the children should be able to induce the rules of inflectional morphology early. We would also expect fewer errors and greater type and token diversity. In the following, we discuss the production patterns of each of these verbal inflections in the speech of the two subjects in the study[3]. As we discuss each inflection, it may be noted that the figures show the token frequencies of each TAM inflection in the children's transcripts[4] and the utterances in (67) - (99) show how each inflection is employed by the children.

The data in Table 5.2 provides a summary list of the type and token frequencies of the different TAM inflections attested in the acquisition data across all the transcripts. As we will see shortly, there are no omission errors attested for any of the TAM inflections. Thus, our criterion for productivity only includes the use of the marker with three or more unique stems rather than the usual criterion of 90% occurrence in obligatory contexts in three consecutive recordings (Cazden, 1968) for the markers that are omitted.

5.3.1 Tense

A summary of the production of the present tense marker -*uṉṉu* in terms of the number of tokens in each transcript is given in Figure 5.1. There are no omission errors attested for the marker. In A's data, 51 unique types and 173 tokens are attested for the present tense morpheme. The morpheme -*uṉṉu* is present once in the first transcript (1;9.14) and is seen continuously after 2;0.16 in the transcripts. The type and token frequencies of the marker also increase progressively with age as may be expected. At the age of 2;1.15, A uses the marker with three different stems. A's type frequency shows a subsequent increase and she uses the inflection with 13 different stems at the age of 2;7.2. The productivity criterion is met at the age 2;1.15.

[3]It has to be noted that A has been very much loquacious compared to H during the recording sessions. Consequently, there are greater number of types and tokens for the affixes in A's transcripts in general. However, the MLU values of the children at the corresponding ages or the percentages of each inflection in their transcripts do not show much difference from each other.

[4]The graphs are without any pattern since the token frequencies of different inflections vary with and across each session. Also, there is a general dip in the number of tokens at the ages 1;10.15, 2;0.16, 2;1.15, and 2;3.16. These local variations are more often the result of the child's mood during those particular recording sessions, where she has been either unwell or sleepy, as mentioned in Chapter 3.

Table 5.2 TAM inflections in A's and H's speech.

			A's Transcripts		H's Transcripts	
	Category	**Marker**	**Frequency**		**Frequency**	
			Type	**Token**	**Type**	**Token**
T	Present	*-uṉṉu*	51	173	33	90
	Future	*-um*	64	485	24	76
	Past 1	*-i*	44	260	30	100
	Past 2	*-ṯu*	28	172	24	84
	Past 3	*-ṉṯu*	15	214	9	81
A	Perfective	*-iṯṯi*	38	122	7	16
	Imperfective	*-uka* + *aaɳi*	24	143	14	37
M	Optative	*-aṯṯe*	13	33	15	29
	Conditional	*-aal*	2	3	2	2
	Imperative 1	*-aan*	14	38	2	2
	Imperative 2	*-ee*	13	52	8	28
	Imperative 3	*-oo*	11	30	9	13

(67) koŋŋan-e kaaɳ-iṉṉ-illa (A, 2;1.15)
monkey-ACC see-PRS-NEG
koraŋŋan-e kaaɳ-uṉṉ-illa [adult form]
monkey-ACC see-PRS-NEG
'The monkey is not seen.'

(68) kaakk-eṭe bukkɨ aaɳ-eṉṉɨ ṯooṉṉ-uṉṉu (A, 2;9.3)
crow-GEN book be.PRS-COMP feel-PRS
[enikkɨ] ṯooṉṉ-uṉṉu [iṯi] kaakkay-uṭe bukkɨ aaɳ-eṉṉɨ [adult form]
[I.DAT] feel.PRS [this] crow-GEN book be.PRS-COMP
'I think it is the crow's book.'

The inflection occurs in 33 types and 90 tokens in H's transcripts. The marker is present continuously in all the transcripts except in the earliest two at 2;3.28 and 2;4.14. H's first productive use of the marker is attested at the age of 2;6.16 where he uses it with three different types. At the age of 2;10.17, he is seen to be using the marker with 10 different stems.

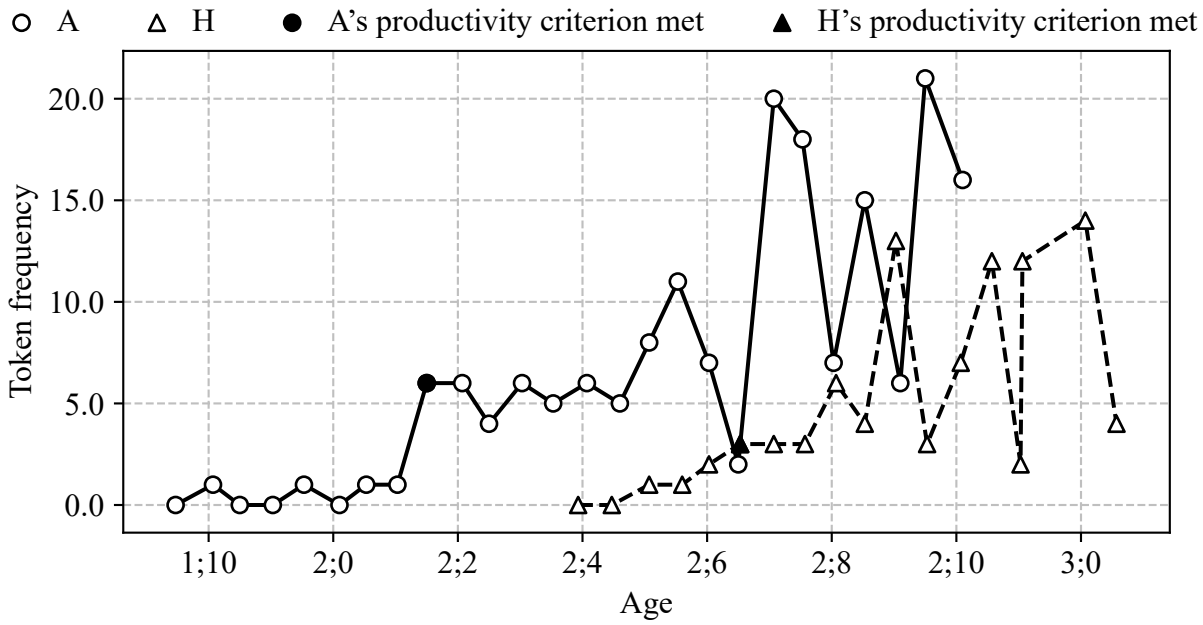


Figure 5.1 TAM inflections - Present *-uɳɳu*.

(69) kaaɳ-iɳɳ-ill-ee? (H, 2;8.2)
see-PRS-NEG-Q
[ceecciy-kki] kaaɳ-uɳɳ-ill-ee? [adult form]
[sister-DAT] see-PRS-NEG-Q
'Doesn't sister see it?'

(70) ke-ɳɳi oŋŋ-uɳɳi (H, 3;0.2)
lie-PTCP sleep-PRS
[accan] keʈa-ɳɳi oraŋŋ-uɳɳu [adult form]
[dad] lie-PTCP sleep-PRS
'Dad is sleeping.'

The future tense marker *-um* is present in the very first session (1;9.14) itself in A's productions. Across all transcripts, A has 64 types and 485 tokens for the future marker *-um*. The first productive use of the marker in A's transcripts is attested at the age of 1;11.1 where she employs the marker with three different types. At the age of 2;7.16, the marker has an attested type frequency of 24.

(71) kaɳɳ-ii kuʈʈ-aani ʋimaanam ʋay-um (A, 1;10.2)
eye-LOC stab-INF aeroplane.NOM come-FUT

ʋimaanam	[n̪amm-uʈe]	kaɳɳ-il	kuʈʈ-aan	ʋar-um	[adult form]
aeroplane.NOM	[we-GEN]	eye-LOC	stab-INF	come-FUT	

'The aeroplane will come to stab (us) in the eye.'

(72)

iicc-eem	koʈuk-um	pa-ɳɳi	pook-um	(A, 2;5.16)
fly-CORD	mosquito-CORD	fly-PTCP	go-FUT	
iiccay-um	koʈuk-um	paraɳɳ-i	pook-um	[adult form]
fly-CORD	mosquito-CORD	fly-PTCP	go-FUT	

'The fly and the mosquito will fly away.'

H produces the future marker *-um* with 24 different verb types over 76 tokens. The marker is attested at least once in all the transcripts except at the age of 2;4.14. The marker meets the productivity criterion at the age of 2;5.18 where H uses it with four different verb stems. The marker has the greatest type frequency at the age of 2;9.1 where he uses it with nine different stems. Figure 5.2 shows the distribution of the token frequencies of the marker in the children's transcripts.

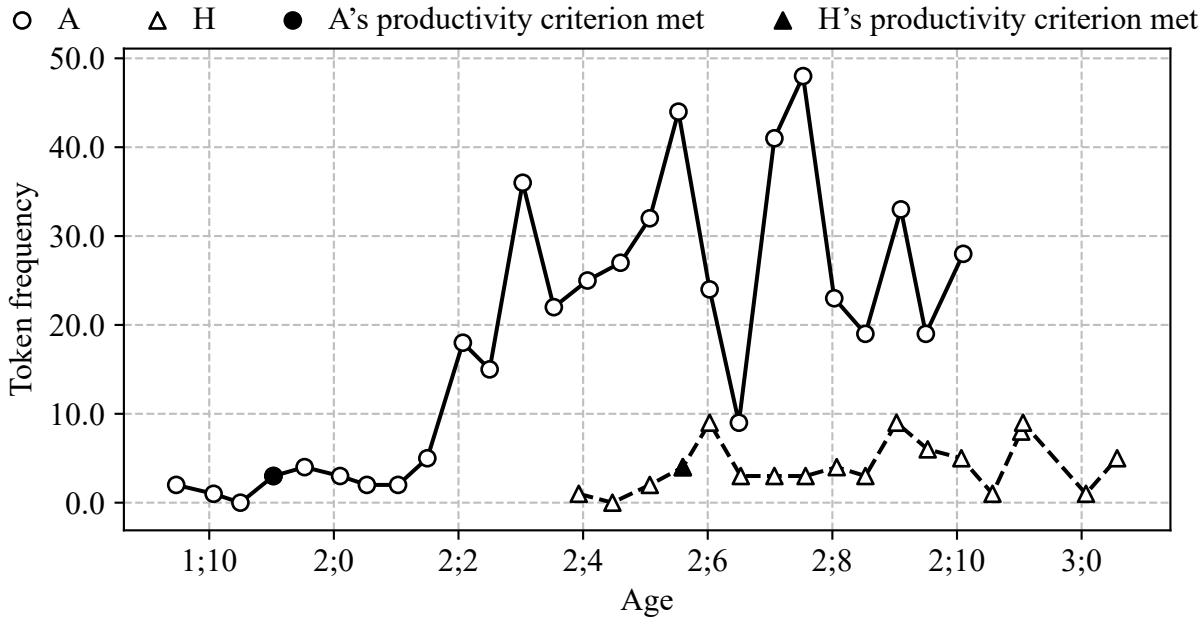


Figure 5.2 TAM inflections - Future *-um*.

(73) ʈaaɻe ʋiiɻ-um (H, 2;6.16)
down fall-FUT
[ɲaan] ʈaaɻe ʋiiɻ-um [adult form]
[I.NOM] down fall-FUT
'I will fall down.'

(74) kiir-i kaḷay-um (H, 2;11.1)
tear-PTCP lose-FUT
[ɲaan] [aʈi] kiir-i kaḷay-um [adult form]
[I.NOM] [that] tear-PTCP lose-FUT
'(I) will tear (it).'

The past tense marker *-i* occurs with 44 verb types and 260 tokens in A's transcripts, and with 30 verb types and 100 tokens in H's transcripts. The productivity criterion is met at 1;9.14 for A and at 2;4.14 for H, where they use the marker with three different stems.

(75) aʈi paaʈ-i (A, 2;10.2)
it.NOM sing-PST
'It sang.'

(76) ootɨ-kaaɾ-an pooy-i (H, 2;3.28)
auto-PL-M.NOM go-PST
ooʈʈoo-kaaɾ-an pooy-i [adult form]
auto-PL-M.NOM go-PST
'The auto-rickshaw driver left.'

The past tense variant *-ʈʈu* occurs with 28 types and 172 tokens in A's productions and 24 types and 84 tokens in H's. A reaches type frequency of three, meeting the productivity criterion, at 2;2.2 and H does so at 2;5.2.

(77) kuɲɲi puucca-kuɲɲ-um ʋalya puucca-kuɲɲ-um ca-ʈʈu (A, 2.9.15)
small cat-baby-CORD big cat-baby-CORD die-PST
'The big kitten and the small kitten died.'

(78) boo⟮ eʈu-ʈʈ-illa (H, 2;6.16)
ball take-PST-NEG
[ɲaan] boo⟮ eʈu-ʈʈ-illa [adult form]
[I.NOM] ball take-PST-NEG
'(I) didn't take the ball.'

The third past tense variant *-ɳʈʈu* is seen in 214 instances involving 15 verb stems in A's transcripts. The marker inflects nine stems for a total of 81 tokens in H's transcripts. A begins to employ the marker productively with five different stems at the age of 1;11.1. In H's transcripts, the marker *-ɳʈʈu* has the type frequency of four at the ages 2;5.2, where he meets the productivity criterion.

(79) ela pa-cc-ɨ kayi-ɲɲu (A, 1;11.1)
leaf pluck-PTCP finish-PST
[ɲaan] ela pari-ccɨ kaɻi-ɲɲu [adult form]
[I] leaf pluck-PTCP finish-PST
'(I) am done plucking leaves.'

(80) ʈaaɻe ʋii-ɳɨ (H, 2.8.16)
down fall-PST
[ɲaan] ʈaaɻe ʋii-ɳu [adult form]
[I] down fall-PST
'I fell down.'

Past tense markers are a key issue in any discussion of Malayalam verbal inflections. We return to the grammar and acquisition of past tense marking in Chapter 6 since there are interesting consequences for acquisition.

5.3.2 Aspect

Children's tokens of the perfective and imperfective aspectual inflections are shown in Figures 5.3 and 5.4, respectively. In A's transcripts, aspect marking does not really surface until 2;0.3 when she uses the perfective marker. The first instance of the imperfective suffix is at 2;1.1. By the end of the period of recording, we see 38 types and 122 tokens of the perfective suffix and 24 types and 143 tokens of the imperfective, respectively. The productivity criterion is met for the perfective marker at 2;2.2 with three types and the imperfective at 2;1.1 with four

types.

(81)	koŋŋ-an	poo-y-iʈʈɨ	akkriim	kayikk-aam	(A, 2;4.18)
	monkey-M.NOM	go-PTCP-PFV	icecream	eat-MOD	
	koɾaŋŋ-an	poo-y-iʈʈɨ	aiʂkriim	kaɻiykk-aam	[adult form]
	monkey-M.NOM	go-PTCP-PFV	icecream	eat-MOD	
	'Let's eat ice cream after the monkey leaves.'				

(82)	amma		paalɨ	koɽukk-uʊ-aa	(A, 2;6.15)
	mother.NOM		milk	give-PTCP-AUX.IPFV	
	amma	[ʊaaʊay-kkɨ]	paalɨ	koɽukk-uʊ-aa	[adult form]
	mother.NOM	[baby-DAT]	milk	give-PTCP-AUX.IPFV	
	'Mom is feeding (the baby).'				

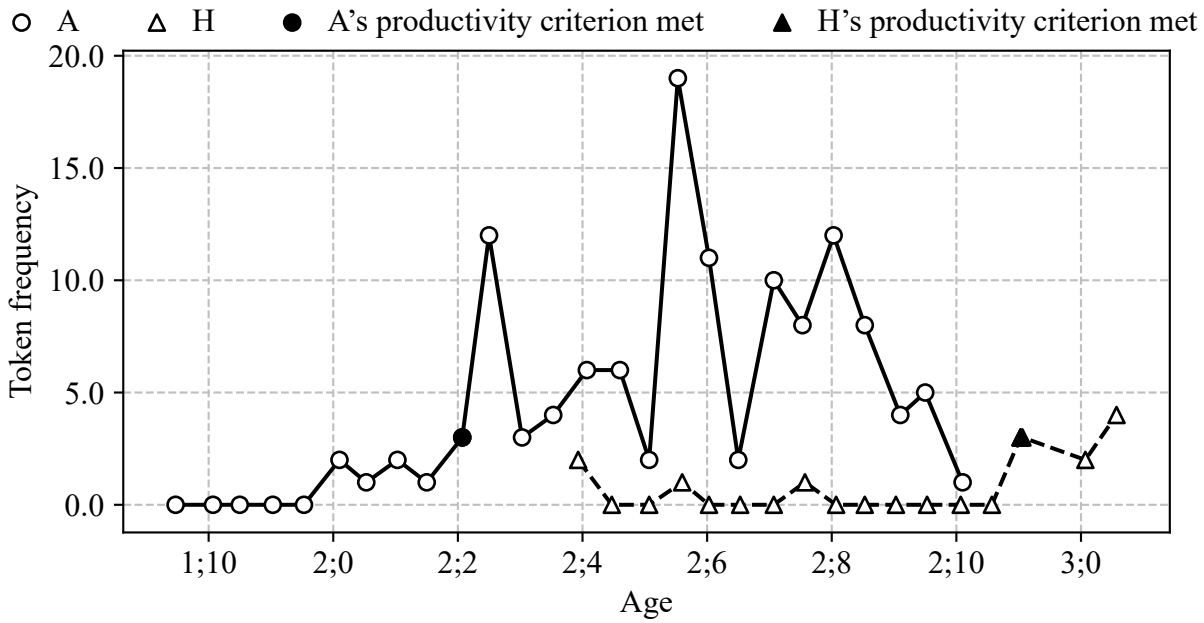


Figure 5.3 TAM inflections - Perfective -*iʈʈɨ*.

In H's transcripts, the perfective marker -*iʈʈɨ* occurs sporadically compared to the imperfective -*uka* which is attested in all the transcripts except two (at 2;6.16 and 2;11.1). The former affix is seen on seven types and 16 tokens while the latter is used with 14 types and 37 tokens. The perfective marker is attested with only three different types in any single transcript, and the imperfective with only four. Both markers meet the productivity criterion, the former at 2;11.1 and the latter at 2;9.1.

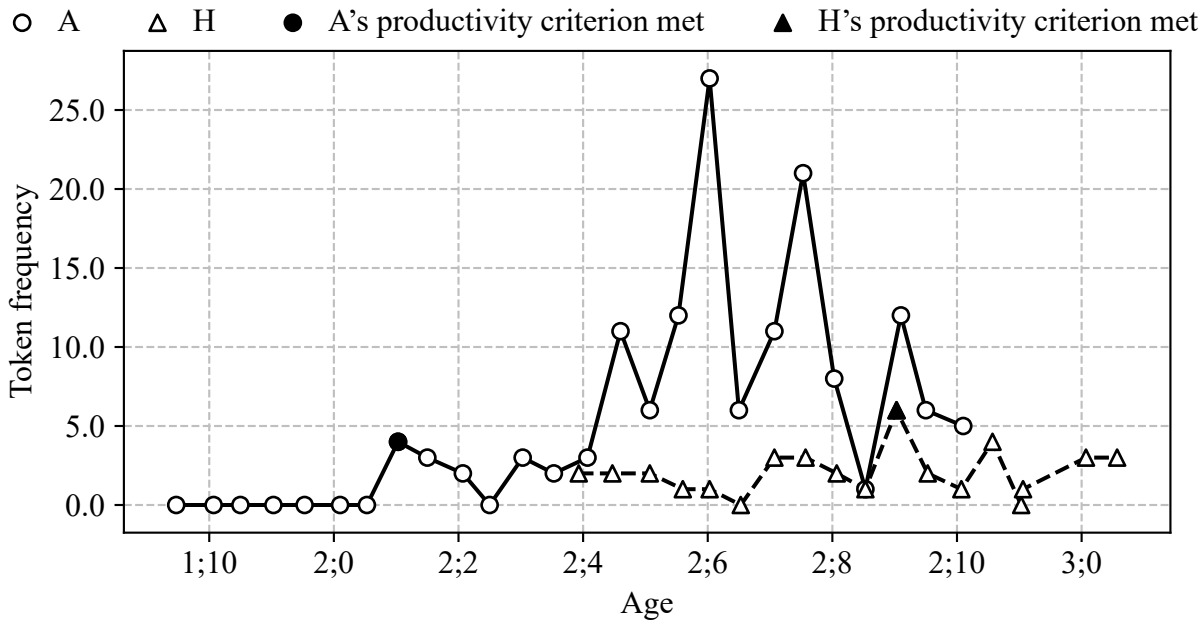


Figure 5.4 TAM inflections - Imperfective *-uka* + *aaɳi*.

(83)	caaya	ku-cc-iʈʈi	poo	(H, 2;11.18)
	tea	drink-PTCP-PFV	go.IMP	
	caaya	kuʈi-cc-iʈʈi	poo	[adult form]
	tea	drink-PTCP-PFV	go.IMP	
	'Have tea and then go.'			

(84)	naan	ʋaɾ-uʊ-aa		(H, 3;0.2)
	I.NOM	come-PTCP-AUX		
	ɲaan	ʋaɾ-uʊ-aa		[adult form]
	I.NOM	come-PTCP-AUX		
	'I am coming.'			

5.3.3 Mood

Child-directed speech typically includes extensive use of the imperatives. Given its prolific use in the ambient data, we expect that children will both comprehend and utilise these affixes readily. Recall that Malayalam has three different affixes as well as a bare form to mark the imperative. Both the children employ the different imperative forms frequently in their speech. The least emphatic form *-oo* is used in 30 tokens and over 11 types in A's data and 13 tokens involving nine types in H's transcripts. Figure 5.5 shows the production details of this marker.

The maximum number of stem types attested in any single transcript does not exceed three for A and four for H. The suffix meets the productivity criterion in A's transcripts at 2;8.1 with three stems and in H's at 2;3.28 (which is the first transcript) with four stems.

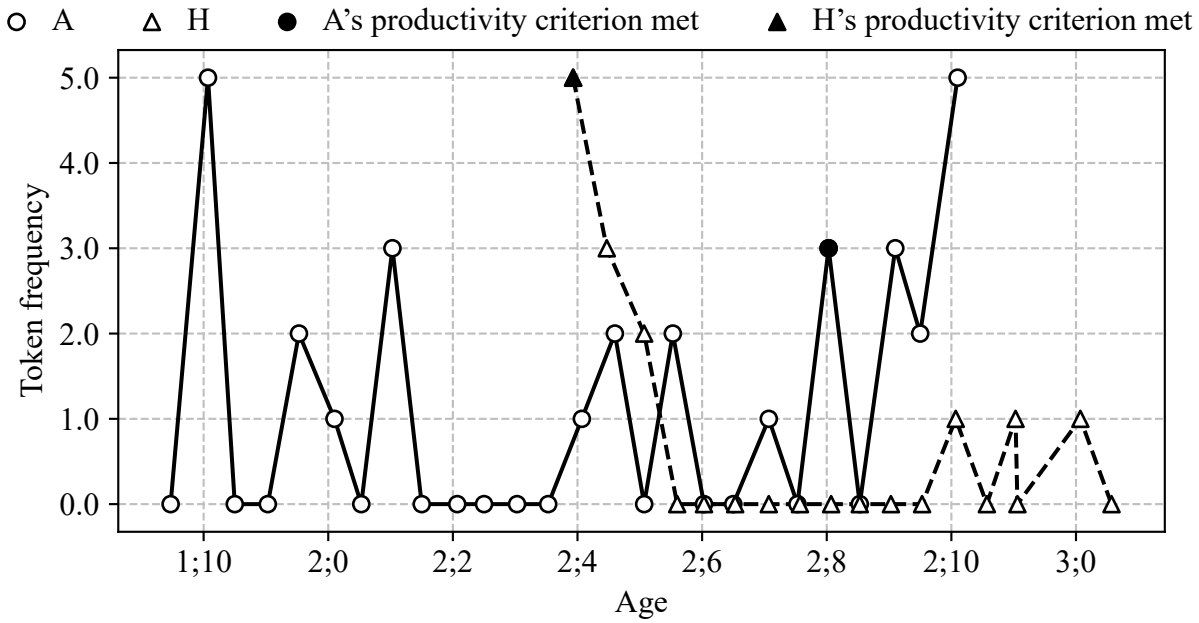


Figure 5.5 TAM inflections - Imperative *-oo*.

(85) ii baatti piṭicc-oo (A, 2;10.3)
this bat hold-IMP
'Hold this bat.'

(86) kucc-oo amm-ee (H, 2;3.28)
drink-IMP mom-VOC
kuṭicc-oo amm-ee [adult form]
drink-IMP mom-VOC
'Mom, drink.'

The more emphatic imperative marker (in the degree of command) *-ee* is seen on 13 types and 52 tokens, and eight types and 28 tokens, in A's and H's transcripts, respectively. The distribution of the marker is shown in Figure 5.6. The maximum number of stem types to which the marker *-ee* is attested in any of A's transcripts is four. A meets the productivity criterion by employing the marker with four different stems at 2;0.3 while the first instance of the productive use of the marker in H's transcripts occurs at the age 2;8.2 with three different stem types.

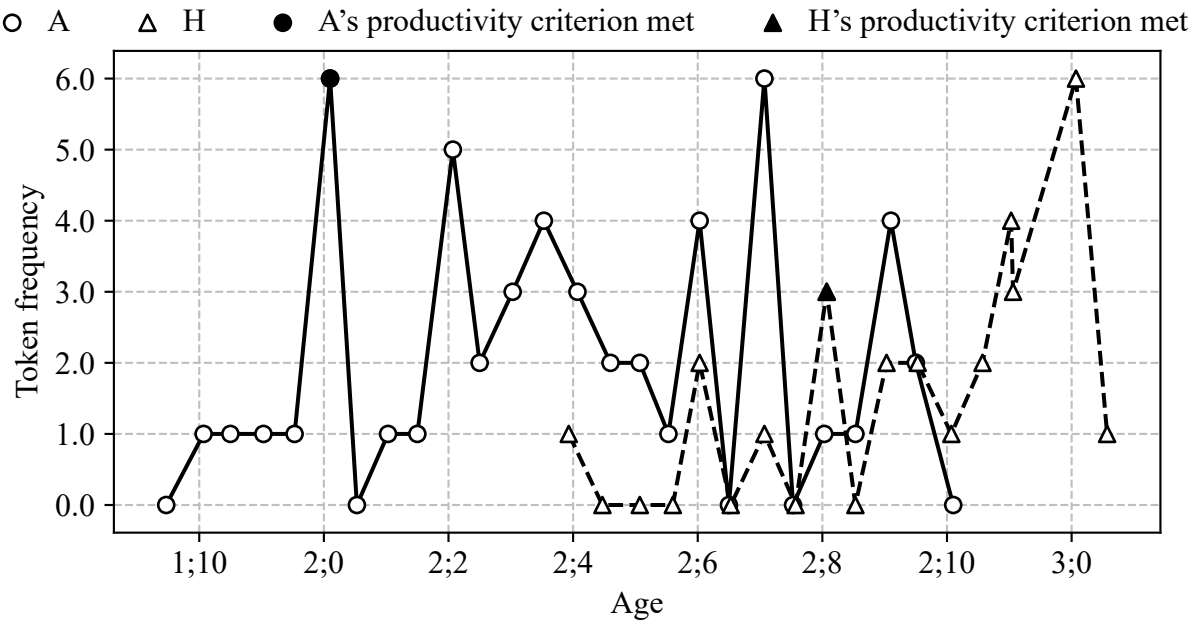


Figure 5.6 TAM inflections - Imperative *-ee*.

(87) muṯṯaayi kaaɳi-cc-ee (A, 2;6.1)
candy see-TR-IMP
'Show me the candy.'

(88) a-cc-i̇ iṯṯ-ee (H, 2;8.2)
close-TR-PTCP put-IMP
aṭa-cc-i̇ iṯṯ-ee [adult form]
close-TR-PTCP put-IMP
'Keep the door closed.'

The most emphatic marker *-aan* is seen frequently in A's transcripts while it only occurs twice in H's speech with just one instance each in two of the transcripts (at 2;8.2 and 2;11.1). Figure 5.7 shows the distribution of the marker in A's transcripts. A uses it with 14 different verb stems in 38 instances. The marker meets the productivity criterion in A's transcripts at 1;11.16 with three different stems while its highest type frequency does not exceed four in any of the transcripts.

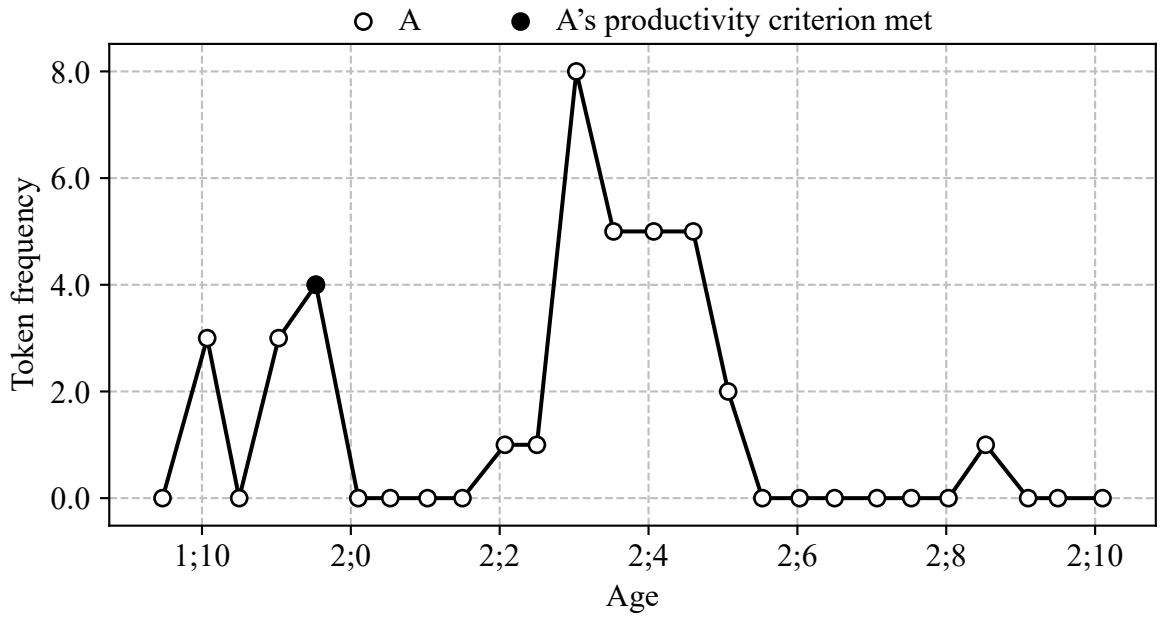


Figure 5.7 TAM inflections - Imperative *-aan*.

(89)	baag-in-aaṯṯi	ʋekk-aani	bukki	(A, 2;3.1)
	bag-GEN-inside	put-IMP	book	
	bukki	baag-in-akaṯṯi	ʋekk-aan	[adult form]
	book	bag-GEN-inside	keep-IMP	
	'Put the book inside the bag!'			

(90)	i-ṯṯi	ṯaɾ-aan	(H, 2;8.2)
	put-PTCP	give-IMP	
	'Throw it (to me)!'		

The children are frequently seen to express the imperative mood using the bare forms of the verb. In fact, bare-imperatives exceed the other three inflectional markers both in type and token frequencies in both the children's transcripts. A and H have 99 and 81 instances of such bare-imperatives, respectively, while the former employs them with 16 different verb types and the latter with 13 verb types. Of the total bare-imperative tokens, A has 39 vowel-final forms and 60 consonant-final forms followed by the release vowel [ɨ], whereas H has 64 vowel-final tokens and 17 consonant-final ones. Figure 5.8 shows the distribution of the bare-imperative tokens in the children's transcripts. Both A and H meet the productivity criterion for the bare imperatives at 1;9.14 with four unique types and at 2;3.28 with three unique types, respectively.

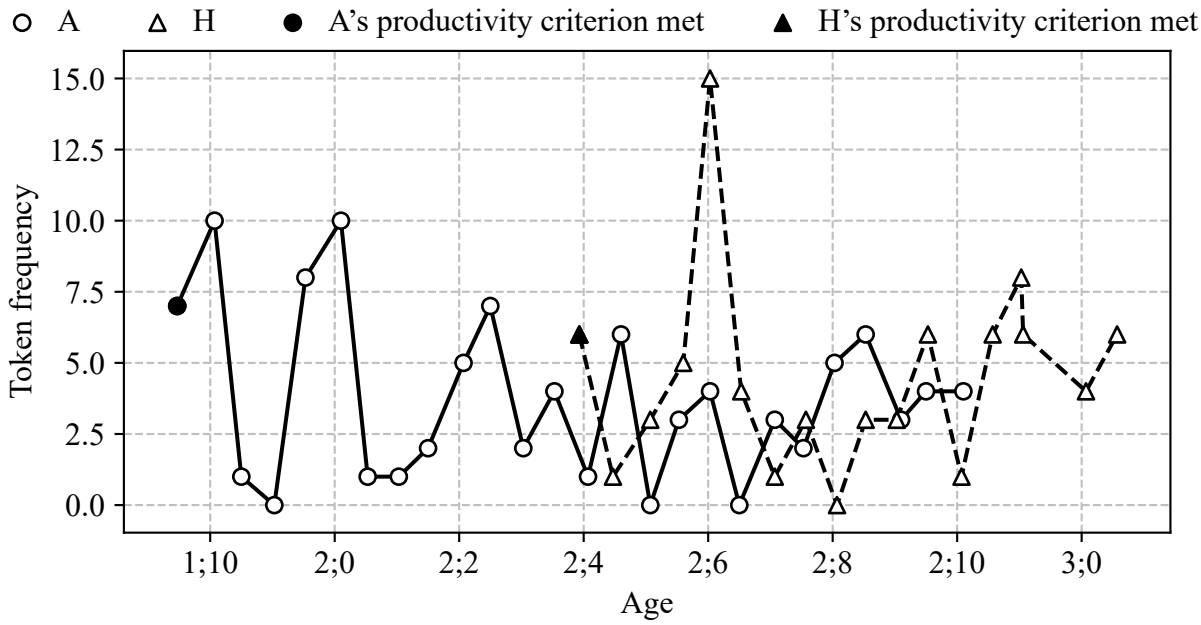


Figure 5.8 Production of bare imperatives.

(91) amm-ee poʈʈ-icc-ɨ ʈaa (A, 2;6.1)
mom-VOC break-TR-PTCP give.IMP
'Mom, open this for me.'

(92) aŋŋ-ooʈʈɨ maa-tt-i ʋekkɨ (A, 1;11.2)
there-towards remove-TR-PTCP put.IMP
'Keep it there.'

(93) amm-aa iʋiʈe iɾi (H, 2;5.18)
mom-VOC here sit.IMP
'Mom, sit here.'

(94) iʈɨ e-ʈʈɨ maattɨ (H, 2;10.17)
this take-PTCP move.TR.IMP
iʈɨ eʈu-ʈʈɨ maattɨ [adult form]
this take-PTCP move.TR.IMP
'Remove this.'

As can be seen of the four kinds of imperatives, the most frequent in the children's speech is the bare form followed by *-ee* > *-aan* > *-oo*, and *-ee* > *-oo* > *-aan* in A's and H's transcripts, respectively.

Both children use the optative morpheme to express their suggestions. A has 13 types and

33 tokens of the optative morpheme *-aṯṯe* in her transcripts. The optative is in regular, error-free use. H expresses the optative with 15 types and 29 tokens. He productively uses the optative marker with eight different stems at 3;0.2 whereas A's production of the types does not exceed more than four stems in any of the transcripts. A and H meet the productivity criterion at the ages 2;3.1 and 2;9.1, respectively, by employing the marker with three unique stems.

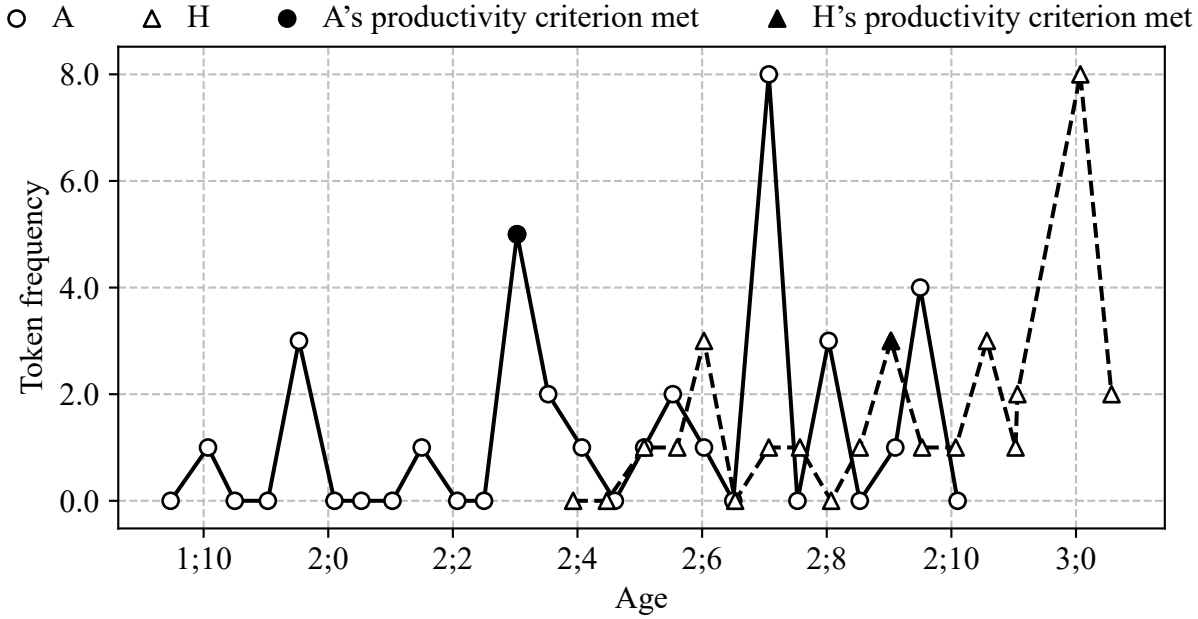


Figure 5.9 TAM inflections - Optative *-aṯṯe*.

(95)		ṉookk-aṯṯe		(A, 2;7.2)
		look-OPT		
	[ɲaan]	ṉookk-aṯṯe		[adult form]
	[I.NOM]	look-OPT		
	'Lemme see.'			
(96)		keer-aṯ-ee	maṟaṯṯ-ee?	(H, 2;8.2)
		climb-OPT-Q	tree-LOC	
	[ɲaan]	maṟaṯṯ-il	keer-aṯ-ee?	[adult form]
	[I.NOM]	tree-LOC	climb-OPT-Q	
	'Shall I climb the tree?'			

The conditional morpheme *-aal* is the least frequently employed morpheme both in terms of types and tokens in both the children's transcripts. In A's and H's speech, the conditional morpheme comes much later and only towards the end of the period of recording. The children

employ the conditional with just two different verb types with a total of three tokens in A's data and two tokens in H's data. The marker does not meet the productivity criterion in both the children's transcripts. This is not entirely surprising since the conditional marker is also one of the least frequently employed TAM inflections in adult speech, for, unlike the other finite inflections, the contexts for conditionals are less frequent in day to day speech. The conditional marker occurs at 2;3.1, 2;6.1, and 2;7.2 in A's transcripts while it occurs at 2;9.1 and 3;0.2 in H's transcripts.

(97) pooy-aa iŋŋane ʋekk-um (A, 2;10.2)
go-COND like this put-FUT
[aṯi] pooy-aal [ɲaan] [aṯi] iŋŋane ʋekk-um [adult form]
[that] go-COND [I] [that] like this put-FUT
'If (it) goes, (I) will keep (it) like this.'

(98) icc-uʋ-aa naan ʋanʈi oo-cc-aa? (H, 2;9.1)
hit-FUT-Q I.NOM vehicle run-TR-COND
ɲaan ʋanʈi ooʈi-cc-aal iʈiykk-um-oo? [adult form]
I.NOM vehicle run-TR-COND hit-FUT-Q
'Will it crash if I drive?'

5.3.4 Errors in TAM marking

In this section, we are not discussing the errors involving past tense formation, which will be discussed in greater detail in the next chapter. Across all the transcripts we find only one over-generalisation error associated with mood involving the TAM inflections other than the past tense markers. A forms the bare-imperative stem *var-* 'come' with the addition of the release vowel [ɨ], instead of forming the expected bare form *vaa*.

(99)	accay-oo[ɨ]	paɾay-aɳam		mee-cc-ooŋ[ɨ]	(A, 2;5.16)
	dad-SOC	say-MOD		buy-PTCP-with	
	*ʋaɾɨ-eɲɲɨ				
	come.IMP-COMP				
	acc-an-oo[ɨ]	paray-aɳam	[baagɨ]	mee[i-cc-ooŋ[ɨ]	[expected form]
	dad-M-SOC	say-MOD	[bag]	buy-PTCP-with	
	ʋaa-eɲɲɨ				
	come.IMP-COMP				
	'Should tell dad to buy (the bag).'				

5.4 Verbal Inflections: Summary

All the major TAM inflections except the conditional (and the imperative *-aan* in H's transcripts) are frequently used by both the children and are contextually appropriate in their usage. There are no omission errors attested for any of these markers. The only incorrect instance of TAM marking among the inflections discussed in this chapter involves the overgeneralisation of the bare stem imperative. It is clear that the kids are able to analyse the verb morphology successfully, induce the rules, and generate the word forms as required. Since the majority of the TAM inflections are uniformly and exceptionlessly marked and are without allomorphs, the performance by the children is not surprising and is congruent with the findings in the other typologically similar languages. The selection and acquisition of Malayalam past tense markers will be discussed in detail in the next chapter.

Chapter 6

Past Tense Morphology and Acquisition

6.1 Introduction

Past tense marking in Malayalam is a complex phenomenon and has been a topic of much interest to linguists over several decades (Asher, 1969; Prabodhachandran Nayar, 1972; Kunjan Pillai, 1965). Phonological conditioning of the past tense markers during the course of Malayalam's linguistic development, historically, has resulted in the formation of a variety of surface structures.

We argue that there are two past tense markers: *-i* and *-ṯu*, with the latter varying predictably with *-ṉṯu*. The allomorphs *-ṯu* and *-ṉṯu* themselves have several phonetic alternants, with the former surfacing phonetically as one of *-ṯu*, *-ṯṯu*, *-ttu*, *-ccu*, *-ṭu*, and *-ṭṭu* and the latter as one of *-ṉṯu*, *-ṉṭu*, *-ɲɲu*, *-ṉṉu*, *ṉṉ*, and *-ṇu* as shown in Table 6.1. These surface structures are the results of regular phonological accommodations[1]. We argue that the difference in the choice of the allomorphs is neither sonorant/consonant attributes of the stem or transitive/intransitive characterisations as argued in the previous accounts, but between underived and derived verbs. The class of derived verbs includes verbs that are derived from nouns (denominal), adjectives (deadjectival) and other verbs. Verbs can undergo morphological valency change derivations with either affixation and/or sound changes to the stem. For example, *karuppi*$_{\text{adj}}$ 'blackness'→ *karukk-*$_{\text{verb}}$ 'blacken' and *parakk-*$_{\text{intr}}$ 'fly' → paraṯṯ-$_{\text{tr}}$ 'fly'. Our aim in this chapter is to provide a unifying account of Malayalam past tense formation by analysing the selectional restrictions on a particular marker as enforced by a particular class of verbs. Rather than sorting verb stems into different (paradigmatic) classes based on the past tense marker they select, we focus on

[1] We do not discuss the derivation of these surface structures in this thesis.

the morpho-phonological constraints which determine the selection of a specific past tense allomorph.

Table 6.1 Malayalam past tense.

Surface Structures of *-ṯu*		**Surface Structures of *-ṉṯu***	
Past Form	**Gloss**	**Past Form**	**Gloss**
koy-**ṯu**	'cut'	ʋe-**ṉṯu**	'boiled'
ʋiir-**ṯṯu**	'inflated'	ṉaʈa-**ṉṉu**	'walked'
ʋi-**ttu**	'sold'	kaɻi-**ɲɲu**	'finished'
ciri-**ccu**	'laughed'	ko-**ɳʈu**	'hit'
iʈ-**ʈu**	'put'	ʋii-**ɳu**	'fell'
kee-**ʈʈu**	'heard'	ʋaar-**ṉu**	'trickled'

Linguists have differed in their estimate of the actual number of past tense allomorphs found in Malayalam, and have invariably attempted to classify Malayalam verbs into morphological classes (paradigms) on the basis of the number of past tense allomorphs that they identify in their individual analyses. Kunjan Pillai (1965) proposes sixteen verb classes in Malayalam to account for the differing surface forms based on the surface structure of the verb stems. Wickremasinghe and Menon (1927) proposed eight classes, Sekhar and Glazov (1961, as cited in Asher & Kumari, 1997) twelve, Asher (1969) four, Prabodhachandran Nayar (1972) four, and Valentine two (1976, as cited in Asher & Kumari, 1997).

All the above analyses have emphasised the selection of the past tense marker as determined by the phonological features of the verb stems, but they also list a number of exceptional verbs for each class that do not conform to their proposed class divisions. Asher and Kumari (1997, p. 317) state that this variation in the number of classes (from two to 16) is the outcome of varying assessments of how many of the different markers can be predicted by phonological conditioning. Valentine's (1976, as cited in Asher & Kumari, 1997) two-class division (similar to our claim) is based on whether a verb takes the marker *-i* or *-ṯu*. Class I contains verbs marked with *-i* and Class II, the ones marked with *-ṯu*. All the non-derived stems (verbs that are not formed by the addition of valency-changing suffixes according to his classification) are assigned to either Class I or II before the application of the phonological rules that yield the suffix's actual phonetic realisation. He also assigns derived stems (formed by the addition of derivational suffixes such as the causative markers) to one of these two classes. He proposes

that when a geminate in a stem is replaced with another geminate in order to increase the verb's valency, as in *iṛikk-* 'sit' ~ *iṛuṯṯ-* 'seat', the marker will be *-i* and thus it will belong to Class I. On the other hand, if a geminate like *-kk* is added directly to the verb stem in order to effect valency change, the past tense marker will be *-ṯu* and hence such verbs will belong to Class II. These rules seem ad hoc and do not really explain the grammatical motivation behind the choice of the allomorph.

Like Valentine, Rajaraja Varma (1896) also groups Malayalam verbs into two classes. He does this based on whether or not they have the *-kk* morpheme incorporated into their stems[2]. He posits two past tense markers: (i) *-ṯu* for the transitive *-kk* stems that terminate in vowels or sonorant consonants, with a variant *-ṉu* for sonorant final stems that do not have the *-kk*, and (ii) *-i* for all the remaining verb stems. However, there are drawbacks to this analysis. There are intransitive verbs that end in *-kk* that take the marker *-ṯu* such as *ciṛiykk-* 'smile' and *-i* occurs with stems that may or may not end in a sonorant (e.g. *kiiruka* 'tear' – *kiiri* 'tore'; *kaɻukuka* 'wash' – *kaɻuki* 'washed') which are not predicted by Rajaraja Varma. Further, the variant *-ṉu* is assigned to sonorant final stems but not *-ṉṯu*. Where the nasal marker surfaces as *-ṉṯu*, *-ɳṭu* etc, he analyses the forms as having euphonic nasals that occur with *-ṯu*. These rules are also posited as exceptions to his initial assumptions. However, neither sonorancy of the stem-final phoneme nor the transitivity-intransitivity feature can provide a fully satisfactory account of the past tense marking phenomenon, precisely because these markers straddle morphological and phonological classes.

These earlier accounts of Malayalam past tense formation are focused on forming classes of assorted verbs as they occur with a particular marker. In other words, the presence of an affix determined lexical class membership. The exception to this approach is Rajaraja Varma (1896). Rajaraja Varma's account, however, fails to provide a definitive account of the underlying rules as we saw above. The other accounts including Valentine's (1976, as cited in Asher & Kumari, 1997) do not make a distinction between the allomorphs *-ṯu* and *-ṉṯu* either. For Rajaraja Varma, they are exceptional in some sense. All the proposals except Kunjan Pillai (1965) treat the nasal in *-ṉṯu* as a euphonic nasal that surfaces with *-ṯu* and attempt to discover the rules for the insertion of the euphonic nasal. Kunjan Pillai treats the *-n* as a stem final nasal consonant. Valentine (1976, as cited in Asher & Kumari, 1997) reports that the nasal is evident when the stems end in either [*i*, *e*, *r*, *l*, *ɭ*, *ɻ*, *akk*, *y*, *v*] (the last two only if the stem has at least two consonantal segments).

[2]This is a transitivising suffix in Dravidian languages (Krishnamurti, 2003).

A major flaw of all these accounts is that there is no phonetic motivation for the assumption of a euphonic nasal. This suggests that the morpheme -*ṉṯu* is an allomorph whose assignment is subject to morpho-phonological rules as with the other variants.

Attempts to predict the phonetic realisation of the past tense marker depending on the nature of the stem, and creating lexical classes based on the past tense marker they choose puts the cart before the horse. It is necessary to explain why a particular marker is assigned to a particular class of stems in the first place, i.e, prior to the actual realisation of the surface forms. Therefore, in these accounts, the assignment of the marker remains arbitrary and accounts are not generative. This is problematic from a language acquisition point of view. A child will not know what the past tense marker may be unless he/she hears it in use. This runs counter to the predictive power of grammatical rules and their use in developing grammars.

In this chapter, we will demonstrate that the past tense marking works somewhat differently in the two proposed classes of verbs: a) in the underived class, the selection of the past tense marker is driven by syllable weight constraint and b) in the derived class where verbs are derived either through affixation or through sound change, the morphological rules of derivation take precedence over phonological constraints in the context of affixation, and vice versa in the context of sound change. Therefore the verbs derived through sound change behave like the underived class and syllable weight constraint is of primary importance in such verbs.

Krishnamurti (2003) says that -*ntu* was the Proto-Dravidian intransitive past tense marker, reconstructed from cognate forms in the different daughter languages. The fact that Proto-Dravidian had separate past tense markers for different morphological classes suggests that the selection of a particular past tense allomorph was determined primarily by the morphology. We argue that Malayalam is on its way to collapsing these historical class differences in past tense marking. We employ the acquisition data to a) substantiate these arguments, and b) demonstrate the various effects these proposals have on language development itself, i.e., on the acquisition patterns of the past tense markers.

6.2 Past Tense of Underived Verbs

Past tense marking in underived stems is determined by the syllabic weight constraint in Malayalam which we discussed in Chapter 4. Malayalam has a general tendency to break up heavy syllables and resyllabify the coda consonants to the adjacent onset positions. This is especially

reflected in the choice of affixes. It has to be noted that most of the inflectional affixes in Malayalam are vowel-initial ones. Krishnamurti (2003, p. 296) notes that a similar constraint holds in Proto-Dravidian on the addition of *-i*, which is a past tense marker for heavy stems in the proto-language.

Underived verb stems in Malayalam comprise both monosyllabic (e.g. *poo-* 'go') and bisyllabic (e.g. *ṯu.ḷumb-* 'brim') ones. Stems which comprise more than two syllables are either compounds or reduplicated structures in general (e.g. *ṯuḷuṯuḷumb-* 'brim and spill'). The syllable structures of verb stems belonging to the underived class are listed in Table 6.2. It has to be noted that apart from a small class of stems of CVC structure, which constitutes high-frequency verbs in general, the rest of the underived verbs choose the past tense marker *-i* as shown in Table 6.2[3]. These exceptional CVC stems choose the past tense marker *-ṯu* as in *ṉaṭ-ṭu* 'planted' from the stem *ṉaṭ-* 'plant', except in *ṯaṙ-* 'give' and *vaṙ-* 'come', which choose *-ṉṯu* instead, forming the past forms *ṯaṉṉu* 'gave' and *vaṉṉu* 'came', respectively, possibly to match the sonorant feature of the stem-final consonant. In Malayalam, the marker *-i* occurs with such heavy stems, except for the CVC class, in conformity with the syllabic weight constraint. For example, the heavy CVVC stem *maar-* 'move$_{intr}$' is resyllabified to form *maa.ri* 'moved' in the past tense form as in (100).

(100)

C	V	V	C	+	V	-	C	V	V	.	C	V
m	**a**	**a**	**r**		**i**		**m**	**a**	**a**	.	**r**	**i**

6.3 Past Tense of Derived Verbs

Derived verbs are of two types: i) denominal and deadjectival verbs, examples of which are listed in Table 6.3 and ii) verbs that are derived from other verbs (through valency changing operations). There are two operations to create derived verbs, a) by affixation (e.g.: *ooṭ-* 'run$_{intr}$' → *ooṭikk-* 'chase$_{tr}$') and b) by sound change on the stem (e.g. *aaṭ-* 'swing$_{intr}$' → *aaṭṭ-* 'swing$_{tr}$'). Affixation can be used to derive verbs of either type but sound change is used only in valency changing processes. This means of derivation affects the selection of the past tense morpheme. Verbs which are derived through affixation (with attendant lexical change) always select *-ṯu* or *-ṉṯu*, while verbs derived through sound change such as stem-final consonant gemination always choose the past tense marker *-i* as we will show in the following.

[3]However, the high-frequency verb *caa-* 'die' which selects the marker *-ṯu* ,forming the past tense *caṯṯu* 'died' is an exception to this. It is unclear whether the underlying form of the verb is *caa-* or *ca-*.

Table 6.2 List of underived verb stems in Malayalam.

Stem Type	Example	Past Tense Marker	Past Form
CVC (exceptional)	ɲaʈ- 'plant'	-ʈu	ɲaʈʈu 'planted'
	ʈaɾ- 'give'	-ɲɲu	ʈaɲɲu 'gave'
CVV	poo- 'go'	-i	pooyi 'went'
(C)VVC	ʈeeʈ- 'search'	-i	ʈeeʈi 'searched'
CVCC	minn- 'strobe'	-i	minni 'strobed'
(C)V.CVC	ciʈaɾ- 'scatter'	-i	ciʈari 'scattered'
(C)V.CVCC	ʈuɭumb- 'brim'	-i	ʈuɭumbi 'brimmed'
CV.CVVC	ʈalooʈ- 'caress'	-i	ʈalooʈi 'caressed'

Table 6.3 Examples of derived verbs in Malayalam.

Root Word	Derived Verb
ciɾi$_{noun}$ 'laugh'	ciɾiykk- 'laugh'
puka$_{noun}$ 'smoke'	pukay- 'smoke$_{intr}$'
puka$_{noun}$ 'smoke'	pukaykk- 'smoke$_{tr}$'
ʈuɻa$_{noun}$ 'paddle'	ʈuɻay- 'paddle'
karuppiɨ$_{adjective}$ 'black'	karukk- 'blacken'
ʋeɭuppiɨ$_{adjective}$ 'white'	ʋeɭukk- 'whiten'

(Rajaraja Varma, 1896)

Rajaraja Varma (1896) and Asher and Kumari (1997) state that verbs are derived from nouns by either affixing *-kk*[4] to the nominal stem or by zero-derivation (as English *knife* ~ to *knife*). Formally we notate zero-derivation as an addition of a null-affix because it allows us not only to describe the complementary distribution of the allomorphs *-ʈu* and *-ɲɲu* but also to capture the uniformity of the derivational processes involving both affixation and sound change. For example, the intransitive verb *ʈiɾiy-* 'turn$_{intr}$' is assumed to be derived from the noun *ʈiɾi* 'turn' by zero-derivation, while its transitive counterpart *ʈiɾiykk-* 'turn$_{tr}$' is formed through the

[4]The *-kk* suffix is also used to derive Malayalam verbs from Sanskrit roots. Examples of such verbs are *ciɲʈikk-* 'think', *ɲamikk-* 'bow' etc. derived from the Sanskrit roots *ɲam-* and *ciɲʈ-*, respectively (Rajaraja Varma, 1896). However, such derived verbs are used in formal contexts in general, and are typically not found in child-directed speech. Here, we focus only on the role of *-kk* in deriving verbs from native (Dravidian) roots.

affixation of *-kk* to the noun. Burrow and Emeneau's *A Dravidian Etymological Dictionary* (1984) confirms and offers corroborative evidence that a majority of these verbs are derived from nouns or adjectives.

Unlike the underived verbs, whose past tense formation is driven by phonology, it is the morphological rules that determine the past tense morphemes of derived verbs. The past tense formation rules may be formulated as follows:

(101) [[nominal stem + -kk]$_{verb}$ -ṯu]$_{past}$

(102) [[nominal stem + -ø]$_{verb}$ -ṉṯu]$_{past}$

Tables 6.4 and 6.5 provide examples of past tense formation in denominal verbs derived through (101) and (102), respectively.

Table 6.4 Past tense formation using *-ṯu* in *-kk* affixed stems.

	Derived Verb	**Past Tense Marker**	**Past Form**
[1562][5]	ciɾiy**kk**- 'laugh'	-ṯu	ciɾi-ccu 'laughed
[5496]	ʋeɭu**kk**- 'whiten'	-ṯu	ʋeɭu-ṯṯu 'whitened'
[5421]	ʋil**kk**- 'sell'	-ṯu	ʋi-ttu 'sold'
[2017]	kee⌊**kk**- 'hear'	-ṯu	kee-ṯṯu 'heard'

Table 6.5 Past tense formation using *-ṉṯu* in zero-derived stems.

	Derived Verb	**Past Tense Marker**	**Past Form**
[859]	eriy- 'throw'	-ṉṯu	eri-ɲɲu 'threw'
[3770]	ṉeeɾ- 'vow'	-ṉṯu	ṉeer-ṉu 'vowed'
[3793]	ṉoo- 'pain'	-ṉṯu	ṉo-ṉṯu 'pained'
[5432]	ʋiɭ- 'crack'	-ṉṯu	ʋi-ɳṯu 'cracked'
[5372]	ʋaaɻ- 'rule'	-ṉṯu	ʋaa-ɳu 'rule'

Diachronically, the rules in (101) and (102) derived the transitive and intransitive (typically, unaccusative) verbs, respectively. Evidence for this assumption is available in the history of Dravidian languages. Krishnamurti (2003) postulates *-kk-* as the transitivity affix and *-ntu* as

[5]The square bracketed numbers correspond to the entries in *A Dravidian Etymological Dictionary* (Burrow & Emeneau, 1984) which show the existence of a related noun.

the past tense marker for Proto-Dravidian intransitive verbs. In Malayalam, such a distinction is not present synchronically and the context in which these rules apply are apparently unpredictable. However, an examination of nouns that lend themselves to both the derivations shows us that this is exactly how the nouns behave. A noun can derive either the transitive verb with *-kk* or an intransitive one through zero-derivation and the past tense morpheme selected aligns with that derivational change, reflecting the diachronic difference, covertly, in the synchronic morphology. Table 6.6 presents examples of such denominal transitive and intransitive pairs derived from the same base noun. This is one of the strongest arguments for analysing derived stems in a manner distinct from underived ones.

Table 6.6 Derivation of transitive-intransitive pairs through affixation.

	Noun	**Derivation**	**Past Tense Form**
[1709]	kuɳir 'small mounds[6]'	kuɳir- + -kk	kuɳir-ʈʈu [soak-PST]$_{tr}$
		kuɳir- + -ø	kuɳir-ɳu [soak-PST]$_{intr}$
[3246]	ʈiɾi 'turn'	ʈiɾiy- + -kk	ʈiɾi-ccu [turn-PST]$_{tr}$
		ʈiɾiy- + -ø	ʈiɾi-ɲɲu [turn-PST]$_{intr}$
[811]	eɾi 'fire'	eɾiy- + -kk	eɾi-ccu [burn-PST]$_{tr}$
		eɾiy- + -ø	eɾi-ɲɲu [burn-PST]$_{intr}$
[652]	uɾi 'hide'	uɾiy- + -kk	uɾi-ccu [skin-PST]$_{tr}$
		uɾiy- + -ø	uɾi-ɲɲu [shed-PST]$_{intr}$
[5411]	ʋiɾi 'spread'	ʋiɾiy- + -kk	ʋiɾi-ccu [spread-PST]$_{tr}$
		ʋiɾiy- + -ø	ʋiɾi-ɲɲu[spread-PST]$_{intr}$
[3682]	ɳira 'fill'	ɳiray- + -kk	ɳira-ccu [fill-PST]$_{tr}$
		ɳiray-+ -ø	ɳira-ɲɲu [fill-PST]$_{intr}$
[4760]	mara 'cover'	maray- + -kk	mara-ccu [hide-PST]$_{tr}$
		maray- + -ø	mara-ɲɲu [hide-PST]$_{intr}$
[1817]	kuɻa 'mixture'	kuɻay- + -kk	kuɻa-ccu [mix-PST]$_{tr}$
		kuɻay- + -ø	kuɻa-ɲɲu [mix-PST]$_{intr}$
[277]	aɻi 'ruin'	aɻiy- + -kk	aɻi-ccu [untie-PST]$_{tr}$
		aɻiy- + -ø	aɻi-ɲɲu [loosen-PST]$_{intr}$
[1818]	kuɻi 'pit'	kuɻiy- + -kk	kuɻi-ccu [dig-PST]$_{tr}$
		kuɻiy- + -ø	kuɻi-ɲɲu[shrink-PST]$_{intr}$

As we can see from Table 6.6, *kuɻa* 'mixture' is a noun. Affixing *-kk* to the nominal stem derives the transitive verb form *kuɻaykk-* whereas zero-derivation results in the intransitive verb form *kuɻay-*. The difference in valency change in the two derived verb forms is shown in (103) and (104).

(103) *kuɻaykk-* 'mix$_{tr}$' e.g. ɲaan maaʋɨ kuɻa-ccu
I.NOM dough mix-PST
'I mixed the dough.'

(104) *kuɻay-* 'mix$_{intr}$' e.g. maaʋɨ kuɻa-ɲɲu
dough mix-PST
'The dough got mixed.'

Additional support for the rules in (101) and (102) comes from another past tense formation pattern. The morphological marking of verbs that are homophonous in the transitive and the intransitive forms such as *vaaɾ-* and *uuɾ-* are of particular interest. As underived transitive verbs *vaaɾ-* and *uuɾ-* mean 'to take something by handfuls' and 'remove something', respectively. As denominal intransitive verbs (through zero-derivation) the verbs mean 'to trickle down' and 'to crawl', respectively. Table 6.7 summarises the past tense selection in such homophonous verb alternants. It is important to note that while the underived, transitive version of these verbs take the past tense marker *-i*, forming *vaaɾ-i* and *uuɾ-i* respectively, the intransitive version take the past tense marker *-n̪t̪u* forming *vaar-n̪u* and *uur-n̪u*. The fact that while marking homophonous verbs for past tense, the marker *-i* is selected in the underived verbs observing the syllabic weight constraints and the marker *-n̪t̪u* is selected by the morphological class (zero-derivation of intransitives) in the derived verbs, makes it clear that different principles are in operation in the derived and underived verb classes. This also underscores and strengthens our analysis of *-n̪t̪u* as the morphologically selected, intransitive past tense marker.

A possible alternative analysis is that the *-t̪u* marker is selected because of the obstruent suffix *-kk* and *-n̪tu* because of sonorancy, thus making the tense marker selection entirely phonology driven and undermining the case for a morphological selection. However, if that were the case, any stem that has *-kk* in it will automatically be deemed to be heavy and will violate the syllabic weight constraint as in *ciriykk-* 'smile'. Given such cases, we would expect the default marker to always be *-i*. This is incorrect and does not cohere with the facts. *ciriykk-*

[6]Small mounds of cultivable earth in rice fields which are irrigated and thus moist. The intransitive verb derived from the noun means 'get soaked' while the transitive verb means 'soak something'.

Table 6.7 Differences in past tense marking in homophonous stems.

Verb Form	Morphological Class	Marker	Past Tense Form	Meaning
ʋaar-	underived transitive	-i	ʋaaɾi	‘scooped up’
	derived intransitive	-n̪tu	ʋaarn̪u	‘trickled down’
uur-	underived transitive	-i	uuɾi	‘removed’
	derived intransitive	-n̪tu	uurn̪u	‘crawled down’

forms the past tense *ciri-ccu* and not **cirikk-i*. Moreover, there are unergative verbs that pattern in a manner that locates them between transitives and unaccusatives such that they take the past tense marker *-nṯu* despite having the obstruent *-kk* in the verb stem, as we shall see below. Thus, phonology cannot be the primary determiner of past tense affix selection in all these cases.

The morphological behaviour of unergative verbs like *iɾiykk-* ‘sit’ is somewhat like that of transitive verbs and somewhat like that of unaccusative verbs. Like the transitive verbs they are marked by the *-kk* morpheme, but pattern with unaccusative verbs in selecting the past tense marker *-nṯu*, though we might expect them to pattern with the transitive verbs since they are derived through *-kk* affixation. Examples of such verbs are given in Table 6.8. In Modern Malayalam, it is not just these unergatives that show this in-between pattern. Transitive verbs like *karakk-* ‘to milk’ and unaccusatives like *paɾakk-* ‘to spread’ form the past tenses *kara-nnu* [milk-PST] and *paɾa-nnu* [spread-PST], respectively rather than the expected **karaṯṯu* and **paraṯṯu*. We consider verbs that select both *-kk* and *-nṯu* as exceptional verbs.

Table 6.8 Past tense formation of *kk*-derived verbs that select the past tense marker *-nṯu* instead of *-ṯu*.

	Derived Verb	Past Tense Marker	Past Form
[4020]	parakk- ‘fly’	-n̪tu	para-nnu ‘flew’
[4760]	marakk- ‘forget’	-n̪tu	mara-nnu ‘forgot’
[1385]	karakk- ‘milk’	-n̪tu	kara-nnu ‘milked’
[480]	iɾiykk- ‘sit’	-n̪tu	iɾi-nnu ‘sat’
[3675]	n̪ilkk- ‘stand’	-n̪tu	n̪i-nnu ‘stood’

A third piece of evidence for *-kk* derivation of denominal verbs comes from the formation of bare imperatives in Malayalam. In the discussion of TAM markers we saw that Malayalam has

both inflected and bare-imperative forms (see 5.1). Verbs formed through *-kk* derivation have two bare imperatives, one with *kk* and another without. These are freely alternating without any difference in meaning. For example, the verb *ɪɾiykk-* 'sit', has the forms *iɾi* as well as *ɪɾiykkɨ* both meaning [sit.IMP]. Additional examples of such imperatives are shown in Table 6.9.

Table 6.9 Bare stem imperatives of denominal verbs.

Verb Stem	**Bare Imperatives**		**Gloss**
	with *-kk*	**without *-kk***	
ɲilkk- 'stand'	ɲilkkɨ	ɲillɨ	stand!
ʊeykk- 'put'	ʊeykkɨ	ʊey	put!
oɻiykk- 'pour'	oɻiykkɨ	oɻi	pour!
ɲanaykk- 'water'	ɲanaykkɨ	ɲana	water!
aʈiykk- 'beat'	aʈiykkɨ	aʈi	beat!
ɲaʈakk- 'walk'	ɲaʈakkɨ	ɲaʈa	walk!
koʈukk- 'give'	koʈukkɨ	koʈɨ	give!
ʈoʈaykk- 'wipe'	ʈoʈaykkɨ	ʈoʈa	wipe!
kuʈiykk- 'drink'	kuʈiykkɨ	kuʈi	drink!
ʊiɾiykk- 'spread'	ʊiɾiykkɨ	ʊiɾi	spread!

Thus, the past tense formation patterns in homophonous transitive-intransitive verbs and derivation of denominal intransitive and transitive pairs from the same base noun together suggest that these rules ((101) and (102)) must have originated as clear transitive-intransitive indicators. It appears that subsequent reanalysis through overextension and semantic change has obfuscated their original specifications. Since these are not transitive verbs, the *-kk* addition might have been overridden as part of paradigm levelling of some kind and got reanalysed as stem internal consonants. The derivational nature of these verbs is only transparent through the past tense marking that is selected. This makes this class of verbs synchronically deviant in its past tense pattern.

Whether a derived verb is transitive or intransitive is dependent on the semantics of the source noun. When the rules in (101) or (102) are overextended to nouns that do not conform to the original transitive-intransitive distinction, the rule necessarily loses its predictability. For example, the overextension of the intransitive formation rule (102) to nouns such as *eri* [859] 'throw' and *parai* [4301] 'word' which only lend themselves to transitive verbs such as *eriy-*

'throw' and *paray-* 'say' by virtue of their semantics, contributes to the loss of the rule's predictability which in turn blurs the transitive-intransitive distinctions in Malayalam past tense marking. In addition, semantic changes cause valency changes as well. For example, *vaaɻ-* originally meant 'flourish' and was zero-derived with the past tense marker *-ṉṯu*, as we expect. However, in Modern Malayalam *vaaɻ-* has undergone semantic shift and transformed into a transitive verb meaning 'rule', but retains the intransitive past tense marker *-ṉṯu*. This results in conflicting characterisations of its morphology and semantics. These changes, as well as mismatches, do not ensure ease of acquisition.

It also has to be noted that most of the intransitive verbs were sonorant ending in Proto-Dravidian, and they chose the past tense marker *-ntu* (Krishnamurti, 2003). In Modern Malayalam, where the original premises of the derivational rules (101) and (102) are obscure, many sonorant ending verbs select the past tense marker *-ṉṯu*, the verb *eriy-* 'throw' is an example of such a verb. It is possible that the morphological rules were trumped by the phonological feature matching between the stem-final and the affix-initial consonants as an outcome of the reanalysis of past tense formation rules. Malayalam speakers must have generalised the past tense rule from the phonological features of the verb stem instead of the original morphological class differences over the course of time. Thus, synchronically, the past tense marker *-ṉṯu* occurs with sonorant-ending verbs straddling different morphological classes.

We now to turn to the other type of derived verbs, i.e., verbs derived through valency changing operations which can involve either affixation or stem internal phonological changes. Transitive verbs and causative verbs can be derived from intransitives, with an increase in valency and a modification of the event structure of the verb as shown in (105) and (106). This increases the valency by the addition of [cause] to the event structure of the verb and mapping a [cause] theta-role to a new argument in the syntax. These derived transitive and causative verbs are formed by the regular addition of *-kk* to the verb stems which do not have a morpheme internal *-kk-* sequence as part of the verb's phonological representation as in (105), and *-pp-* to the stems with a stem internal *-kk-* as in (106). Such derived transitive verbs also take the past tense marker *-ṯu* as shown in Table 6.10.

(105) *ooɻ-* [run.INTR] e.g. avan ooɻ-uṉṉu
he.NOM run-PRS
'He is running.'

*ooʈiy-**kk*** [run-TR] e.g. avan ʋanʈi ooʈiy-kk-uɳɳu
he.NOM vehicle run-TR-PRS
'He is driving the vehicle.'

(106) *kaɻiy**kk***- [eat.TR] e.g. avа-ni coorɨ kaɻiykk-aɳam
he-DAT rice eat-MOD
'He wants to eat rice.'

*kaɻi**ppykk***- [eat-CAUS] e.g. eni-kkɨ avan-e coorɨ kaɻippykk-aɳam
I-DAT he-ACC rice make eat-MOD
'I want to make him eat rice.'

Table 6.10 Valency change through affixation.

Base Verb	**Past Tense**	**Derived Verb**	**Past Tense**
[paaʈ-]$_{tr}$- 'sing'	paaʈi	[paaʈiykk-]$_{caus}$ 'make sing'	paaʈi-ccu
[paray-]$_{tr}$ 'say'	paraɲɲu	[parayikk-]$_{caus}$ 'make say'	miɳʈi-ccu
[caɾiy-]$_{intr}$ 'tilt'	cariɲɲu	[caɾiykk-]$_{tr}$ 'tilt'	caɾi-ccu
[ʋekk-]$_{tr}$ 'put'	ʋeccu	[ʋeppiykk-]$_{caus}$ 'make put[7]'	ʋeppi-ccu
[ʋiirkk-]$_{intr}$ 'inflate'	ʋiirʈʈu	[ʋiirppiykk-]$_{tr}$ 'inflate'	ʋiirppi-ccu

Table 6.10 provides further examples which collectively demonstrate that verbs undergoing valency change through affixation choose the past tense marker -*ʈʈu* like the verbs derived from nouns through affixation, without exception. Thus, affix-derived verbs (whether denominal or valency changed) behave uniformly with respect to the past tense marker. We must also note that all the transitive verbs formed from intransitives through affixation select the past tense marker -*ʈʈu* and never -*ɳʈʈu* as in Table 6.10. This is further corroborative evidence to our earlier claim that -*ɳʈʈu* is indeed an intransitive past tense marker for the derived verbs. Thus, past tense affix selection in derived verbs involving affixation is primarily driven by the morphology and not phonology unlike in the underived verbs.

Valency change can also be effected through phonological modification of the verb stem by either a) augmenting the coda or b) modifying the features of the coda consonants while preserving the overall original syllable structure. A summary of these changes is provided in (107) - (111).

[7]Make someone put something.

(107) *iɭa**k***- ‘stir$_{intr}$’ —gemination→ *iɭa**kk***- ‘stir$_{tr}$’

(108) *ana**ŋŋ***- ‘move$_{intr}$’ —denasalisation→ *ana**kk***- ‘move$_{tr}$’

(109) *maa**r***[8]- ‘change$_{intr}$’ —coda augmentation→ *maa**tt***- ‘remove$_{tr}$’

(110) *para**kk***- ‘fly$_{intr}$’ —*-kk* to -*tt*→ *para**tt***- ‘fly$_{tr}$’

(111) *vii.ɻ**ø***- ‘to fall$_{intr}$’ —*ø* to -*tt*→ *vii.ɻ**tt***- ‘to fell’.

These phonologically derived verbs pattern with the underived verb class in choosing the past tense marker *-i*, whose selection as we established earlier is determined by the syllabic weight requirements. It is easy to see why. With the incrementation or modification of the codas, these stems become heavy syllabled. The syllable quantity triggers the selection of the past tense marker *-i*. Therefore, the past tense marker selection in verbs derived through sound change is essentially a phonologically driven selection unlike in the verbs derived through affixation. Thus, underived verbs and phonologically derived verbs pattern together and in both, the past tense marker selection is determined by the phonology and, specifically, by the syllable weight constraint. In contrast, affixally derived denominal and valency-increased verbs select the past tense marker given the morphological feature [±TRANSITIVE].

Finally, verbs with stem final phonological sequence -*tt* require additional consideration. Table 6.11 provides examples of -*tt* derivation in denominal verbs derived by the rules (101) and (102). Only denominal verbs derived through the affixation of *-kk* and zero-derivation undergo this valency change involving -*tt*. *parakk*- ‘fly’ is a denominal verb that is derived by affixing *-kk* to the nominal base. The remainder of the denominal verbs (typically intransitives) that are not derived through *-kk* affixation are derived by zero-derivation.

Valentine (1976, as cited in Asher & Kumari, 1997) argues that the *-kk* in the intransitive stem undergoes a sound change from *-kk* → -*tt* as in *parakk*- ~ *para**tt***-, and -*tt* is added to the zero-derived verb stems as an affix as in *vii.ɻ*- ~ *vii.ɻ**tt***- without realising that these are denominal verbs that are being further derived. An argument against such an affixation account is that both stem types select the past tense affix *-i* to form *para**tt**i*$_{tr}$ ‘flew’ and *vii.ɻ**tt**i*$_{caus}$ ‘make fall’. This is where our null-affix account brings in clarity to the derivation processes, especially since both stem types choose the phonologically selected past tense marker *-i*. Therefore, a more economical account treats both types of denominal verbs alike with a single rule as in (112).

[8]K. P. Mohanan and Mohanan (1984) posit *maar*- to have the underlying structure maar$_{C}$r- which changes to [tt] as in *maatt*-. However, we just assume that there is an extra consonant at the syllabic level in the valence-changed form *maatt*-, which is the surface form of the underlying *maatr*-.

Table 6.11 Examples of -*t̪t̪* derivation.

Derivation			**Past Tense Form of the Derived Verb**
iɾikk- 'sit'	→	iɾut̪t̪- 'seat'	iɾu-t̪t̪-**i** [sit-TR-PST]
kiʈakk- 'lie'	→	kiʈat̪t̪- 'lay'	kiʈa-t̪t̪-**i** [lay-TR-PST]
n̪aʈakk- 'walk'	→	n̪aʈat̪t̪- 'make walk'	n̪aʈa-t̪t̪-**i** [walk-TR-PST]
ʋiiɻø- 'fall'	→	ʋiiɻt̪t̪- 'fell'	ʋiiɻ-t̪t̪-**i** [fall-TR-PST]
paʈarø- 'spread'	→	paʈart̪t̪- 'spread'	paʈar-t̪t̪-**i** [spread-TR-PST]
t̪aaɻø- 'sink'	→	t̪aaɻt̪t̪- 'sink'	t̪aaɻ-t̪t̪-**i** [sink-TR-PST]

Valentine's analysis, with sound change account for one class and an affixation account for the other, loses out on a more general account of denominals:

(112) kk, ø → t̪t̪/ _____ [suffix]

6.4 Intermediate Summary

The discussion so far has demonstrated that the past tense marker selection primarily depends on whether a verb is underived or derived. In the underived verbs, the syllabic weight constraint governs the morpheme selection. The underived verb stems are heavy stems in general and they choose the past tense marker *-i* except for a certain small class of high-frequency CVC verbs which typically choose -*t̪u* instead.

As we have also seen, denominal verbs that are affixally derived select the allomorphs -*t̪u* (when derived with *-kk-*) and -*n̪t̪u* (for zero derivation). Diachronically, these past tense allomorphs were sensitive to the transitive/intransitive distinction and have since lost their original morphological specifications as a result of overextensions and reanalyses. The morphological class differences in the inflectional assignment have thus collapsed, though we see its reflex in the past tense marker that is instantiated. Valency change is effected either by affixation (of *-kk* or *-pp*) or by modifying the stem to create superheavy syllables. In the latter case, the verbs pattern with underived verbs that are heavy and the past tense marker is *-i* which allows for suitable syllable reconfiguration. The former pattern with denominal verbs and select -*t̪u*.

It can be seen from the discussion above that a distinction has to be made in the selection of the past tense marker in different verbs. The marker is either selected because of the syllable

weight considerations, or because of the morphological properties of the verbs (derivation from nouns or for valency change etc.). This *mode of selection*, as we call it, has consequences for language acquisition. We will see in Section 6.5 that Malayalam is undergoing a synchronic shift towards phonological selection as the rules determining morphological class are losing their predictability.

6.5 Interface Patterns in Affixed Verbs and Phonological Levelling

Malayalam is on its way to collapsing the morphological class differences in past tense marking. The collapse of morphological class distinctions can be seen from the past tense formation patterns exhibited by certain denominal verbs that are derived through the morphological rule of zero-derivation. In such verbs, we see that instead of past tense marker selection determined by the morphology and the feature [±TRANSITIVE], the verbs make a phonological selection of the past tense marker. For example, *vell-* 'challenge' and *coll-* 'say' are heavy stems but are also denominal, zero-derived forms (rule (102)). Interestingly, these verbs have two past tense forms *ve-ṉṉu* and *vell-i* both meaning 'challenged', and *co-ṉṉu* and *coll-i*, both meaning 'said'. This pattern is already reported by Rajaraja Varma (1896) and is treated as special behaviour. He, however, does not take cognizance of the nature of those stems, i.e., whether they are derived or underived. In these verbs, *vell-* and *coll-*, the past tense forms marked with *-ṉṯu* are predicted according to the original derivation rules. However, these *-ṉṯu* forms are being used less frequently. Instead, the past tense forms with *-i* are in current and increasing use, suggestive of a shift towards the phonologically driven past tense marking using *-i*. Unsurprisingly, when the morphological class information begins to be less transparent, a phonological levelling is to be expected. Additionally, these are low-frequency verbs. So the levelling (regularisation) is even more likely.

Nonetheless, the past tense forms of a few high-frequency zero-derived stems ending in lateral geminates like *cell-* 'go' and *koll-* 'kill', are *ce-ṉṉu* 'went' and *ko-ṉṉu* 'killed', respectively, instead of **cell-i* and **koll-i* which is what we would expect given a) their phonological shape and b) the absence of morphological information regarding the derivation. Burrow and Emeneau in *A Dravidian Etymological Dictionary* (1984) in fact attest the noun *kolli* [2132] as

'killing' although that noun is not in use in Modern Malayalam[9]. Hence, the verb *koṉṉu* could be derivationally related to either the noun *kola* (in current use) or the noun *kolli* (both meaning 'killing'). This means that the verb belongs to the denominal verb class and given the sonorancy of the stem, using the marker *-ṉṯu* is appropriate. Likewise, *Sabdatharavali* (Padmanabha Pillai, 1918) has the entry *cellɨ* to mean both 'debt' and 'journey' related to the verb *cell-* from which it could have been derived, which in turn explains its nasalised past tense marker *-ṉṯu*. Unlike the earlier discussed *vell-* and *coll-*, which are phonologically similar to *cell-* and *koll-*, the latter are high-frequency verbs and thus escape phonological levelling as is true of 'irregular' but high-frequency forms.

This kind of phonological levelling is not unusual in languages, the past tense selection of English irregular verbs being a case in point. Pinker (1999) proposes that frequency can play a crucial role in determining the existence of an irregular verb. He notes that when the verbs are of low-frequency, there may exist doublets of regular/irregular pairs such as *slayed*/*slew*, *strived*/*strove*, *dived*/*dove*, *dreamed*/*dreamt* etc. or else they may completely go out of use such as *chid*, *dempt*, and *abode* (Pinker, 1999, p. 131).

Another example of such conflict between the phonological and morphological patterns in Malayalam is the co-existence of the verbal forms *vaaŋŋ-* and *vaaŋŋikk-* both meaning 'buy'. The former takes the past tense marker *-i* to form *vaaŋŋ-i* as determined by the phonological rules and the latter forms the past tense with *-ṯu* as in *vaaŋŋi-ccu* as per the morphological rules of affixation, both meaning 'bought'. These are cases where the past tense marking in verbs are driven by conflicting requirements.

According to Pinker (1999, p. 131) "if a memory entry is faint or blurry because the word is uncommon, the matching and fetching will be especially erratic, and often the rule may not be braked in time". He says that people might start producing forms like *slayed* as speech errors, but the listeners may store it in their memory as a genuine past tense form. This can be attributed to Malayalam past tense formation as well, especially, for past tense forms that exist as doublets. In the next section, we look at the acquisition of the past tense markers and we will discover whether the phonological levelling impacts the acquisition of Malayalam past tense markers and whether the modes of selection affect the course of development.

[9] Although the noun *kolli* does not exist independently, few forms like *aaɭekkolli* 'man killer', which typically refers to animals, still exist in Malayalam.

6.6 Acquisition of Past Tense

As we described at some length in Chapter 5, Malayalam verbs are inflected for tense, aspect, and mood, but do not show overt agreement with their subjects for person, number and gender, unlike the other Dravidian languages. The morphemes *-uṉṉu* (present tense) and *-um* (future tense) apply exceptionlessly. The past tense marker selection is of particular interest in language acquisition because of the morphophonological rules at play and the conflicting paradigmatic behaviour of many verbs. In the following sections, we will examine the acquisition of Malayalam past tense markers and argue that children appear to be sensitive to the multiple constraints at work in Malayalam. The acquisition data will be seen to support our analysis of Malayalam past tense morphology.

Of the total inflectional forms produced (both nominal and verbal), verbal inflection constitutes 52% of A's and H's total inflectional production as can be seen in Table 6.12. Analysing the transcripts of the two children, we find that 37% of A's total TAM[10] inflection forms or 19% of the total inflectional production comprises past tense markers. In H's speech, past tense markers comprise 47% of the TAM inflections or 25% of the total inflectional production. The details of the two children's inflectional production in absolute numbers, including the past tense tokens, are given in Table 6.12.

Table 6.12 Subjects' inflectional production.

Tokens	**A's Transcripts**	**H's Transcripts**
Total inflections	3346	1069
Verbal inflections (TAM)	1725 (52%)	558 (52%)
Tense inflections	1304 (39%)	431 (40%)
Present tense markers	173 (5%)	90 (9%)
Future tense markers	485 (14%)	76 (7%)
Past tense markers	646 (19%)	265 (25%)
-i	260	100
-ṯu	172	84
-ṉṯu	214	81

We have already presented the grammar and acquisition patterns of each TAM inflection

[10]Tense Aspect Modality (which are the main verbal inflections in Malayalam).

in Malayalam in the previous chapter. We wish to discuss the past tense markers here, keeping in mind the extensive allomorphy and the number of constraints that underlie the allomorph selection. Figure 6.1 summarises the difference in past tense formation in the derived and underived verbs in Malayalam.

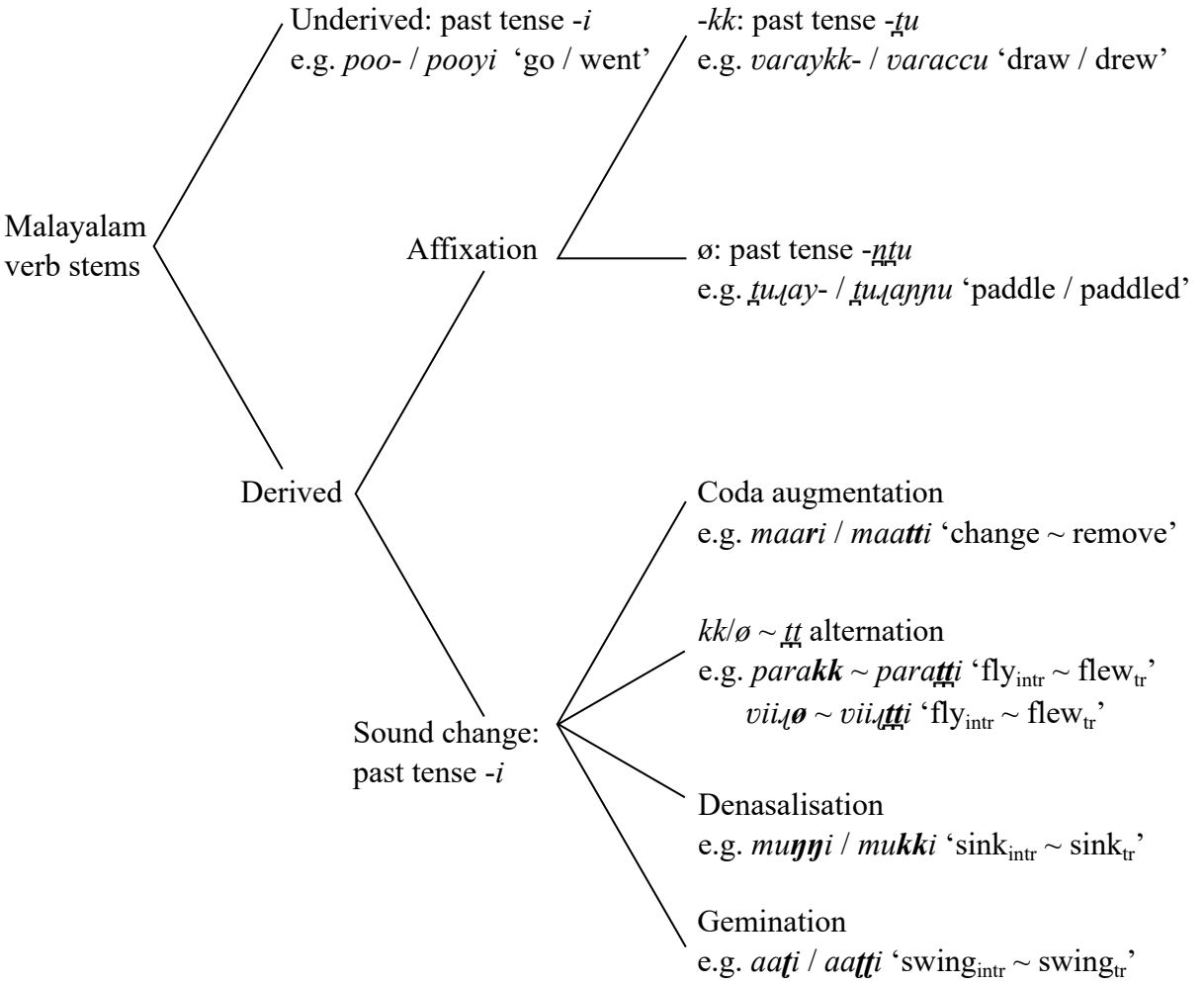


Figure 6.1 Malayalam verb stems.

Malayalam does confer a typological advantage in acquisition given the favourable properties of Malayalam inflections such as phonological and morphological transparency, biuniqueness, and productivity. We have also seen that children are phonologically mature and are equipped to discover the target phonotactics even before they begin their speech production (Guasti, 2002; Demuth, 2001; Jusczyk et al., 1993; Mehler et al., 1988; Werker & Tees, 1984, etc.). Phonologically driven inflectional assignment should be relatively easier to acquire than those inflections that require other information such as lexical features or paradigmatic class or morphological complexity of the stem. This should increase the formal complexity of the rules and thus impact learning. Given the past tense allomorphy, we expect the *-i* marker, being

phonologically determined, to be the most productive and error free. Since the original conditions of affixation are not recoverable synchronically, children cannot predict which affix might be appropriate with which set of verbs. As a result, we should expect errors to be present in the production of the allomorphs -*ṯu* and -*ṉṯu*. We examine each past tense marker in turn in the following[11].

6.6.1 Production of the past tense affix -*i*

The choice of the marker -*i* is governed by the syllable weight considerations. All the underived stems (except a small class of CVC stems) and the stems that are derived through sound change (valency change operation) take the marker -*i*. It is also the marker that has the most number of types and tokens in both the children's speech.

The marker -*i* is present in both the children's speech from the very first recording session onwards at the ages of 1;9.14 in A and 2;3.28 in H, as seen in Figure 6.2. More than 80% of the tokens comprises the past tense form *pooy-i* [go-PST] in both children's transcripts. While A starts productively using the marker with three different stem types during the first recording itself (1;9.14), H starts productively using it at the age of 2;4.14 with three different stem types[12].

(113) amboottiilɨ pooy-i (A, 1;9.14)
temple-LOC go-PST
[ɲaan] ambalaṯṯ-il pooy-i [adult form]
[I] temple-LOC go-PST
'I went to the temple.'

(114) ootti-kaaɾ-an pooy-i (H, 2;3.28)
auto-PL-M go-PST
oottoo-kaaɾ-an pooy-i [adult form]
auto-PL-M go-PST
'The auto-rickshaw driver left.'

In A's transcripts, there are 231 instances of 31 underived verb types inflected with the past tense marker -*i* while in H's transcripts, we find 89 instances of 22 types. Both A's and

[11] Appendix D contains the complete list of past tense forms present in the children's transcripts.

[12] As mentioned earlier in the previous chapter, there is a random pattern to the graph since the child produces more past tense forms in some sessions than in the others. This cannot be controlled for since it is not an experimental data. This does not have any bearing on the child's morphological competence per se.

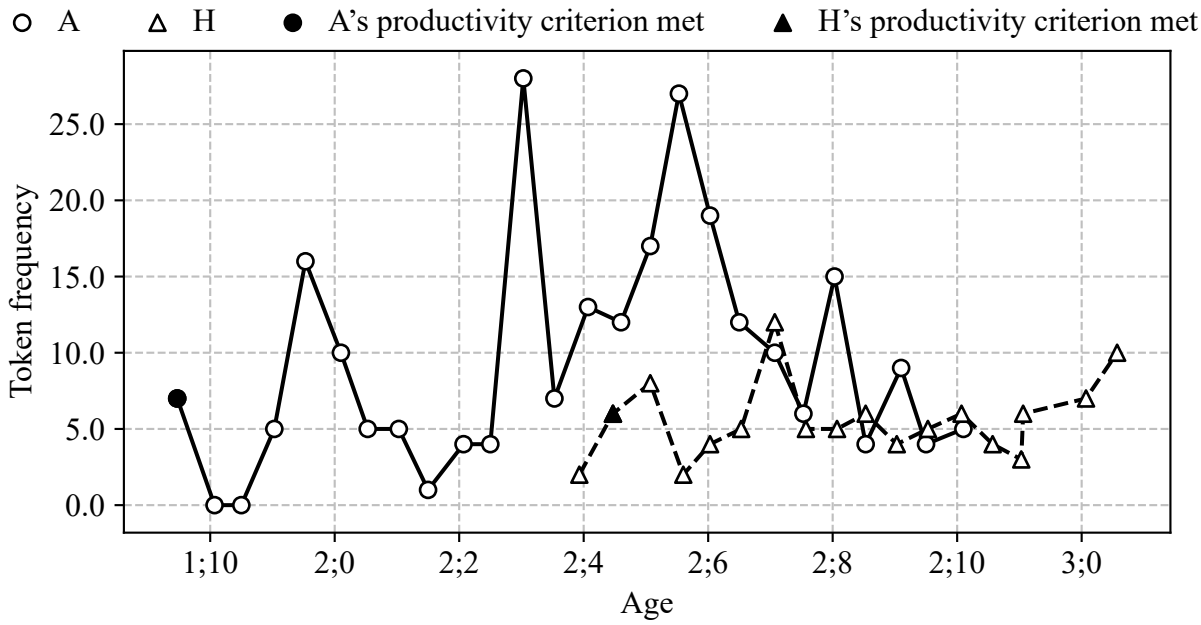


Figure 6.2 TAM inflections - Past *-i*.

H's past tense productions involving *-i* in underived stems include heavy syllabled verbs such as *keer-i* 'entered', *pooy-i* 'went', *kiʈʈ-i* 'got' etc. All the 89 instances are correctly marked. Of the valency changing derived verbs that take *-i*, A's transcripts have 13 types and 22 tokens while H's transcripts have eight types and 11 tokens. These derived verbs comprise the ones that have undergone gemination (e.g. *aaʈʈ-i* 'swung'), coda augmentation (e.g. *maatt-i* 'removed'), denasalisation (e.g. *ɲiikk-i* 'moved'), and *kk*/ø ~ *ʈʈ* alternation (e.g. *pa-ʈʈ-i*[13] 'flew', *kama-ʈʈ-i* 'toppled'). These are all correctly marked with the expected marker *-i* as well. The children's performance is hardly surprising since it is the phonological shape of the stem that determines affix selection. In addition, children are also seen to use derived, heavy stems appropriately.

6.6.2 Production of the Past Tense suffixes *-ʈu* and *-ɳʈu*

While the marker *-ʈu* is first attested in A's speech at 1;11.1, it is already in place in H's speech in the first transcript at the age of 2;3.8 (first recording) and continues to be in use till the end of the recording sessions at the age 3;0.17 as shown in Figure 6.3. The marker *-ʈu* meets the productivity criterion of marking three different stems at 2;2.2 and 2;5.2 in A's and H's transcripts, respectively. Children's transcripts show that they do not have issues in forming the different surface structures of *-ʈu* and *-ɳʈu*, as can be seen from the list of past tense forms in Appendix D.

[13] Adult Form: *para-ʈʈ-i* [fly-TR-PST]

This is not surprising since the formation of different surface structures results from regular phonological accommodations. This shows children's competence in their target phonology.

(115) kuɲɲi puucca-kuɲɲ-um ʋalya puucca-kuɲɲ-um ca-t̪̪u (A, 2;9.15)
small cat-baby-CORD big cat-baby-CORD die-PST
'The big kitten and the small kitten died.'

(116) kaɲɲi ku-cc-aa? (H, 2;5.18)
porridge drink-PST-Q
[n̪̪ii] kaɲɲi kuɽi-cc-oo? [adult form]
[you] porridge drink-PST-Q
'Did you have porridge?'

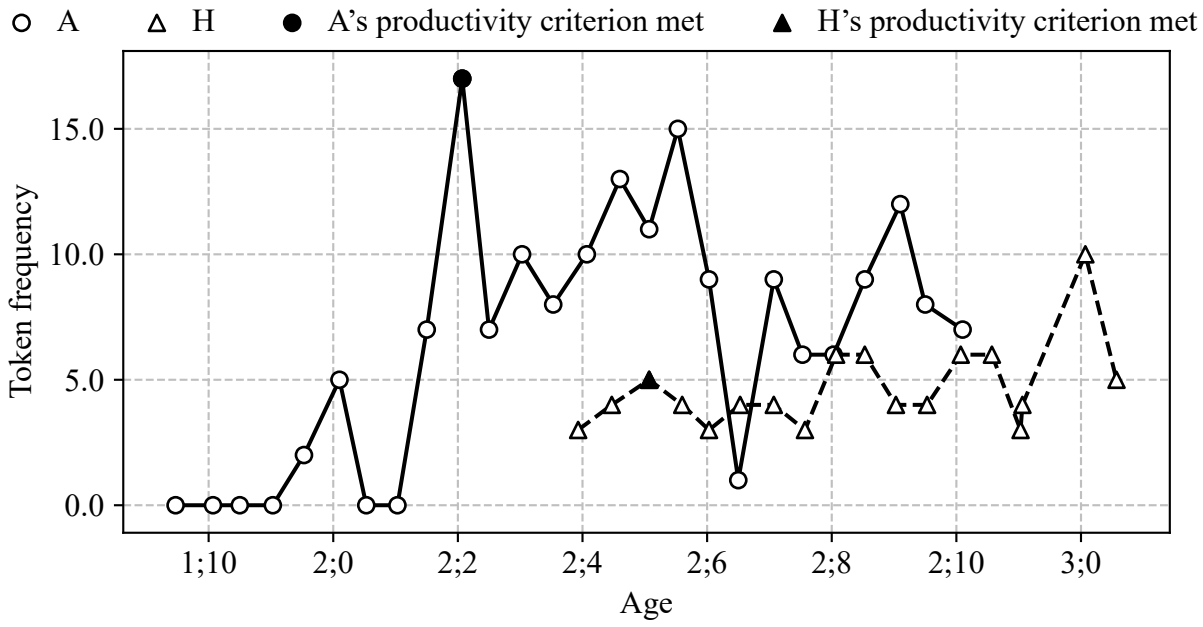


Figure 6.3 TAM inflections - Past -*t̪u*.

The marker -*n̪t̪u* is present from the first transcript onwards at the age of 1;9.4 for A. The marker occurs in restricted environments in Malayalam and is thus expected to be the least productive compared to its counterpart -*t̪u* and -*i*. Although -*n̪t̪u* has in absolute terms the maximum tokens, it is mainly the verb form *va-n̪n̪u* [come-PST] belonging to the underived class that comprises the majority of those tokens. The marker is productively used with five different stem types at the age of 1;11.1 for A, while the marker meets the productivity criterion at 2;5.2 in H's transcripts. The token frequencies of the marker in each transcript for both the children are shown in Figure 6.4.

(117) biykkatɨ paappam ʈi-ɳɳu (A, 1;9.2)
biscuit food eat-PST
[ɲaan] biʂkkatɨ ʈi-ɳɳu [adult form]
[I] biscuit eat-PST
'I ate biscuits.'

(118) ʈaaɻe ʋii-ɳɨ (H, 2;8.16)
down fall-PST
[ɲaan] ʈaaɻe ʋii-ɳu [adult form]
[I] down fall-PST
'I fell down.'

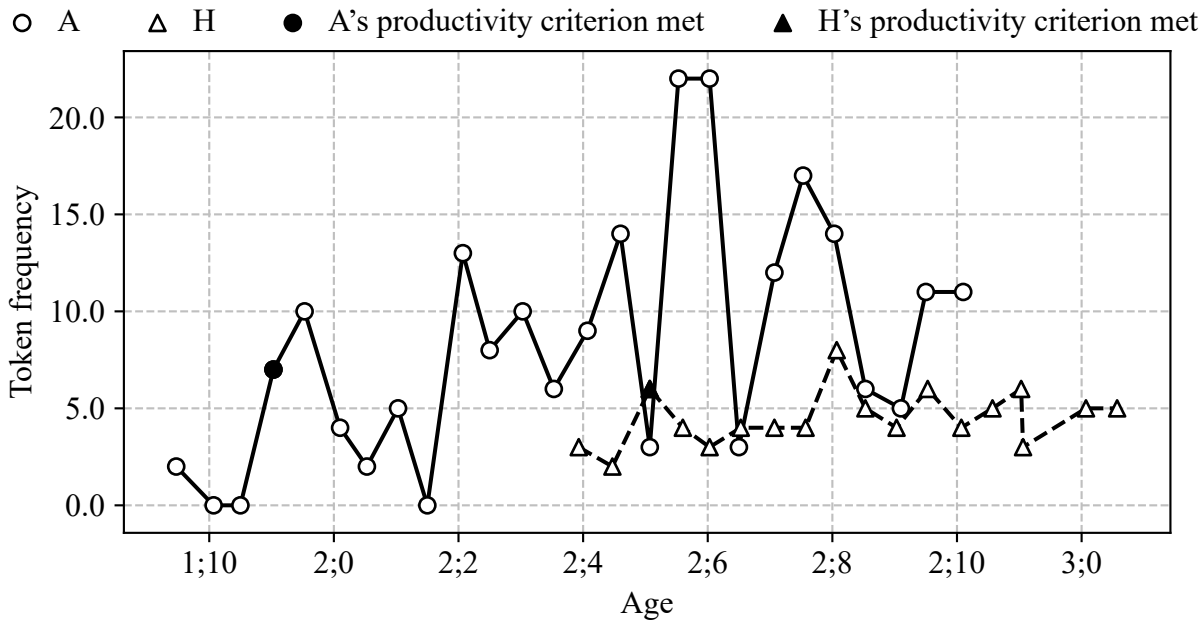


Figure 6.4 TAM inflections - Past -*ɳʈu*.

We established in Section 6.2 that the selection of the allomorphs -*ʈu* and -*ɳʈu* is determined in two different ways in the underived and derived class. Crucially, the selection of the affix is phonologically driven in the underived class of verbs (based on constraints related to syllable weight). In the derived verbs (denominal and valency-changing), the morphological component determines the affix selection (the rules of affixation being determined by the morphological feature [±TRANSITIVE]).

In the derived class, verbs that are derived through the addition of -*kk* choose the allomorph -*ʈu* and the ones that are derived by null-affixation choose -*ɳʈu*. We have already seen

that the past tense marking involving the phonologically driven marker *-i* is error-free in all the obligatory contexts. We expect that the different modes of selection of the past tense affix with different classes of verbs increase the complexity of affix selection. This may influence children's use of these affixes, which we focus on next.

There are four underived verb stems each in A's and H's transcripts that select the past tense marker *-ʇu*. The past tense forms present in A's transcripts are *ca-ʇʇu* 'died', *cey-ʇu* 'did', *iʈ-ʇu* 'put', and *viʈ-ʇu* 'released' whereas H's transcripts have *pey-ʇu* 'poured', *iʈ-ʇu* 'put', *ʇoʈ-ʇu* 'touched' and *iʂʈappeʈ-ʇu* 'liked'. While there are 22 tokens of *-ʇu* marking in the derived class in A's transcripts, H has seven instances.

Only two underived past tense types that choose the allomorph *-ɳɳu* are attested - *va-ɳɳu* [come-PST] with 60 instances in A's and 17 instances in H's speech and *ʇa-ɳɳu* [give-PST] with 22 and three instances in both A's and H's transcripts, respectively. The vastly smaller number of verb types reflects the restricted application of *-ɳɳu* assignment in underived Malayalam verbs, and the large number of tokens reflects the high-frequency of such verbs in the language. Both the children correctly mark all instances of the underived verbs of different kinds of stems with the expected past tense marker without exception.

Among the derived class that choose *-ʇu* as the past marker, A's transcripts have 24 types and 150 tokens while H has 20 types and 77 tokens. These verbs include both the verbs derived from nouns by *-kk* affixation as well as the transitives derived from intransitives by the *-kk*/*-pp* affixation. Examples of such verbs present in A's transcripts include *kuɭ-ccu*[14] 'bathed', *ci-ccu*[15] 'smiled', *oʈi-ccu* 'broke' etc. and H's transcripts include *oʈʈi-ccu* 'stuck', *kaʈi-cc*ɨ[16] 'bit', *ku-cc*ɨ[17] 'drank' etc.

There are 13 past tense types with 132 correct instances and seven past tense types with 59 correct instances of the derived class that select *-ɳɳu* in A's and H's transcripts, respectively. While H's past tense forms include the ø derived forms such as *vii-ɳu* 'fell', *ʇi-ɳɳu* 'ate', *ka-ɲɲu*[18] 'discarded' etc., A also has the exceptional forms mentioned in Section 6.2. The past tense forms belonging to this class are *iɾu-ɳɳu* 'sat' and *ɳaɻa-ɳɳu* 'walked' in addition to the ø derived forms such as *kaɾa-ɲɲu* 'cried' and *ɳana-ɲɲu* 'got wet'. As expected, there are errors involved in the assignment of the expected past tense marker in the derived forms involving both

[14]Adult Form: *kuɭi-ccu* [bathe-PST]
[15]Adult Form: *ciɾi-ccu* [smile-PST]
[16]Adult Form: *kaʈi-ccu* [bite-PST]
[17]Adult Form: *kuʈi-ccu* [drink-PST]
[18]Adult Form: *kaɭa-ɲɲu* [discard-PST]

-*ṯu* and -*ṉṯu* which will be discussed in detail in the next section. The type and token frequencies of each past tense marker are summarised in Table 6.13.

Table 6.13 Details of children's past tense marker production.

	A's Transcripts		**H's transcripts**	
Marker	**Underived**	**Derived**	**Underived**	**Derived**
-*i* (e.g. *keeri*, 'entered')	31 (231)[19]	13 (29)	22 (89)	8 (11)
-*ṯu* (e.g. *koṯuṯṯu*, 'gave')	4 (22)	24 (150)	4 (7)	20 (77)
-*ṉṯu* (e.g. *eriɲɲu*, 'threw')	2 (82)	13 (132)	2 (22)	7 (59)

6.6.3 Errors and Error Analysis

Notably, there are no omission errors or uninflected verb forms in any of the transcripts of the children. As we have seen, present and future tense marking are unexceptional and error-free. However, past tense marking in both the children is interspersed with commission errors which typically involve the use of the wrong past tense morpheme. Given our discussion, we predict that these are to be expected in acquisition since there are two kinds of past tense marking, one regular, productive and phonologically determined, and the other, morphologically determined but, synchronically, unpredictable and non-productive. Children's early competence in the phonology of the language (especially phonotactics, which is critical to word learning as well) helps them master the past tense inflections that are within the ambit of phonology. This primarily includes the past tense marker -*i*. The morphological rules determining past tense assignment in Malayalam is not synchronically transparent as to the relevant morphological features and even in the adult productions, the classes are collapsing and moving towards the phonologically conditioned marking. We find multiple past tense forms for some verbs and seemingly irregular forms for others.

As predicted, the errors in inflectional assignment involve the past tense marking of the verb class derived through affixation of -*kk* and ø which are arbitrary processes synchronically. There are no incorrect instances involving any of the verbs belonging to the underived class or the ones derived through sound change (valency changing). One specific error that is attested in both subjects in several sample sentences involves the morpheme -*i* which is used instead of

[19] Token frequencies are given in parentheses beside type frequencies.

the morpheme *-ɲɲu* with the verb *koll-* 'kill' (a zero-derived verb) (119). While A produces this commission error once each at 2.4.18 and 2.9.3, H produces it once each at 2.8.2 and 2.9.16, and twice at 2.8.16. We provide the utterances in (120)-(122)

(119) $kola_{\text{noun}} \xrightarrow[\substack{\text{denominal verb}\\(\text{coda augmentation})}]{} koll\text{-}_{\text{verb}} \xrightarrow[\text{heavy syllabled}]{} koll\text{-}i$ (predicted) / *ko-ɲɲu* (actual form)

(120)

		aʈi-ccu	*koll-i	(A, 2;4.18)
		beat-PTCP	kill-PST	
[ɲaan]	[iiccay-e]	aʈi-ccɨ	ko-ɲɲ u	[expected form]
[I]	[fly-ACC]	beat-PTCP	kill-PST	

'I killed the fly.'

(121)

	caanam	cooɳ-e	i-ccɨ	*koll-i	(H, 2;10.7)
	thing	phone-ACC	hit-PTCP	kill-PST	
[aa]	ʂaaɖʱanam	fooɳin-e	iʈi-ccɨ	ko-ɲɲu	[expected form]
[that]	thing	phone-ACC	hit-PTCP	kill-PST	

'That thing killed the phone.'

The same pattern is observed in the cross-sectional data as well.

(122)

	iiccay-e	*koll-i	(R, 4;2.8)
	fly-ACC	kill-PST	
[ɲaan]	iiccay-e	ko-ɲɲu	[expected form]
[I]	fly-ACC	kill-PST	

'I killed the fly.'

While studies in the acquisition of Malayalam (and other Dravidian languages) are few, Girija Devi (1972) also observes in her study that her subject produces the incorrect past tense form **kolli* 'killed' instead of the correct form *koɲɲu* 'killed'. She attributes it to learning by analogy through similarly structured verbs like *ʈall-* 'beat' with the past tense form *ʈalli* 'beat'. However, as we saw in Section 6.2, despite ending in geminates, verbs like *vell-* 'challenge' and *coll-* 'say' which are lexically similar to *koll-* 'kill' form their past tenses using *-ɲɲu* because they have been originally derived from nouns albeit through null-affixation/zero derivation. However, the fact that *vell-* 'challenge' and *coll-* 'say' have two alternating past tense forms *ve-ɲɲu* [challenge-PST] and *vell-i* [challenge-PST] and *co-ɲɲu* [say-PST] and *coll-i* [say-PST] demonstrate the increasing shift in usage towards the predictable form derived by phonological rules rather than the one based on diachronically visible morphological constraints which are arbitrary

in Modern Malayalam. This shift is not unlike the English past tense form *dived* (*dive-dived*) in place of the irregular *dove* (*dive-dove*). Interestingly, A's production reveals this bias. She uses *-i* on a geminate-ending high-frequency verb stem *koll-* 'hit' instead of retaining the arbitrary marker *-ṉṯu* as shown in (123). A has 5 instances of this verb in her speech out of which she correctly forms the past tense *koṇṯu* 4 times.

(123) bukkɨ ʋaay-cc-appam kaɳɳ-ii *koll-i (A, 2;3.16)
book read-PST-ADV eye-LOC hit-PST
[ammu] bukkɨ ʋaay-cc-appam [bʱaɖ ɾay-uṯe] kaɳɳ-il [expected form]
[ammu] book read-PST-ADV [Bhadra-GEN] eye-LOC
ko-ɳṯu
hit-PST
'The book hit (Bhadra) in the eye when (I) read it.'

The overgeneralised form **koll-i* is attested in the cross-sectional data as well.

(124) urumbi-nɨ kuṯṯɨ *koll-i (R, 4;2.8)
ant-DAT stab hit-PST
urumbi-nɨ kuṯṯɨ ko-ɳṯu [expected form]
ant-DAT stab hit-PST
'The ant got stabbed.'

Such overextension of the phonologically selected marker *-i* is seen in the bilingual acquisition of English and Malayalam as well (Raghunathan, unpublished data). The bilingual child D is seen to over-extend the marker *-i* to the zero-derived, denominal verb *vee-* 'boil'. The child incorrectly substitutes the phonologically selected marker *-i* instead of the expected morphologically selected marker *-ṉṯu*.

(125) aḍɨ *ʋeeʋ-i-iʈ-ill-e? (D, 2;6.5)
that boil-PTCP[20]-PFV-NEG-Q
aṯɨ ʋe-ṉṯ-iʈ-ill-ee? [expected form]
that boil-PTCP-PFV-NEG-Q
'Isn't it boiled?'

(Raghunathan, unpublished data)

Another example from the same bilingual corpus involves the formation of the incorrect past tense *ʈaɳukk-i** 'cool-PST' of the denominal verb *ʈaɳukk-* 'cool' formed by *-kk* derivation.

[20]This is the past participle which has the same form as the corresponding past tense marker in Malayalam.

Here, instead of the expected morphologically selected past tense marker *-ṯu*, the child chooses the marker *-i*. This is not unwarranted since *ṯaɳukk-* meets the superheavy syllable criterion and the selection of *-i* shows the child's adherence to the syllable weight constraint.

The above errors show clearly that children overextend the more regular, phonology driven rules of inflectional assignment. The errors underscore the difficulty in acquiring a synchronically non-productive past tense rule (i.e., the selection of *-ṯu* or *-ɳṯu* depending on the morphological feature [±TRANSITIVE]) and the relative ease in acquiring the phonologically driven inflectional selection of *-i*.

Two other examples of A's overgeneralisation errors at 2;5.16 and 2;6.1 involve the exceptional verbs *iɾiykk-* 'sit' and *ṉilkk-* 'stand' that select the past tense marker *-ɳṯu* rather than *-ṯu* despite these being derived denominal verbs via *-kk* affixation. Here, instead of selecting the grammatically correct form *-ɳṯu*, she uses the default past tense form *-ṯu* chosen by the verbs that are derived through *-kk* affixation as explained in Section 6.2.

(126) ceecc-i ṉookk-i-*iɾi-cc-appam oɾu kaʈ-eelɨ oɾu kokkin-e (A, 2;5.16)
sister-F look-PTCP-sit-PST-ADV one shop-LOC one crane-ACC
kaɳ-ʈu
see-PST
ceecc-i ṉookk-i iɾu-ɳṉ -appam oɾu kaʈ-eelɨ oɾu kokkin-e [expected form]
sister-F look-PTCP sit-PST-ADV one shop-LOC one crane-ACC
kaɳ-ʈu
see-PST
'When sister looked, she saw a crane.'

(127) iʊiʈe *ṉicc-appoo kot-eem koɳʧi-pook-uʊ-aayiɾuɳṉu (A, 2;6.1)
here stand-PST-ADV umbrella-CORD with-go-PTCP-AUX.PST
iʊiʈe ṉi-ɳṉ-appooɭ kot-eem koɳʧi-pook-uʊ-aayiɾuɳṉu [expected form]
here stand-PST-ADV umbrella-CORD with-go-PTCP-AUX.PST
'While I stood here, (she) was walking with an umbrella.'

The errors in the acquisition data reveal the mismatch between the child's grammar and the target grammar. We expect zero-derivation or null affixation itself to be a problem for the child. However, we do not have enough acquisition data to statistically validate this assumption. Given the complexity of the past tense morphology in Malayalam and the synchronic arbitrariness of the morphological constraints involved, the young learner is expected to produce errors

in the output showing a bias towards the predictable rule-bound inflectional assignment governed by the phonological component. If the unpredictability of the derivation rules has an effect on the acquisition of past tense assignment as evidenced by the errors discussed above, it is only logical that it will have an effect on the acquisition of verbs and verbal inflections in general. Although we do not have enough errors and correct instances to contrast this with zero-derived denominal/deadjectival verbs, we do have enough data involving the *-kk* derived verbs. Additionally, A is seen to omit the *-kk* morpheme in her transcripts at the age of 2;4.18. The omission of the marker apart from showing incomplete learning of the derivation of denominal verbs, demonstrates that the marker is not part of the underived stem, i.e., the root is *vey-* and not *veykk-* 'put' and the child is aware of this.

(128) ivi[e *ʋeyy-aam (A, 2;4.18)
here put-MOD
[ŋamu-kki] [bukki] ivi[e ʋeykk-aam [expected form]
[we-DAT] [book] here put-MOD
'Let's put the book here.'

In order to examine whether the unpredictability of the derivation rules has an effect on the verb acquisition, a Fisher's exact test has been run separately on each child's speech samples by analysing the verb forms that have *-kk* and the ones that do not have *-kk* in their stems, since verbs derived through *-kk* affixation are the ones with more exceptions and show greater arbitrariness in inflectional selection. Interestingly enough, the data show that both the children produce greater number of stem errors in verbs with *-kk* than in the ones without *-kk* . Tables 6.14 and 6.15 provide the details of stem errors present in A's and H's speech, respectively.

Table 6.14 Stem errors in A's verbs.

	with *-kk*	**without *-kk***
Incorrect	18	11
Correct	702	1947

Table 6.15 Stem errors in H's verbs.

	with *-kk*	**without *-kk***
Incorrect	14	0
Correct	174	693

Fisher's exact test for the verb productions of A and H give the p-values 6.81×10^{-8} and 2.74×10^{-10}, respectively, which suggest that the correct forms to error ratio is statistically significant. This offers further validation to our argument that the arbitrariness of derivation rules that force the children to depend solely on statistical learning resulting in the incorrect inflectional

assignment since phonologically similar stems take different past tense markers in Malayalam. This complicates the course of the acquisition of verbal inflections, especially, past tense marking since the other verbal inflections are assigned invariably without exceptions. This also provides a strong reason for why there is a shift towards phonologically driven past tense marking in modern Malayalam, especially since children are the agents of change in a language's grammar.

6.7 Conclusion

The analysis of past tense formation patterns show that there are two past tense markers: *-i* and *Tu*, with the latter realising two variants *-ṯu* and *-ṉṯu* in predictable environments. The synchronic arbitrariness of the derivation rules poses problems for acquisition and creates a shift from the inflectional marking based on morphological class distinctions towards the predictable, phonotactically driven inflectional selection. The acquisition data, especially the types of errors found, give support to the idea that past tense marking works differently in the two proposed classes and that Malayalam is slowly shifting towards inflectional assignment that is phonologically driven. Despite being a language that typologically enables inflectional development, the fact that past tense marking is interspersed with errors calls for a closer look at the factors that determine the relative ease/difficulty in acquiring a marker. With the factors that aid inflectional development already in place, we see that mode of selection of the morpheme has a crucial role to play in the acquisition of past tense markers: i.e, the phonologically selected past tense markers are acquired relatively easily and uniformly, compared to the morphologically selected past tense patterns.

Chapter 7

The Grammar and Acquisition of Malayalam Nominal Inflections

7.1 Malayalam Nominal Inflections

Languages vary in the extent to which they mark grammatical information on nominals directly using affixes. Case, number, and gender are the three main nominal inflectional categories. Case inflections encode the grammatical role of a noun in a sentence such as the subject, the object, the indirect object, possessor, location, instrument, and so on. More precisely, case morphemes show the relationship between an argument and the predicate. Theoretically, morphological case derives from an argument's structural position in a sentence (*structural case*) or is assigned lexically (*inherent case*, including quirky case on subjects and objects).

Many languages like Finnish, Turkish, Japanese, Tamil, Malayalam etc. have rich case morphology and a number of inflections that overtly mark them. For example, in the following Japanese sentence (129), *-ga*, *-ni*, and *-o* signal the subject, indirect object, and the direct object, respectively. Since there are separate markers to denote the different arguments, scrambling of the arguments does not change the core meaning of the sentence[1].

(129) taroo-**ga** hanako-**n**i hon-**o** age-ta
taro-NOM hanako-DAT book-ACC give-PST
'Taroo gave a book to Hanako.'

Thus, the sentence *Taroo-ga hon-o Hanako-ni age-ta* also means 'Taroo gave a book to Hanako'. Certain languages like English do not have an elaborate case-marking system and,

[1]Repositioning of arguments can, of course, add focal meaning to the moved argument etc.

typically, such languages depend on word-order to encode the grammatical relations of the arguments. For example, the sentence *The kids liked the Avengers* does not mean the same thing as *The Avengers liked the kids*.

Malayalam, like the other Dravidian languages, has a well-developed case system. The case morphemes of Malayalam are phonologically salient, morphologically transparent, and can be easily separated from their stems. Malayalam inflects for seven cases: nominative, accusative, dative, genitive, sociative, locative, and instrumental, as shown in Table 7.1. Among the case markers, the nominative is unmarked or phonologically null. Although the nominative typically instantiates the subject, the accusative the object, the dative the indirect object and so on, case-assignment is a selectional property of the predicate and as such, the match between grammatical roles and case can be different, notably in the case of dative subject predicates, some of which also permit nominative objects.

Table 7.1 Malayalam case markers.

case	***puuca* 'cat'**	***kuyil* 'cuckoo'**
nominative	puucca	kuyil
accusative	puuccay-**e**	kuyili-**ne**
dative	puuccay-**kk*i***	kuyili-**n*i***
genitive	puuccayu-**ʈe**	kuyili-**nte**
sociative	puuccay-**ooʈ*i***	kuyili-**nooʈ*i***
locative	puuccay-il	kuyil-il
instrumental	puuccay-aal	kuyil-aal

The allomorphy in case marking is phonologically conditioned. A uniform pattern is observed in the selection of variants by the different class of nominal stems. Morphemes signalling the accusative, dative, genitive, and sociative have two variants, which are shown in boldface in Table 7.1. The nominal stems which have an underlying palatal off-glide [y][2] choose the allomorphs *-e*, *-k*i*̷*, *-ʈe*, and *-ooʈ*i*̷* and the rest of the stems choose *-ne*, *-n*i*̷*, *-nte*, and *-nooti*̷ to signal the accusative, dative, genitive, and sociative, respectively. The phonological variants comprising the palatal [k], the retroflex [ʈ], the central vowel [e], and the back vowel [o] are chosen by the former class of stems so that both the stem-ending and suffix-initial phonemes

[2]Recall that in Section 1.2, we discussed the loss of the off-glide in word-endings as a feature of Malayalam that distinguishes it from Tamil. This off-glide is there in the underlying structures and surfaces only during affixation.

share the distinctive feature [−anterior]. This difference is shown in Table 7.2.

Table 7.2 Phonologically conditioned allomorphs of Malayalam case markers.

stem	**accusative**	**sociative**	**dative**	**genitive**
kiliy- 'bird'	kiɭiy-e	kiɭiy-ooʈɨ	kiɭiy-kkɨ	kiɭiy-uʈe
ʈaṯṯay- 'parrot'	ʈaṯṯay-e	ʈaṯṯay-ooʈɨ	ʈaṯṯay-kkɨ	ʈaṯṯay-uʈe
puuʋ- 'flower'	puuʋi-ne	puuʋi-nooʈɨ	puuʋi-nɨ	puuʋi-nte
kaʈal- 'sea'	kaʈali-ne	kaʈali-nooʈɨ	kaʈali-nɨ	kaʈali-nte
kayar- 'coir'	kayari-ne	kayari-nooʈɨ	kayari-nɨ	kayari-nte
maan[3]- 'deer'	maa-ne	maa-nooʈɨ	maa-nɨ	maa-nte
iṯaɭ[4]- 'petal'	iṯaɭi-ne	iṯaɭi-nooʈɨ	iṯaɭi-nɨ	iṯaɭi-nte
maraṯṯ[5]- 'tree'	maraṯṯi-ne	maraṯṯi-nooʈɨ	maraṯṯi-nɨ	maraṯṯi-nte

The accusative case marker *-e* is assigned to direct objects. The manifestation of the accusative marker depends on the lexical feature of *animacy* of the noun. The accusative marker is obligatorily, overtly marked for [+ANIMATE] nouns and is obligatorily dropped with [−ANIMATE] nouns. Thus, the accusative morpheme is morphologically selected on the basis of the feature [±ANIMATE].

(130) amma kuʈʈiy-kkɨ paal / puuccay-e koʈu-ṯṯu
mom child-DAT milk-ø / cat-ACC give-PST
'The mother gave milk/ a cat to the child.'

When both the subject and the object are inanimate entities, there is scope for ambiguity, given that Malayalam also allows scrambling. Accusative is overtly marked on inanimate objects only in such cases. Otherwise, it is never overtly marked on inanimate objects even to indicate specificity.

(131) kappal ʈiramaala-kaɭ bʱeeḍi-ccu
ship wave-PL. øACC break-PST

[3]Alternative forms such as *maani-ne* [deer-ACC], *maani-nooʈɨ* [deer], *maani-nɨ*, *maani-nte* also exist.

[4]However, the word-forms ending in the plural suffix *-kaɭ* (e.g. *kiɭi-kaɭ* 'birds') and the feminine suffix *-aɭ* (eg. *mak-aɭ* 'daughter') choose the [−anterior] markers as in *kiɭikaɭ-e* [birds-ACC], *kiɭikaɭ-ooʈɨ* [birds-SOC], *kiɭikaɭ-kkɨ* [birds-DAT], *kiɭikaɭ-ʈe* [birds-GEN], and *makaɭ-e* [daughter-ACC], *makaɭ-ooʈɨ* [daughter-SOC], *makaɭ-kkɨ* [daughter-DAT], *makaɭ-ʈe* [[daughter-GEN], respectively.

[5]This stem derives from *maram* 'tree', which becomes *maraṯṯ-* during affixation. This is seen in all the *-am* ending nominals where the stem ending becomes *-aṯṯ-*, at morpheme boundaries without exception.

The above sentence could either mean 'The ship broke through the waves' or 'The waves split the ship'. In such instances, the potential ambiguity can be resolved by marking the object with the accusative case.

(132) kappal ʈiɾamaala-kaɭ-e bʱeeɖi-ccu
ship wave-PL-ACC break-PST
'The ship broke through the waves.'
(Asher & Kumari, 1997, p. 204)

(133) kappali-ne ʈiɾamaala-kaɭ bʱeeɖi-ccu
ship-ACC wave-PL break-PST
'The waves split the ship.'
(Asher & Kumari, 1997, p. 204)

The dative marker has two variants *-ni* and *-kki* that range over different semantic and grammatical roles. The dative marks the indirect object denoting the recipient of ditransitive verbs (134), as well as semantic roles such as *benefactive* (135), *purpose* (136), *price/value* (137), and *time* (138). The subject of a certain class of verbs can also be marked with the dative. These include what have been called *experiencer* predicates (for example, *viɕakk-* 'be hungry') (139) as well as modal verbs (140) and copular constructions (141).

(134) amma kuʈʈiy-kki paal koʈu-ʈʈu (recipient)
mom.NOM child-DAT milk give-PST
'The mother gave the child milk.'

(135) amma kuʈʈiy-kki ʋeeɳʈi bʱakʂaɳam uɳʈaakk-i (benefactive)
mother.NOM child-DAT for food make-PST
'Mother made food for the child.'

(136) ɾaaman ameeɾikkay-il ɟooliy-kki pooy-i (purpose/reason)
rama.NOM america-LOC work-DAT go-PST
'Rama went to America for work.'

(137) aʋaɭ iɾupaʈi ɾuupay-kki miin ʋaaŋŋ-i (price/value)
she.NOM twenty rupee-DAT fish buy-PST
'She bought fish for twenty rupees.'

(138) ɾaaman muuɳɳɨ maɳiy-kkɨ pooy-i (time)
rama.NOM three time-DAT go-PST
'Rama left at three o'clock.'

(139) eni-kkɨ ʋiɕa-ɳɳu (experiencer)
I-DAT hungry-PST
'I was hungry.'

(140) ɾaama-nɨ pook-aɳam
rama-DAT go-MOD
'Rama wants to go.'

(141) ɾaama-nɨ ɾanʈɨ kuʈʈikaɭ uɳʈɨ
rama-DAT two children have.PRS
'Rama has two children.'

The genitive case has two phonologically conditioned variants *-uʈe* (e.g. *kuʈʈiy-uʈe* child-GEN) and *-nte* (e.g. *puuvi-nte*, flower-GEN). The genitive signals *possession* (142), *quantity* (143) and *material* of which an object is made (144).

(142) kuʈʈiy-uʈe kuʈa (possession)
child-GEN umbrella
'The child's umbrella.'

(143) oɾu paʋa-nte sʋarɳam (quantity)
one sovereign-GEN gold
'One sovereign of gold.'

(144) sʋarɳaʈʈin-te ʋaɭa (material / composition)
gold-GEN bangle
'golden bangle'

The morpheme *-ooʈɨ* marks the sociative case and is lexically assigned to indirect objects of certain verbs of 'saying' like *paray-* 'say/tell' (145), *cooḏiykk-* 'ask' (146), *vivaɾiykk-* 'explain' (147) etc.

(145) ɾaaman ʂiiʈay-ooʈɨ oɾu kat^{h}a para-ɲɲu
rama sita-SOC one story tell-PST
'Rama told Sita a story.'

(146) ɾaaman ʂiitay-ooʈɨ oru coodyam coodi-ccu
rama sita-SOC one question ask-PST
'Rama asked Sita a question.'

(147) ɾaaman ʂiitay-ooʈɨ aa ʂambʱavam ʋivaɾi-ccu
rama sita-SOC that incident explain-PST
'Rama explained that incident to Sita.'

The common locative marker *-il* occurs with both animate and inanimate nouns without allomorphic variation (148). An alternate locative marker *-aṯṯɨ* is used with place names ending in *-am* (149). However, certain nouns like *maɻa* 'rain' take both *-il* and *-aṯṯɨ* as in *maɻay-il* and *maɻay-aṯṯɨ* both meaning 'in the rain', where the former is more formal and the latter is more colloquial. While the marker *-il* can apply exceptionlessly, the assignment of *-aṯṯɨ* is lexically specified for individual items and for place names ending in *-am*. There is also another marker *-kal* which signals the location (150). This is the least frequently used locative marker and is synchronically non-productive in Malayalam.

(148) kuɭaṯṯ-il miin uɳʃɨ (location)
pond-LOC fish have.PRS
'There is fish in the pond.'

(149) ɲaan kottay-aṯṯɨ pooy-i (to-place name)
I.NOM kottayam-LOC go-PST
'I went to Kottayam.'

(150) ɾaaman paʈiy-kkal kaa-ṯṯɨ ɳi-ɳɳu
rama.NOM step-LOC wait-PTCP stand-PST
'Rama waited at the steps.'

The instrumental case is marked by the morpheme *-aal* (151). This is one of the least frequently used case markers in Malayalam, especially since the postposition *koɳʈɨ* 'with' (152) is the one that is used to signal instrument colloquially, thus restricting the marker *-aal* to formal contexts. The postposition *koɳʈɨ* freely alternates with the case marker *-aal* except in passive constructions, where the morpheme *-aal* signals the demoted agent (153).

(151) ɾaaman ɾaavaɳa-ne ʋaaɭ-aal ʋeṯṯ-i
rama.NOM ravana-ACC sword-INS cut-PST
'Rama wounded Ravana with a sword.'

(152) ɾaaman ɾaaʋaɳa-ne ʋaaɭi koɳʈi ʋeʈʈ-i
rama.NOM ravana-ACC sword with cut-PST
'Rama wounded Ravana with a sword.'

(153) ɾaaʋaɳan ɾaaman-aal kolla-ppeʈʈu
ravana.NOM rama-INSTR kill-PASS-PST
'Ravana was killed by Rama.'

In languages that mark grammatical gender, nouns are assigned different genders. Gender may be aligned with sex distinctions (as in English and Tamil, the male is masculine gender, female feminine gender, and everything else is neuter), or arbitrarily two or three genders (e.g. Marathi, French, German[6]) or many more (e.g. Mountain Arapesh with thirteen genders). The gender systems also use phonological cues (e.g. Sanskrit) or semantic classes (e.g. Swahili) to assign a noun to a gender class. Lexical features like animacy or human-ness also play a role. Some common systems of gender division include masculine-feminine (e.g. Celtic languages), masculine-feminine-neuter (e.g. Sanskrit), animate-inanimate (e.g. Anatolian languages), common-neuter (e.g. Swedish) etc. Gender is an inherent lexical feature of a noun[7] but finds morphological presence in agreement of various kinds (adjectival, verbal etc.)

Malayalam has a three-gender system (like the other Dravidian languages and English) - masculine, feminine, and neuter -which corresponds to the natural sex classification. Hence, only the nouns with the semantic feature [+MALE] are assigned the masculine gender and only the nouns with the semantic feature [+FEMALE] are assigned the feminine gender. The rest are assigned the neuter gender including other living organisms and inanimate entities. Hence nouns like *kaɻuʈa* 'donkey', *kuʈa* 'umbrella' and so on belong to neuter gender. In the contexts where the gender of an animal has to be specified, the masculine adjective *aaɳ* 'male' or the feminine adjective *peɳ* 'female' is added. Examples of such contexts are the noun phrases *aaɳ kaɻuʈa* 'male donkey' and *peɳ kaɻuʈa* 'female donkey'. The only exceptions to this classification are the nouns referring to cattle. [+ANIMATE] nouns like *paɛu* 'cow', *eruma* 'she-buffalo' and *kaaɭa* 'ox', *pooʈʈi* 'he-buffalo' are assigned feminine and masculine genders respectively. In Malayalam, gender information must be specified in the lexicon for each noun.

[6]Again, ignoring that by and large female is marked feminine and male as masculine even in languages with random grammatical gender classification.

[7]The predictability of this feature, of course, rests on the degree of transparency that we find in the particular languages.

Gender marking[8] in Malayalam is not phonologically driven but is governed by the morphology depending on the feature [±HUMAN]. However, there are certain exceptional cases where the nominals signalling the animals *koraŋŋ**an*** 'monkey' and *kurukk**an*** 'fox' as well as celestial bodies *s̪uur**yan*** 'the sun' and *caṉḍ**ran*** 'the moon' are also marked with the masculine suffix. The masculine gender inflectional suffix in Malayalam is *-an*. There are three inflectional suffixes to indicate the feminine gender: *-i*, *-aṯṯi*, and *-aḷ*. The marker *-i* is more productive and more frequently occurring while *-aṯṯi* is restricted to fewer number of stems. The distribution of the marker *-aḷ* is further restricted in its distribution. It typically manifests with pronouns, as in the examples *avaḷ* 'she$_{remote}$', *ivaḷ* 'she$_{proximal}$', and very few nouns like *mak-aḷ* 'daughter'. The feminine marker *-aḷ* is synchronically non-productive in Malayalam. Table 7.3 shows the Malayalam gender suffixation system.

Table 7.3 Malayalam gender suffixation.

Masculine	**Feminine**
aaɳɨ 'male'	*peɳɳɨ* 'female'
kaaḷa 'ox'	*paɛu* 'cow'
eruma 'she-buffalo'	*pooṯṯɨ* 'he-buffalo'
makan 'son'	*makaḷ* 'daughter'
kaḷḷ-an 'male thief'	*kaḷḷ-i* 'female thief'
ceeṯṯ-an 'elder brother'	*ceecc-i* 'elder sister'
yaɻamaan-an 'male superior'	*yaɻamaan-aṯṯi* 'female superior'

The final grammatical feature is that of number. Most languages mark number on the noun while there are a few which mark it on the verb (e.g. Rapanui). Languages differ in terms of the number values they are capable of distinguishing. Languages like Hindi, English, and Tamil make a distinction between the singular and the plural. There are systems in which the singular and the plural are supplemented by a dual (e.g. Central Yupik, Modern Arabic and Irish) that refers to 'two distinct real-world entities', as well as systems in which there are four number values (e.g. Larike) such as the singular, the dual, the trial, and the plural, among which the trial refers to 'three real-world entities'. Among the classical languages, only Sanskrit had the dual number while in Greek, the only remnant of the dual was the numeral for 'two', and in

[8]There is no particular suffix to indicate the neuter gender. The affix *-am* which also signals neuter is a derivational affix.

Latin, it was lost as a grammatical category. Certain languages like Baiso, Warlpiri and Serbo-Croatian mark the paucal number which refers to a small set of real-world entities (as opposed to many). Number systems can also make distinctions between collective and distributive plurals in addition to the above and some systems show contrasts in the pronominal system but not on nouns in general.

Malayalam distinguishes the plural with an overt marker and the bare noun is usually singular (except where the noun is an inherently plural or mass noun). The lexical features [±HUMAN] and [±ANIMATE] of the nouns determine the degree of obligatoriness in marking plurality. The plural allomorphs are given in Table 7.4. If a noun is [+HUMAN and +ANIMATE], then the plural marker *-maar* is used. Nouns that bear the features [−HUMAN, +ANIMATE], are marked with the affix *-kaɭ*[9] (e.g. *paɛu-kkaɭ* 'cows'). In stems that end in nasals, the marker also undergoes nasalisation as in *maɾaŋ-ŋaɭ* 'trees' from *maɾam* 'tree'. Finally, plural morpheme can be optionally omitted if the noun is [−ANIMATE, −HUMAN][10].

Table 7.4 Malayalam plural formation.

Lexical Features	Affix	Example
[+HUMAN]	*-maar*	*amma-maar* 'mothers'
[−HUMAN, +ANIMATE]	*-kaɭ*	*kiɭi-kaɭ* 'birds'
[−HUMAN, −ANIMATE]	*ø~-kaɭ*	*ɾaɳʈi kaṯeeɾa* [two chair] 'two chairs'

Malayalam also has two epicene plural markers *-kaar* and *-ar*, which are lexically conditioned as shown in Table 7.5. The former is a humanising marker with a number feature. The marker *-kaar* inflects [−ANIMATE] nouns to humanise them and further signals a set of entities including both male and female members. For example, when *-kaar* inflects the nominal *kaʈa* 'shop', the resulting plural does not mean 'shops'. Instead, it means a group of 'shopkeepers' which include both men and women. In order to include only one particular sex, the gender markers are employed as shown in Table 7.6. The allomorphic variant *-ar* occurs with bound stems of Sanskrit origin such as *gaayak-* 'singer', *n̪arttak-* 'dancer' etc. These *-ar* inflected forms are typically used in formal contexts and thus, are not part of child-directed or colloquial speech.

[9]There are certain exceptions where [+HUMAN] nouns are also marked with *-kaɭ* instead of the expected *-maar*. Three such notable exceptions are *kuʈʈi-kaɭ* 'children', *s̪trii-kaɭ* 'women', *peɳɳiŋ-ŋaɭ* 'women'.

[10]Recall that accusative case was also unmarked for nouns that were [-ANIMATE]. Sometimes in colloquial speech, the marker *-kaɭ* is optionally omitted on [+ANIMATE] nouns as well.

Table 7.5 Epicene plural markers in Malayalam.

stem type	stem	marker	example
free stems	paɳi 'work'	*-kaar*	*paɳi-kkaar* [work-HUM.PL] 'workers'
bound stem	n̪arṯṯak- 'dance'	*-ar*	*n̪arṯṯak-ar* [dancer-PL] 'dancers'

Table 7.6 Differences in plural formation using *-kaar*.

Category	***kaṭa* 'shop'**	Gloss
Epicene plural	kaṭa-kkaar	[shop-HUM.PL] 'shopkeepers'
Masculine singular	kaṭa-kkaar-an	[shop-HUM.PL-M] 'male shopkeeper'
Masculine plural	kaṭa-kkaar-an-maar	[shop-HUM.PL-M-PL] 'male shopkeepers'
Feminine singular	kaṭa-kkaar-i	[shop-HUM.PL-F] 'female shopkeeper'
Feminine plural	kaṭa-kkaar-i-kaɭ	[shop-HUM.PL-F-PL] 'female shopkeepers'

7.2 Cross-Linguistic Patterns in the Acquisition of Nominal Inflections

The acquisition of nominal inflections is well-studied across the world languages and documented for a variety of European languages in particular. Since the ways in which world languages employ the inflections on nouns are variegated, it is interesting and important to see how children grapple with the acquisition of inflections in their ambient language which may employ different inflectional systems. The various properties of nominal inflectional acquisition in English, Arabic, Turkish, etc. are examined in this section. Early English plurals show overregularisation of the irregulars. Arabic plurals include more than one default form like Malayalam. Special focus is given to the acquisition of Turkish owing to its grammatical similarities to the Dravidian languages. Non-canonical case marking in Tamil (Lakshmi Bai, 2004), Telugu (Usha Rani & Sailaja, 2004) and Hindi (Narasimhan, 2005) will be discussed in the next chapter.

Marcus (1995) examined the plural production of ten English speaking children, five boys and five girls, between the ages 1;3 and 5;2. Seven out of the ten children produced overgeneralisation errors like *foots** and *mans** for *feet* and *men*, respectively. These children produced the irregular forms correctly in the beginning before transitioning to produce the overregularised

forms thus adhering to the *U-shaped* developmental pattern (Marcus et al., 1992). Overregularisation errors were not attested in the remaining three children's transcripts. From the acquisition patterns exhibited by the children, Marcus et al. concludes that acquisition of inflections is rule-based, and that once the children deduce the inflectional assignment rule, they tend to apply it by default, across all inputs till they restore the irregularities.

Ravid and Farah (2009) analyse the acquisition of noun plurals in Palestinian Arabic. Palestinian Arabic employs both internal stem change as well as linear suffixation to form plurals. The plurals in the former class are called *broken plurals* (e.g. *kaleb*/*kilaab* 'dog/s') and these irregular plural forms occur in four different patterns. The plurals formed out of linear suffixation are called *sound plurals* and these are of two kinds, sound masculine *-i*:*n* (e.g. *falla*:*H*/*fallaHi*:*n* 'farmer/s') and sound feminine *-a*:*t* (e.g. *kubba*:*y*/*kobaya*:*t* 'glass/es'). A greater number of stems select the feminine plural suffix which is more productive than the masculine one.

Ravid and Farah used 45-60 minutes long speech samples, of a single, monolingual child between the ages 1;8 - 2;8, which were recorded once every week. Among the plural tokens produced by the child, the most frequent were the broken plurals, and then the sound-feminine plurals. The most productive pattern among the plurals was also the sound-feminine plural form as indicated by the number of overgeneralisations and overregularisations. The other plural forms were generally correct in their occurrences. In spite of being a regular plural morpheme like the sound feminine, sound masculine plurals were not only fewer in the overall inflectional tokens produced by the child, but also late to emerge in the child's speech. Since neither morpheme has an advantage over the other in terms of phonological salience or transparency, the authors attribute the lateness in emergence and the limited number of productions of the sound masculine plural to its low frequency in the input (and more generally in the language itself).

Ketrez and Aksu-Koç (2009) study the early development of case and number in Turkish. Being an agglutinative language, Turkish has a regular and transparent inflectional system. The case system of Turkish is illustrated in Table 7.7.

Turkish differentiates six cases. Although Turkish case morphemes are highly frequent, nominal ellipsis is common in Turkish since it has a strong subject-verb agreement. This can negatively affect the token frequency of the case morphemes since the ellipsis of nominals will reduce the number of instances where the case markers manifest overtly. In Turkish, the plural

Table 7.7 Turkish case forms.

Case	***kedi* 'cat'**	***ben* 'I'**
nominative	kediø	benø
accusative	kedi-yi	ben-i
dative	kedi-ye	ben-a
genitive	kedi-nin	ben-im
locative	kedi-de	ben-de
ablative	kedi-den	ben-den
instrumental	kedi-yle	benim-le

(Ketrez & Aksu-Koç, 2009, p. 16)

is formed by adding the morpheme *-lAr*[11] to the singular stem, for example, *kedi - kediler* 'cat/s', *adam – adamlar* 'man/men' etc.

The authors use 20 minutes long speech samples of a monolingual Turkish child who was recorded twice a month from the age 1;3.3 till the age 2;0.4. The criteria for determining productivity of a rule included the following: a) an inflection should appear with at least two different nouns, b) a noun should appear in at least two contrasting inflected forms, c) the number of grammatical productions of an inflection should exceed the number of errors, and d) an inflection should appear in two consecutive recording samples. As the child progressed in age, both the overall number of inflectional paradigms as well as the number of individual members in the paradigms exhibited a significant increase. By the age of 2;0, the child's productions had a total of 39 inflectional paradigms with 24 two-membered ones, and three with seven to nine members. With age, the number of uninflected forms decreased, variations became stable, and the productions started approaching the adult pattern. Among the case forms, the accusative, the dative, and the locative case markers were the first ones to emerge. The accusative case marker was the first one to be productively used, followed by the locative, the dative, the ablative, and the genitive. The plural forms emerged later than the case forms.

The early acquisition of case markers despite their actual low frequency in the input due to nominal ellipsis is attributed to the morphological transparency of the case inflections. The authors ascribe the late emergence of the plural morpheme to its low frequency and to the fact that number is not strictly grammaticised in Turkish. The study stops at two years of age, and

[11] With vowel harmony on the affixal vowel.

the acquisition of nominal inflections was not yet complete at that point and the child was yet to learn the entire syntactic and semantic constraints on the inflections.

From the above studies, we see that the acquisition of inflection is rule-based, as observed in English plural marking (Marcus, 1995). We also see that inflectional acquisition is affected by factors such as frequency (in the input) as seen in Arabic (Ravid & Farah, 2009) and typological features such as morphological transparency as seen in Turkish (Ketrez & Aksu-Koç, 2009). These studies lead us to expect certain patterns in the acquisition of nominal inflections in Malayalam and to account for such patterns in terms of both typological and language specific features. In the rest of this chapter, we discuss the occurrence of each nominal inflection in the current data.

7.3 Acquisition of Malayalam Nominal Inflections

Malayalam, as we just saw, has a variety of case, number, and gender markers. The template of a Malayalam noun in which the different nominal markers are stacked together is **stem** + **gender** + **number** + **case** (154) except in the instances where the humanising marker *-kaar* is added to the stem. In the latter cases, the template is **stem** + **human suffix** + **gender** + **number** + **case** (155).

(154) kaḷ-an-maaṟ-e
thief-M-PL-ACC
'thieves'

(155) miin-kaaṟ-an-maaṟ-e
fish-HUM-M-PL-ACC
'fishermen'

The selection of these affixes is determined by the phonological properties of the stem or the overall syllable shape and, for some affixes, by the lexical properties on the nouns. In this section, we provide a survey of the nominal inflections in Malayalam as they occur in the children's productions. Figures 7.1 to 7.14 show the token frequencies of each nominal inflection in A's and H's transcripts[12]. The utterances (156) - (199) include examples of how each inflection is employed in the data. We begin with Table 7.8 which shows the details of the type and

[12] As noted in the discussion on verbal inflections, the graphs are random since the token frequencies of different inflections vary in each transcript.

token frequencies of each nominal inflection (including the variants). We will demonstrate in the discussion that there are omission errors involving the use of certain nominal inflections in obligatory contexts. The productivity criteria include a) 90% occurrence in obligatory contexts in three consecutive recordings for the markers that are omitted, and b) the use of the marker with three or more different stems.

Table 7.8 Nominal inflections in A's and H's speech.

			A's Transcripts		**H's Transcripts**	
	Category	**Marker**	**Frequency**		**Frequency**	
			Type	**Token**	**Type**	**Token**
Case	Accusative 1	*-e*	25	77	8	14
	Accusative 2	*-ne*	23	50	16	40
	Dative 1	*-nɨ*	9	176	3	35
	Dative 2	*-kkɨ*	24	141	6	9
	Genitive 1	*-nte*	24	234	13	81
	Genitive 2	*-ʈe*	9	48	19	146
	Sociative 1	*-ooʈɨ*	6	8	-	-
	Sociative 2	*-nooʈɨ*	1	2	-	-
	Locative 1	*-il*	53	290	26	113
	Locative 2	*-aṯṯɨ*	7	14	4	9
Gender	Masculine	*-an*	13	148	11	50
	Feminine	*-i*	1	376	1	2
Number	Plural 1	*-maar*	4	32	2	3
	Plural 2	*-kaɭ*	6	6	1	5
	Plural 3	*-kaar*	4	20	2	4

7.3.1 Case

In A's transcripts, the accusative variants *-e* and *-ne* emerge at 1;11.16 and 1;11.1, respectively, with the former having 25 types and 77 tokens and the latter having 23 types and 50 tokens. In H's speech, both the variants are first seen at 2;5.18. The variant *-e* has nine types and 14 tokens across transcripts while *-ne* has 16 types and 40 tokens. Figures 7.1 and 7.2, respectively, show

the instances of the accusative variants *-e* and *-ne* in the children's transcripts. A has 100% correct instances in three consecutive transcripts. H also has 100%, 90%, and 100% correct instances of accusative marking in three consecutive transcripts, respectively. The variant *-e* is productively used with three different stems at 2;0.3 and 2;11.18 in A's and H's transcripts, respectively, while the variant *-ne* is productively employed with three different stems at 2;1.1 and 2;7.17 in A's and H's transcripts, respectively. Recall that the accusative is obligatorily omitted on [–ANIMATE] nouns unless to resolve an ambiguous situation where both the subject and object nominals bear the feature [–ANIMATE]. H has such an instance in his transcripts (158). Both children omit the accusative in obligatory contexts which suggests that the acquisition of the morpheme is not yet complete. We return to these errors in Section 7.4 when we analyse the errors.

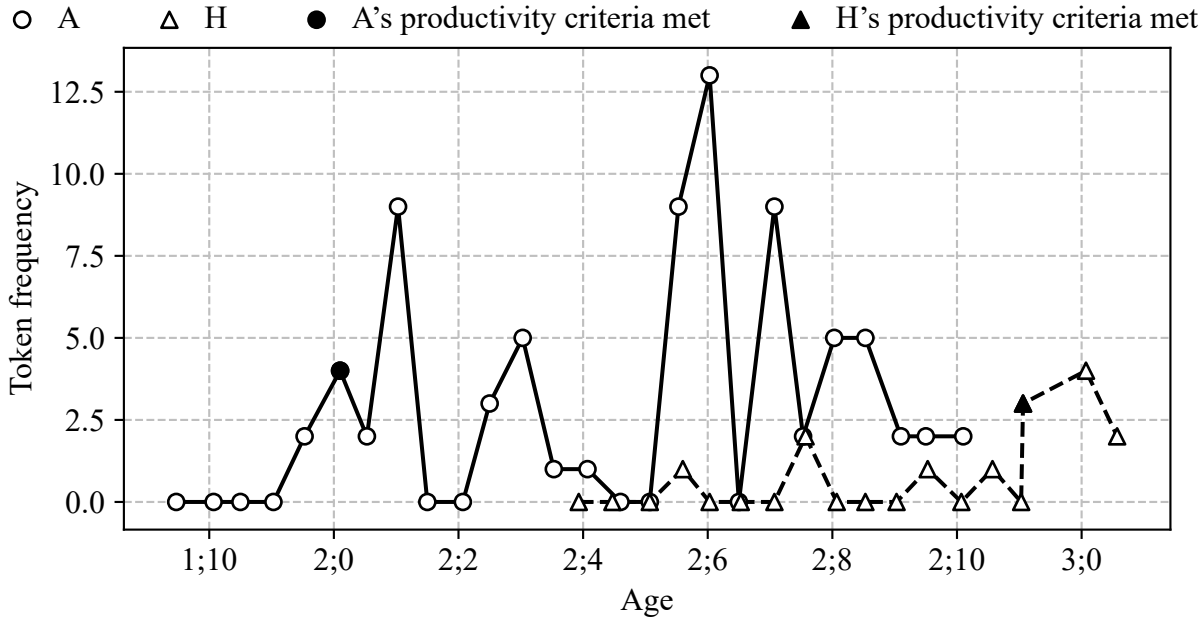


Figure 7.1 Nominal inflections - Accusative *-e*.

(156) amma kuɲɲi-ʋaaʋ-e eʈukk-aan pooy-ii (A, 2;7.2)
mom small-baby-ACC take-INF go-PST
'Mom went to get the baby.'

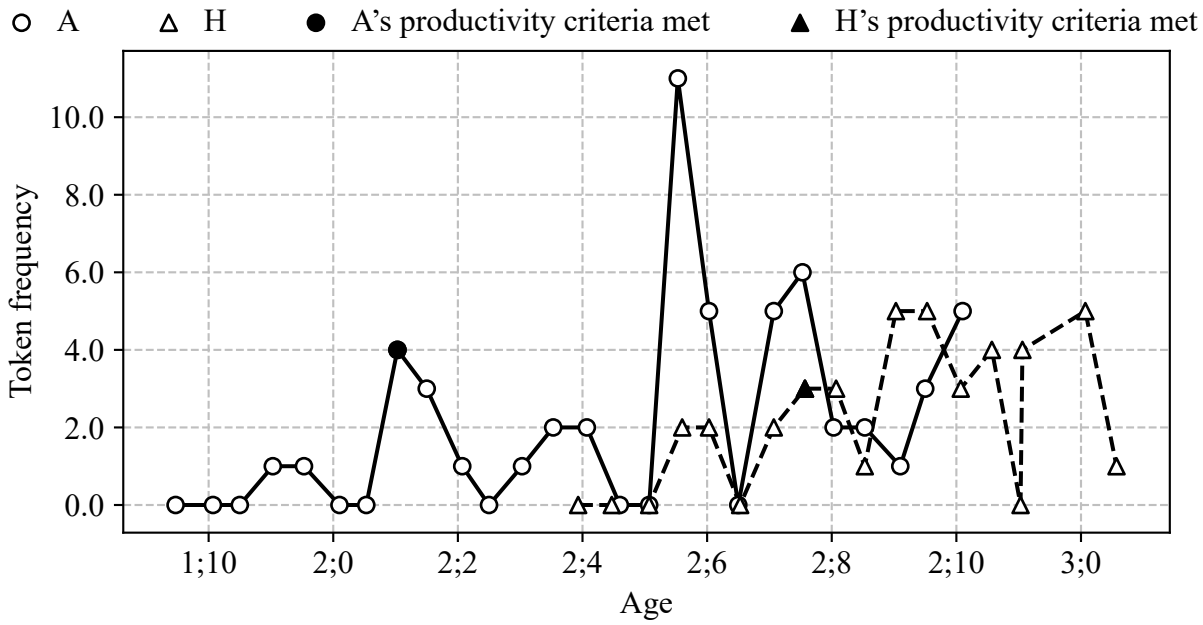


Figure 7.2 Nominal inflections - Accusative *-ne*.

(157)		ʋeere	meecca-ne	kaɳ-ʈ-illa	(H, 2;9.1)
		another	man-ACC	see-PST-NEG	
	[ɲaan]	ʋeere	manuʂya-ne	kaɳ-ʈ-illa	[adult form]
	[I]	another	man-ACC	see-PST-NEG	
	'I didn't see another man.'				

(158)		caanam	**cooɳ-e**	i-cc**ɨ**	koll-i* (H, 2;8.16)
		thing	phone-ACC	hit-PTCP	kill-PST*
	[aa]	ʂaadʱanam	**fooɳin-e**	iʈicc-**ɨ**	ko-ɳɳu [adult form]
	[that]	thing	phone-ACC	hit-PTCP	kill-PST
	'That thing killed the phone.'				

In A's productions, the dative variant *-nɨ* appears in 176 tokens involving nine types and the variant *-kkɨ* in 141 tokens involving 24 types, and in H's productions, *-nɨ* occurs in 35 tokens involving three types and *-kkɨ* in nine tokens involving six types (see Figures 7.3 and 7.4). A has 100%, 100%, and 96% correct occurrences of dative marking in three consecutive transcripts at the ages 2;2.15, 2;3.1, and 2;3.16, respectively, and she employs the dative with seven different nominals in the subsequent transcript at the age of 2;4.2. A's first productive use of the dative variants *-nɨ* and *-kkɨ* is with three unique stems at 2;4.2 and 2;3.1, respectively. No omission errors involving dative marking is attested in H's transcripts and he productively employs both

the variants -*ni* and -*kki* with three different stems at the age of 2;5.18. The dative marker and its acquisition will be explored in great detail in Chapter 8.

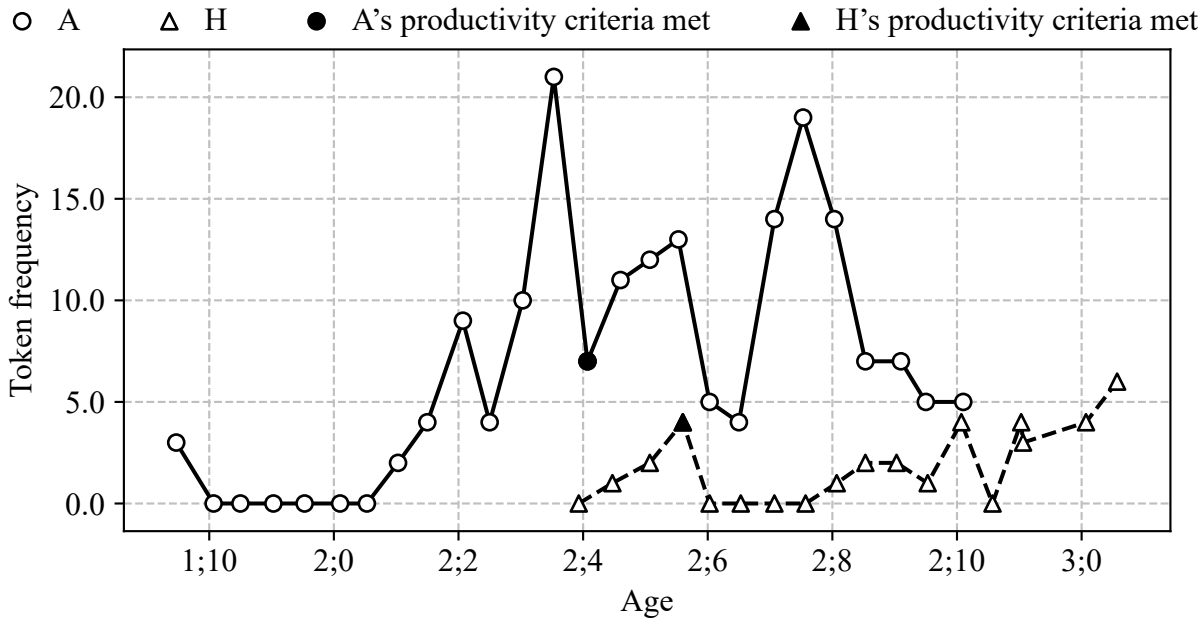


Figure 7.3 Nominal inflections - Dative -*ni*.

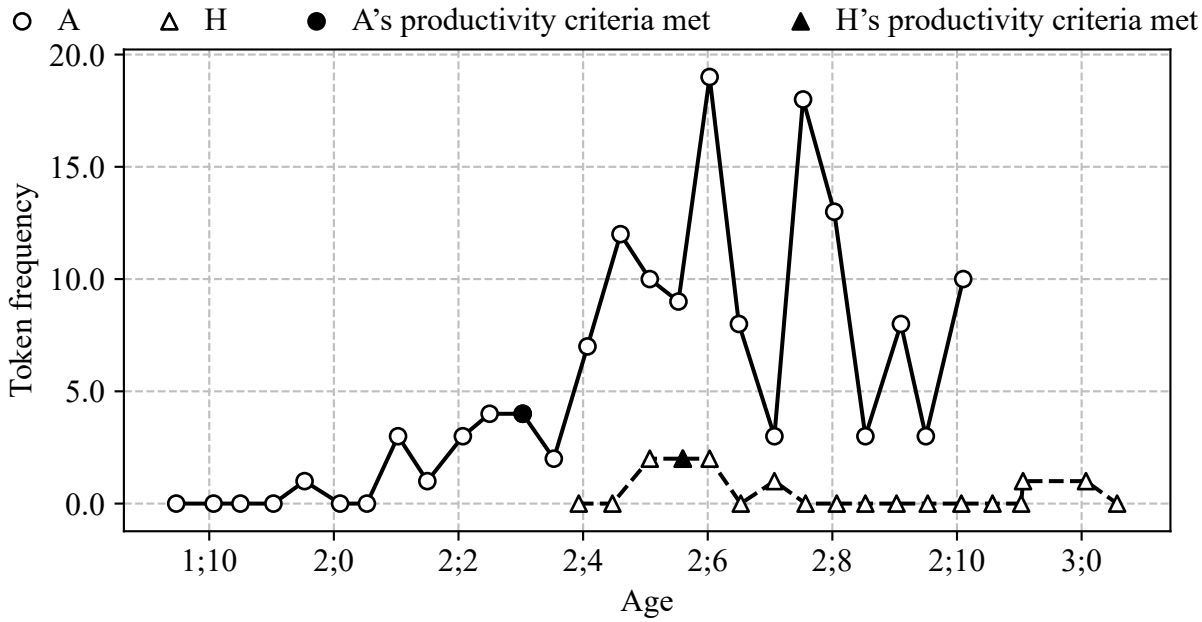


Figure 7.4 Nominal inflections - Dative -*kki*.

(159) iṯi pa-cci ṯaa naa-ni (H, 2;11.1)
this pluck-PTCP give.IMP I-DAT
eni-kki iṯi pari-cci ṯaa [adult form]
I-DAT this pluck-PTCP give.IMP
'Pluck and give it to me.'

(160) ammuu-ni booḷi kaḷiykk-aṇam (A, 2;1.1)
ammu-DAT ball play-want.MOD
'Ammu wants to play ball.'

In A's transcripts, the genitive variant *-nte* appears with 24 types with a token count of 234 instances while the variant *-ṭe* surfaces with 19 types involving a total of 146 tokens. H has 81 instances of the variant *-nte* involving 13 stem types and 48 instances of the variant *-ṭe* involving nine stem types (see Figures 7.5 and 7.6). In spite of having omission errors, A has 100% correct instances in three consecutive transcripts. A meets the productivity criteria for the genitive variants *-nte* and *-ṭe* at 2;1.1 and 2;1.15, respectively, with three different stems. H only omits the genitive once at the age of 2;4.14 and he productively employs genitive variants *-nte* with three different stems at 2;4.14 and *-ṭe* with four different nominals at the age of 2;7.17 thereby meeting the productivity criteria.

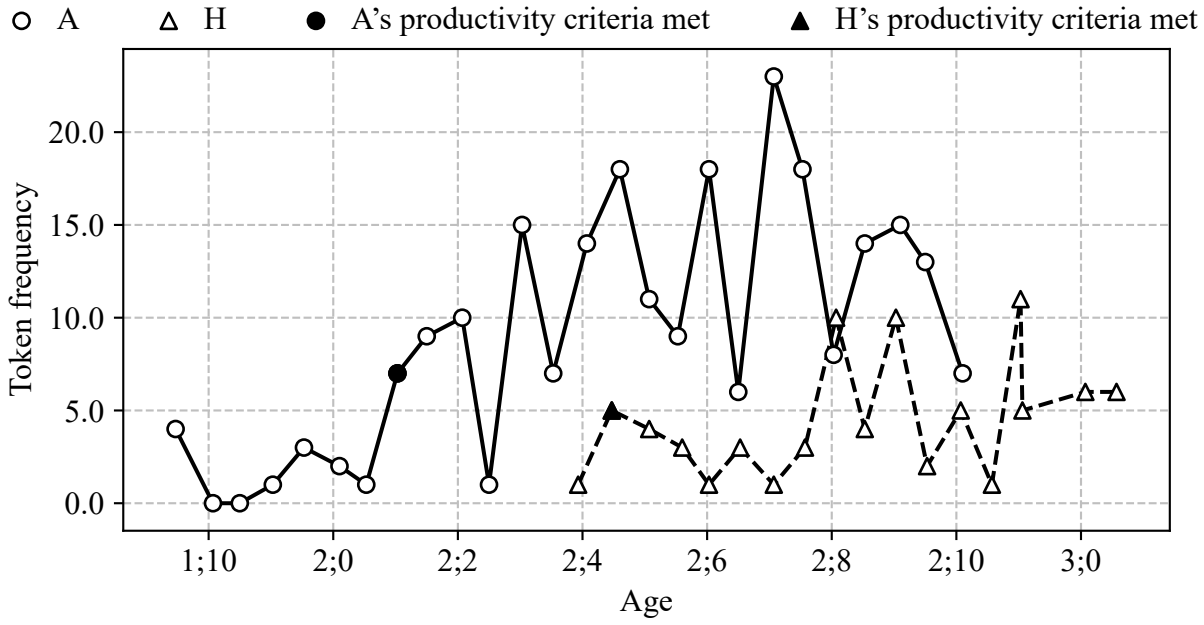


Figure 7.5 Nominal inflections - Genitive *-nte*.

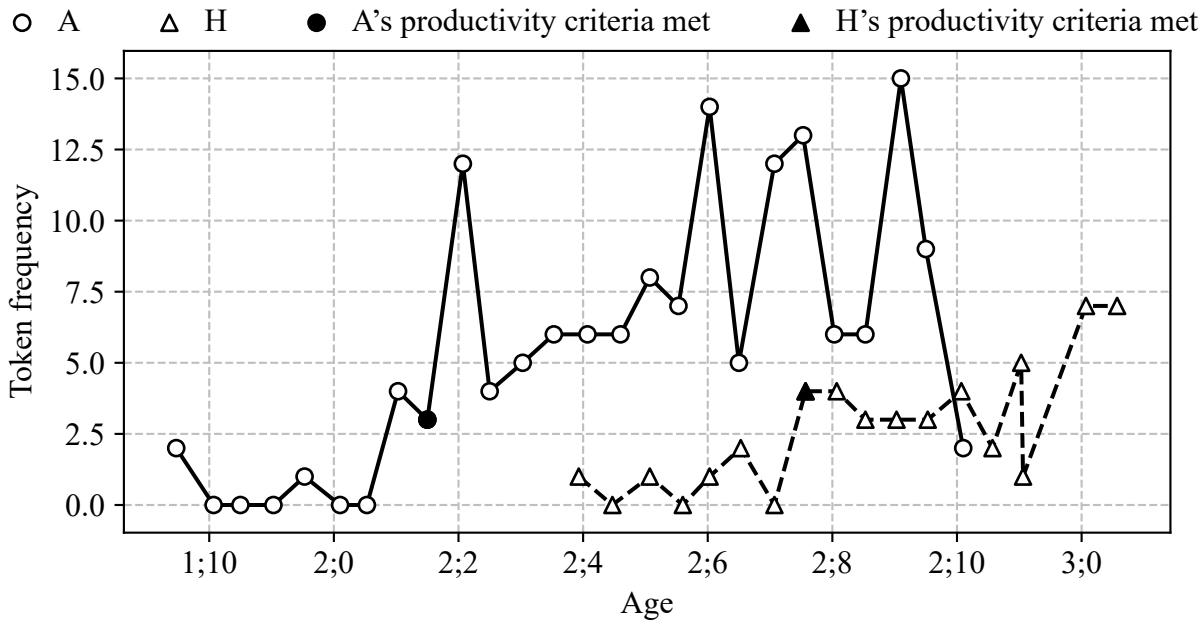


Figure 7.6 Nominal inflections - Genitive *-ʈe*.

(161) aa maam-an-te ʋiiʈʈ-ii pook-uʋ-aa (A, 2;6.1)
that uncle-M-GEN house-LOC go-PTCP-be.PRS
[ɲaan] aa maam-an-te ʋiiʈʈ-il pook-uʋ-aa [adult form]
[I] that uncle-M-GEN house-LOC go-PTCP-be.PRS
'Going to that uncle's house.'

(162) kuɲɲi-piḷḷ-eeɾ-uʈe oɾu aʈɨkriim meeʈiykk-aam (A, 2;7.16)
small-child-PL-GEN one icecream buy-MOD
[ɲamu-kkɨ] kuɲɲi-piḷḷ-eeɾ-uʈe oɾu aiʂkriim meeʈiykk-aam [adult form]
[we-DAT] small-child-PL-GEN one icecream buy-MOD
'Let's buy one of the ice creams for little children.'

(163) naan-te ceppɨ ʈaaɻey-aa (H, 2;8.2)
I-GEN footwear below-be.PRS
en-te ceɾuppɨ ʈaaɻey-aa [adult form]
I-GEN footwear below-be.PRS
'My footwear is downstairs.'

(164)	ammee-ʈe	ʋiiʈʈ-ii	pooy-i		(H, 2;8.16)
	mom-GEN	house-LOC	go-PST		
	[ɲaan]	ammayu-ʈe	ʋiiʈʈ-il	pooy-i	[adult form]
	[I.NOM]	mom-GEN	house-LOC	go-PST	
	'I went to mom's house.'				

The sociative marker occurs with very few verbs in Malayalam. This marker is absent in H's transcripts. In A's transcripts, the sociative case is the last to emerge and is only seen at 2;3.1 as shown in Figure 7.7. It is not frequent in the child's production and is only used with the verb *paray-* 'say' on all occasions. This could be because of the limited distribution of this marker since it only occurs with the *verbs of saying* in Malayalam. Sociative case is assigned to the indirect objects of *verbs of saying* in lieu of the dative marker which is assigned to indirect objects across the board. No omission errors are recorded for this marker in A's productions. The sociative variant *-ooʈɨ* is used with six different nominals in a total of eight instances meeting the productivity criteria at 2;7.2 with four different stems while the variant *-nooʈɨ* only occurs with the proper noun *ammu* just twice across the transcripts. However, this can be attributed to the limited distribution of the marker itself. Since it only occurs with very few verbs, despite being highly frequent, chances of the marker occurring with different stem types are lesser when compared to the other markers.

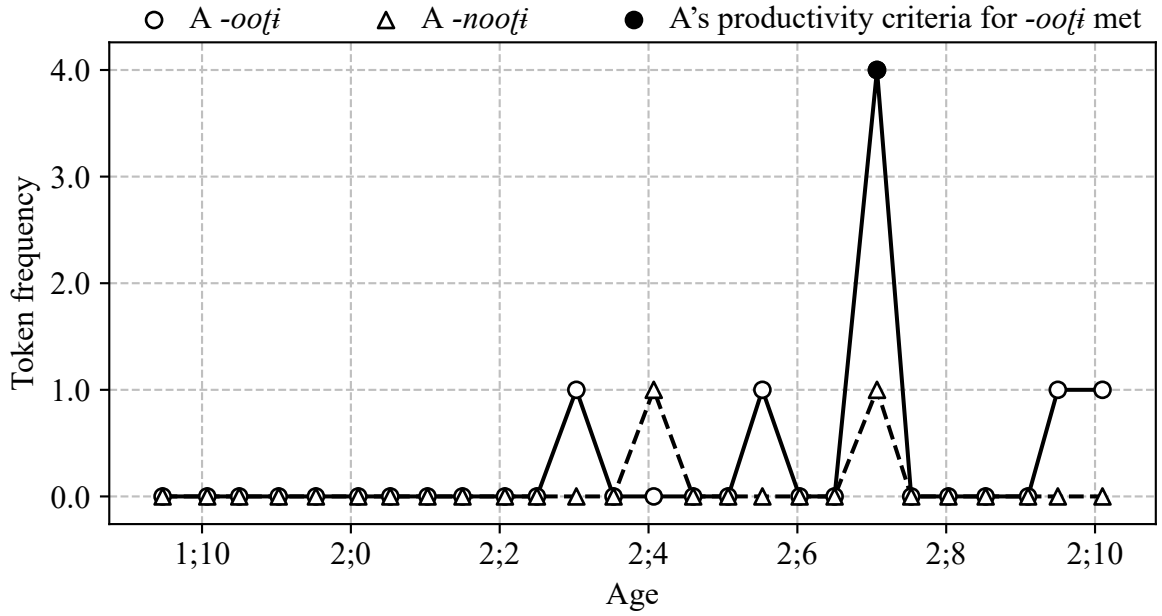


Figure 7.7 Nominal inflections - Sociative *-ooʈɨ* and *-nooʈɨ*.

(165) accay-ooʈɨ fooɳ-ilɨ paray-aɳam ammuu-nɨ (A, 2;3.1)
dad-SOC phone-LOC tell-MOD ammu-DAT
akkriim koɳʈɨʋaɾ-aan
ice cream bring-INF
[ammuu-nɨ] accay-ooʈɨ fooɳ-ilɨ paray-aɳam ammu-nɨ [adult form]
[I.DAT] dad-SOC phone-LOC tell-MOD ammu-DAT
aiʂkriim koɳʈɨʋaɾ-aan
ice cream bring-INF
'Should tell dad to bring ice cream.'

(166) akkɾiim kayikk-aɳʈaa pa-ɲɲu ammuu-nooʈɨ (A, 2;4.2)
icecream eat-MOD.NEG tell-PST ammu-SOC
[ɖookʈar] ammuu-nooʈɨ aiʂkriim kaɻiykk-aɳʈa eɳɳɨ [adult form]
[doctor] ammu-SOC icecream eat-MOD.NEG COMP
para-ɲɲu
say-PST
'The doctor told Ammu not to eat ice cream.'

The children productively use the locative marker *-il* with a number of stems as they progress in age, as can be seen in Figure 7.8. There are 290 instances involving 53 types in A's transcripts for the marker. She omits the marker just twice at 2;2.2 and 2;5.6 and begins to productively employ the marker with three different stems at the age of 1;11.1.

(167) booɭ-eel maɳɳɨ patt-uʋ-oo?
ball-LOC soil get-FUT-Q
'Will the ball get dirty?'

In H's speech, the marker is present in all the transcripts except the very first one. H has 113 tokens of the locative marker *-il* appearing with 26 different nominal stems. There are no omissions and the productivity criteria are met at 2;4.14 with four different stems.

(168) naan-te ʋiiʈʈ-ii oɳʈɨ maɳi (H, 2;10.2)
I-GEN house-LOC have.PRS bell
en-te ʋiiʈʈ-il maɳi oɳʈɨ [adult form]
I-GEN house-LOC bell have.PRS
'There is a bell in my house.'

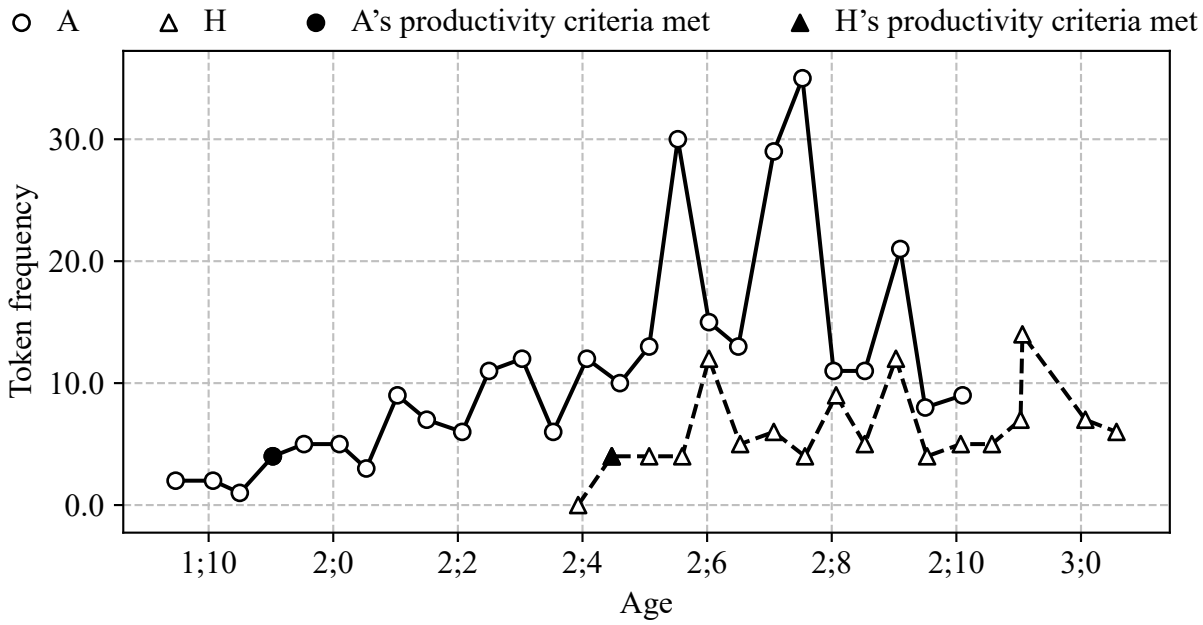


Figure 7.8 Nominal inflections - Locative *-il*.

The use of the locative morpheme *-aṯṯi* is not so frequent in the children's speech (Figure 7.9) although there are no omissions of this marker in the transcripts. There are 14 instances involving seven types and nine instances involving four types in A's and H's productions, respectively. Despite being a marker with restricted distribution, *-aṯṯi* is employed productively with four stems at 2;6.1 and three stems at 2;10.2, in A's and H's transcripts, respectively.

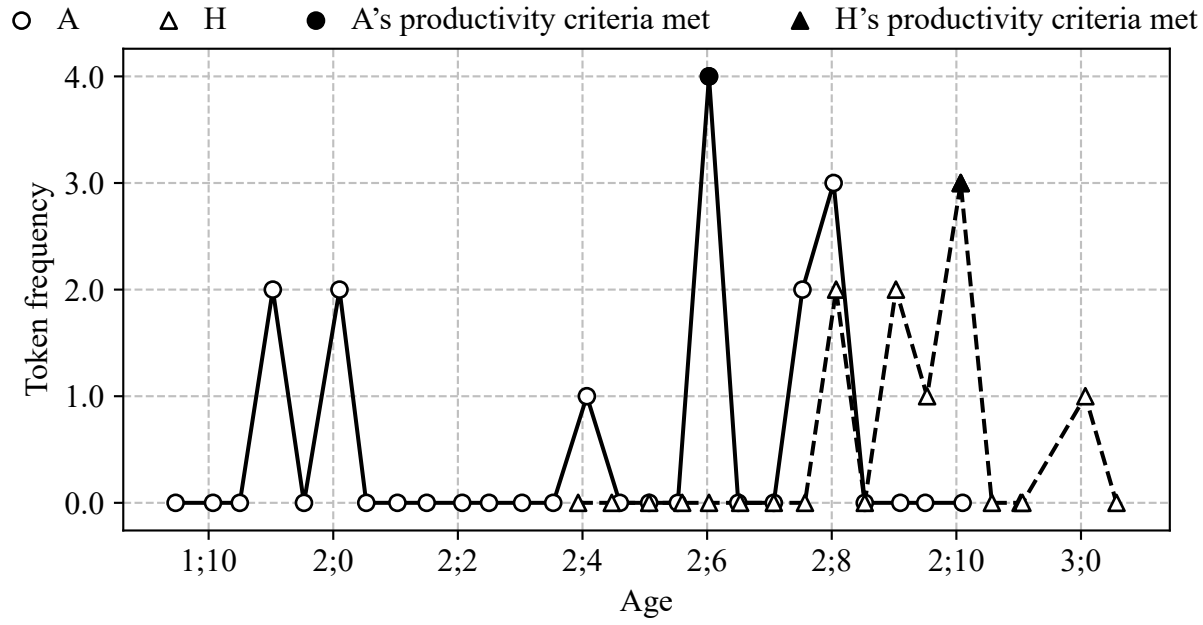


Figure 7.9 Nominal inflections - Locative *-aṯṯi*.

(169) baḍḍeecc-i ʋeyil-aṭṭi ammuu-nte kuuṭe poo-y-ill-ee? (A, 2;6.1)
bhadra-F sunlight-LOC ammu-GEN with go-PST-NEG-Q
'Didn't Bhadra go out with Ammu in the sun?'

(170) iŋŋane caaṭ-i puuccee-ṭe mee-ṭṭi i-ṇṇi (H, 2;10.2)
like this jump-PTCP cat-GEN body-LOC sit-PST
[ɲaan] iŋŋane caaṭ-i puuccayu-ṭe ḍeeh-aṭṭi iɾu-ṇṇu [adult form]
[I.NOM] like this jump-PTCP cat-GEN body-LOC sit-PST
'I jumped and sat on top of the cat.'

It has to be noted that the locative marker *-kal* and the instrumental *-aal* are absent in the children's transcripts. This is not surprising since they are low-frequency morphemes with restricted distribution. The former is employed with very few stems and is synchronically non-productive in Malayalam, while the latter is restricted to formal contexts and thus is not present in the children's or child-directed speech.

7.3.2 Gender

The masculine marker *-an* is present in A's transcripts in 148 instances involving 13 types. The masculine gender marker *-an* surfaces from 1;11.1 onwards. The different stems and the age at which they are first attested are given in Table 7.9.

At 2;6.1 and 2;8.1, the child inflects the largest number of stems with the marker. The marker is never omitted and it meets the productivity criteria at 2;5.2 with three different stems in A's transcripts. The instances that look like omissions are restricted to the three stems *acca* 'dad', *appuuppa* 'grandpa', and *maama* 'uncle'. Those words are used without the overt masculine marker in child-directed speech by her parents and relatives as well and thus do not qualify as omissions in obligatory contexts. The child even uses the marker with the exceptional cases *koɾaŋŋan* 'monkey', *kurukkan* 'fox', and *ṣuuɾyan* 'the sun' which are [–HUMAN] stems that take the masculine marker.

(171) koɾaŋŋ-an keer-uʋ-aa balya maɾaṭṭ-i (A, 2;6.1)
monkey-M climb-PTCP-AUX big tree-LOC
koɾaŋŋ-an ʋalya maɾaṭṭ-il keer-uʋ-aa [adult form]
monkey-M big tree-LOC climb-PTCP-AUX
'The monkey is climbing up the big tree.'

Table 7.9 Masculine word-form types in A's transcripts.

Age	**Noun**	**Gloss**
1;1.11	maaman	'uncle'
2;1.1	ceeʈʈan	'brother'
2;1.1	cittappan	'paternal uncle'
2;2.2	paccakkaʈikkaaɳan[13]	'green grocer'
2;2.15	koŋŋan[14]	'monkey'
2;3.1	accan	'dad'
2;4.18	ammaaʋan	'maternal uncle'
2;5.2	appuppan	'grand father'
2;6.1	cuuɾiyan[15]	'the sun'
2;7.2	kurukkan	'fox'
2;7.2	kaʟʟan	'thief'
2;8.1	manuʂan[16]	'man'
2;8.16	kuuʈʈukaaɾan	'friend'

(172) manuʂy-an ʋa-n̪n̪i en̪t̪i ceyy-um? (A, 2;8.1)
man-M come-PTCP what do-FUT
'What will the man do when he comes?'

(173) kurukk-an-te ɕammaanam-aa (A, 2;9.3)
fox-M-GEN gift-be.PRS
[aʈ̪i] kurukk-an-te ʂammaanam-aa [adult form]
[that] fox-M-GEN gift-be.PRS
'It is the fox's gift.'

In H's transcripts, the marker appears with 11 types and 50 tokens and appears in all the transcripts except one at 2;5.18. The different stem types and the age at which they are first attested in the child's transcripts are given in Table 7.10. H assigns the marker to three distinct stems, meeting the productivity criteria at 2;9.1. Figure 7.10 shows the distribution of the masculine marker, *-an*, in the children's transcripts. There are no omission errors attested

[13] Adult form *paccakkarikkaaran*
[14] Adult form *koraŋŋan*
[15] Adult form *ʂuuɾyan*
[16] Adult form *manuʂyan*. She sometimes uses the adult form correctly as shown in (172).

for the masculine marker in both the children's transcripts.

Table 7.10 Masculine word-form types in H's transcripts.

Age	Noun	Gloss
2;3.8	maaman	'uncle'
2;3.8	oo̲ṯṯookkaaran	'auto rickshaw driver'
2;4.14	accaaccan	'paternal grandfather'
2;4.14	ceeṯṯan	'brother'
2;6.1	appuppan	'grand father'
2;6.16	accan	'dad'
2;7.2	koccaccan	'paternal uncle'
2;8.16	ḏeeṯṯan[17]	proper name (of the subject's brother)
2;9.1	meeccan[18]	'man'
2;9.16	kaḷḷan	'thief'
2;10.2	miinkaaran	'fisher man'

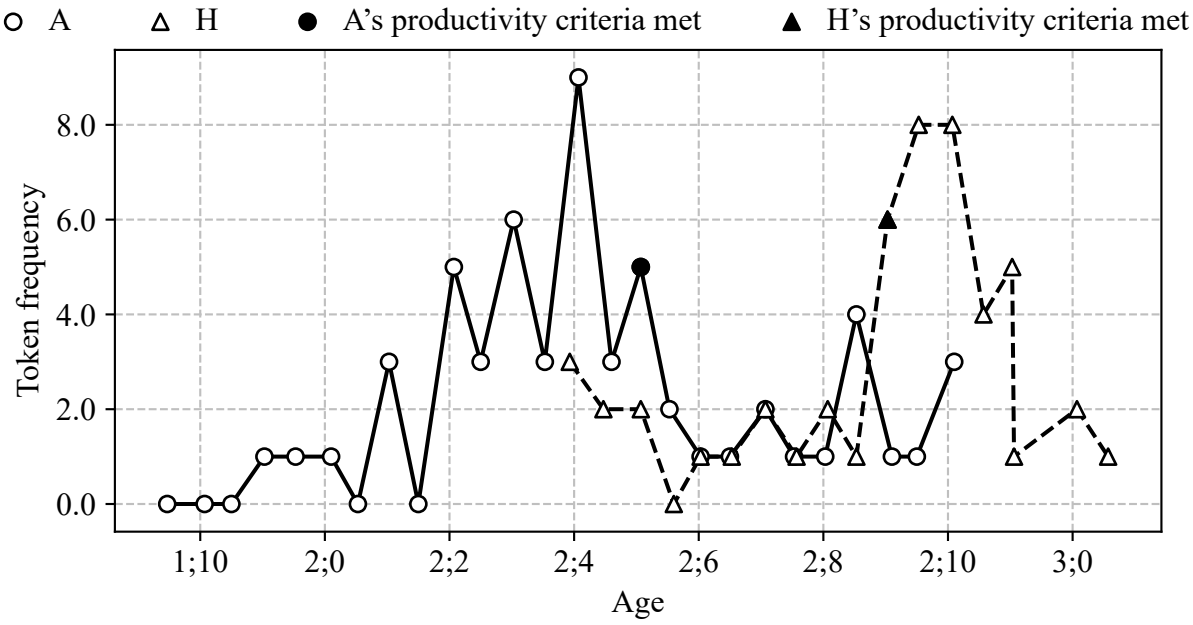


Figure 7.10 Nominal inflections - Masculine *-an*.

[17] Adult form *ḏeevuuṯṯan*
[18] Adult form *manuṣyan*

(174) appuupp-an maaŋ-c-ay-aa aṯṯ (H, 2;1.1)
grandpa-M buy-PST-NMLZ-be.PRS that
aṯṯ appuppan ʋaaŋŋi-cc-aṯ-aa [adult form]
that grandpa-M buy-PST-NMLZ-be.PRS
'Grandpa bought that.'

(175) ṯeŋŋ aa kaḷḷ-an i-cc ka-ɲɲ-ay-aa (H, 3;0.2)
coconut palm that thief-M hit-PTCP lose-PST-NMLZ-be.PRS
ṯeŋŋ aa kaḷḷ-an iṯi-cc kaḷa-ɲɲ-aṯ-aa [adult form]
coconut palm that thief-M hit-PTCP lose-PST-NMLZ-be.PRS
'That thief destroyed the coconut palm.'

In A's transcripts, the feminine marker *-i* occurs once at 1;10.2 and is in regular use from the age of 2;1.1. Nonetheless, it is mostly used with compound stems ending in *ceecc-i*, hence the tokens lack in variety and productivity. There are no omissions but the marker does not meet the productivity criterion.

(176) ceecc-i ʋa-ṉṉu (A, 2;6.15)
elder sister-F come-PST
'Elder sister came.'

(177) baḍṛ-eecc-i oṯaŋŋ-iy-iṯṯi eɳii-tt-iṯṯi ʋaɾa-ṯṯe (A, 2;8.1)
bhadra-sister-F sleep-PTCP-PFV wake up-PTCP-PFV come-OPT
bʱaḍṛa-ceecc-i oraŋŋ-iy-iṯṯi eɳii-tt-iṯṯi ʋaɾa-ṯṯe [adult form]
bhadra-sister-F sleep-PTCP-PFV wake up-PTCP-PFV come-OPT
'Let Bhadra come after she finishes her sleep.'

In H's speech, the feminine marker *-i* appears only in two instances with just one type, *ceecci*, which is the only type present in A's transcripts as well and thus is not productively used. The details of the distribution of the feminine marker are shown in Figure 7.11. The feminine marker *-i* is not omitted but the feminine marker *-aṯṯi* is absent in both children's transcripts.

7.3.3 Number

The details of the production of the different Malayalam plural markers are given in Figures 7.12 to 7.14. In A's transcripts, the first appearance of the [+HUMAN] plural marker *-maar* is at 2;1.1. It appears with four stem types, as shown in Table 7.11 where each type is listed along with the

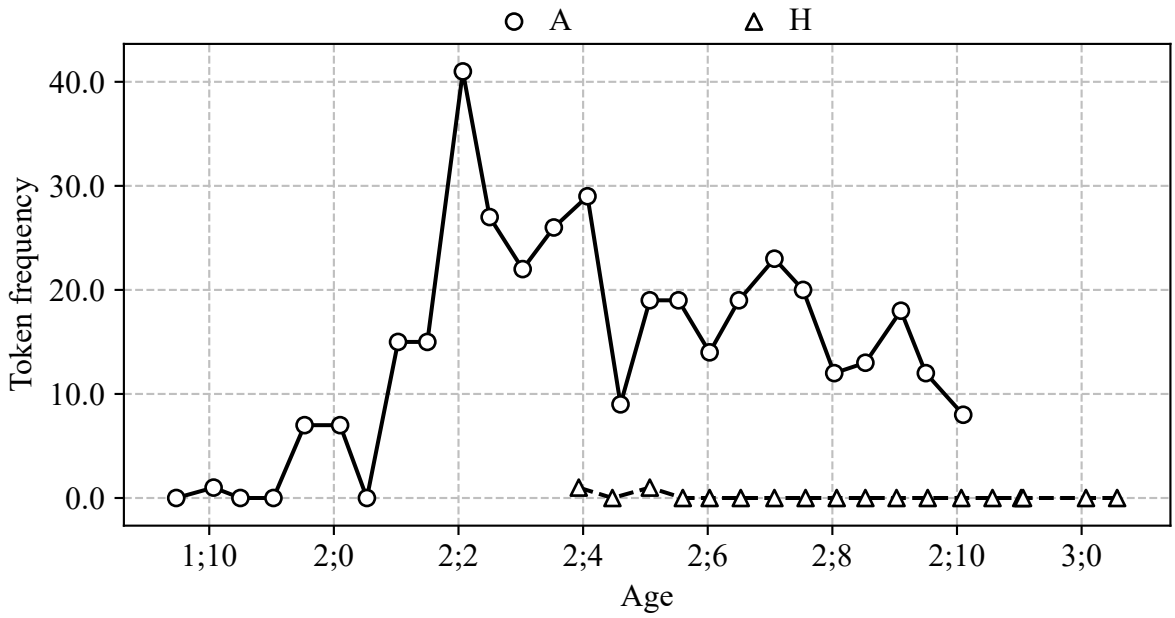


Figure 7.11 Nominal inflections - Feminine *-i*.

age at which it is attested, with a total of 19 instances. The marker meets the productivity criteria at 2;4.2 by inflecting three different stems.

At 2;2.2, the epicene humanising plural marker *-kaar* surfaces and occurs intermittently at later stages between 2;6.1 and 2;10.3. The marker *-kaar* comprises 20 tokens involving four distinct types with only two unique stem types in any of the transcripts. The types along with their age of emergence in the transcripts for the marker *-kaar* is given in Table 7.12. The marker meets the productivity criteria at 2;6.1 by surfacing with three distinct nominals.

Table 7.11 *-maar* plural forms in A's transcripts.

Age	Noun	Gloss
2;1.1	ceeṯṯanmaaṙ	'brothers'
2;2.2	piḷḷaaṙ[19]	'children'
2;4.2	ceeccimaaṙ	'sisters'
2;7.2	kaḷḷammaaṙ	'thieves'

[19]The underlying form of *piḷḷ-aar* is *piḷḷa-maar* [child-PL]. These kinds of elisions happen in Malayalam plural marking. Another example is the epicene form *ṉaaṯṯ-aar*, the underlying form of which is *ṉaṯṯi-kaar* [land-HUM.PL].

Table 7.12 *-kaar* plural forms in A's transcripts.

Age	Noun	Gloss
2;2.2	paccakkaʈikkaaɳan[20]	'green grocer'
2;6.1	aaʟkkaɾɨ	'people'
2;6.1	kuuʈʈukaaɾɨ	'friends'
2;7.16	ʋiiʈʈɨkaaɾɨ	'family'

(178) iṯɨ ceeʈʈ-an-um ceecc-i-maaɾ-um (A, 2;5.16)
this brother-M-CORD sister-F-PL-CORD
'This is elder brother and sisters.' [adult form]

(179) ii ʋiiʈʈu-kaaɾɨ aar-kk-aa koʈukk-uṉṉ-aṯɨ? (A, 2;7.16)
these house-PL who-DAT-be.PRS give-PRS-NMLZ
ii ʋiiʈʈu-kaaɾɨ aar-kk-aa [paiṩa] koʈukk-uṉṉ-aṯɨ? [adult form]
these house-PL who-DAT-be.PRS [money] give-PRS-NMLZ
'Who is this family giving (money) to?'

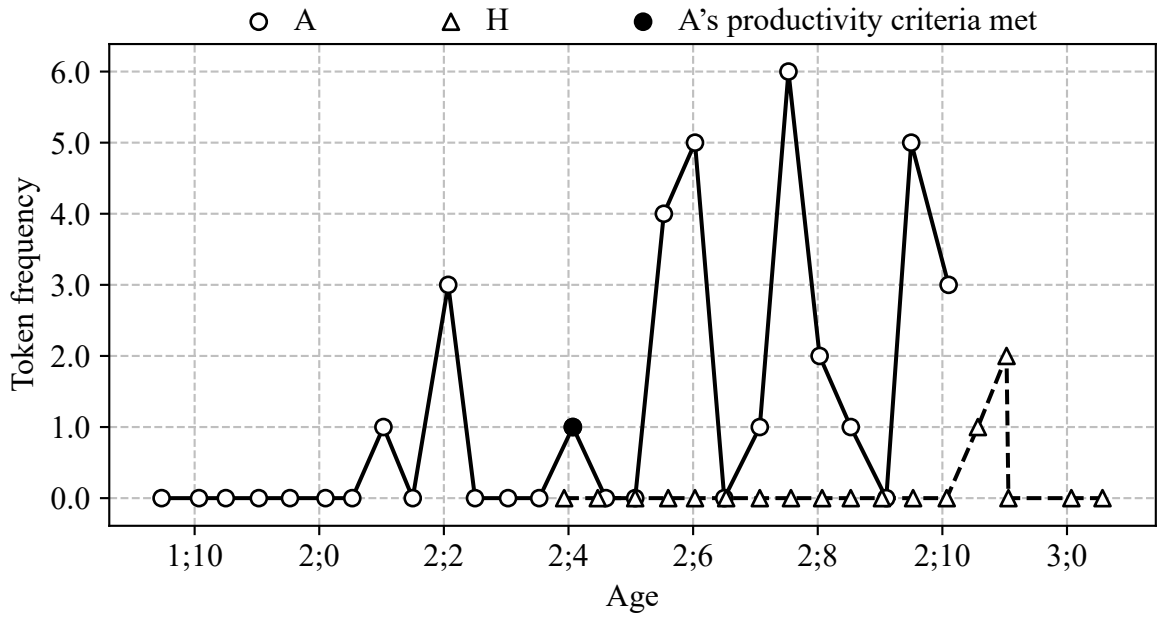


Figure 7.12 Nominal inflections - Plural *-maar*.

The [–HUMAN] marker *-kaʟ* is the last one to emerge with six tokens involving six types. It materialises only at 2;7.16 and is seen in successive recordings, but it never occurs with more

[20] Adult form *paccakkarikkaaran*

than one stem in any of the sessions. The marker meets the productivity criteria at 2;8.16. The late emergence might be due to its optionality. The types with which the marker surfaces and the corresponding ages are given in Table 7.13.

Table 7.13 *-kaḷ* plural forms in A's transcripts.

Age	Noun	Gloss
2;7.16	kaḷippaaṭṭaŋŋaḷ	'toys'
2;8.1	*baccaŋŋaḷ[21]	'foods'
2;8.1	aaṭɨkaḷ	'goats'
2;8.16	ṯavaḷakaḷ	'frogs'
2;9.3	pulikaḷ	'leopards'
2;10.3	ṯaṯṯakaḷ	'parrots'

(180) iṯɨ ṯaṯṯa-kaḷ-aa (A, 2;10.3)
this parrot-PL-be.PRS
'These are parrots.'

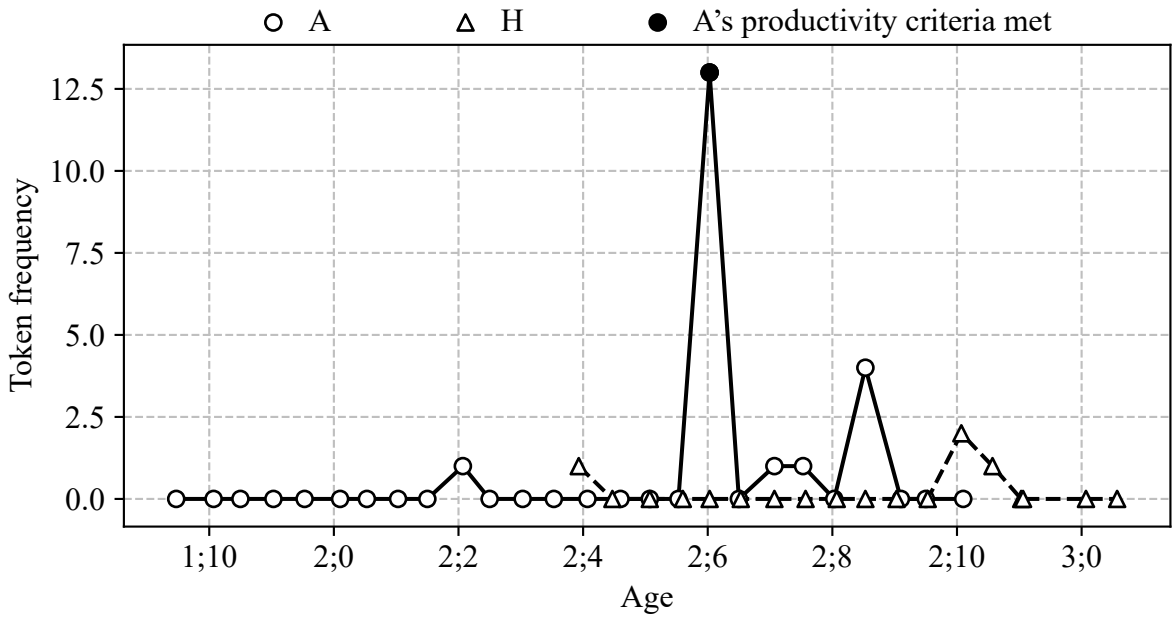


Figure 7.13 Nominal inflections - Plural *-kaar*.

[21] Adult form *bʰakṣaŋam*. This is an instance of overextension. The word *bʰakṣaŋam* 'food' is a mass noun and is not pluralised. However, this shows that the child correctly understood the [-animate] feature of the stem and used the appropriate marker, although failing to take into account the fact that it is a mass noun.

In all of H's transcripts, the plural marker *-maar* occurs once with the stem *ceeʈʈan* 'brother' at 2;11.1 and is incorrectly used with the numeral adjective *oru* 'one'. H also has two instances of *piḷḷaaṛ* 'children' in his speech at 2;10.17 and 2;11.1. Further, the child omits the marker four times[22].

(181) oru* ceeʈʈ-an-maaṛ-aa (H, 2;11.1)
one brother-M-PL-be.PRS
ceeʈʈ-an-maaṛ-aa [adult form]
brother-M-PL-be.PRS
'Those are elder brothers.'

The epicene marker *-kaar* is seen at the age 2;9.16. The former has four tokens involving two types as shown in the examples (182) and (183). None of the epicene markers is used with more than one unique stem in H's speech.

(182) miin-kaaṛ-an aaɳ-aa? (H, 2;10.2)
fish-PL-M-be.PRS-Q
'Is that the fisherman?' [adult form]

(183) ootti-kaaṛ-an pooy-i (H, 2;3.28)
auto-HUM-M go-PST
oottoo-kkaaṛ-an pooy-i [adult form]
auto-HUM-M go-PST
'The auto rickshaw driver left.'

The [−HUMAN] marker *-kaḷ* emerges towards the end of the recordings at 2;11.18. The marker has five tokens, all of them involving a single type. None of the plural markers meets the productivity criteria in H's transcripts since they do not occur with more than two distinct stems in any of the recordings or across transcripts.

(184) en-te ʋiiʈʈ-ii puucca-kuɲɲu-ŋŋaḷ oɳʈ-alloo (H, 2;11.18)
I-GEN house-LOC cat-child-PL have.PRS-EMPH
'There are kittens at my home.' [adult form]

It has to be noted that children correctly follows the template for affix concatenation in nominals. Also, the epicene plural marker *-ar* occurring with Sanskrit bound stems is not attested in the children's transcripts. This is not surprising since the *-ar*-inflected nominals only occur in formal contexts.

[22] We return to this in the discussion on errors.

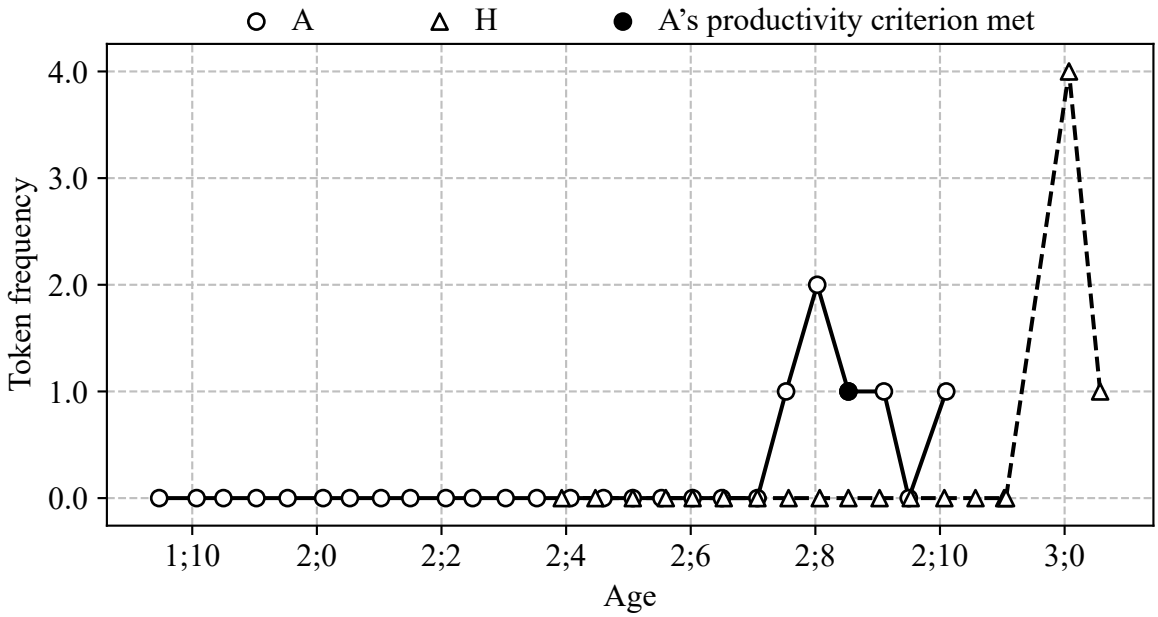


Figure 7.14 Nominal inflections - Plural *-kal*.

7.4 Nominal Inflection: Errors in the Data

In this section, we discuss the production errors involving various nominal inflectional suffixes. We see that all the case markers except the sociative, which occurs very rarely in the transcripts, are omitted in obligatory contexts at least once in A's speech. H is also seen to omit different case markers. Table 7.14 shows the number of correct instances and omissions for the different case markers that have been omitted in obligatory contexts by the children.

Table 7.14 Omissions vs correct uses in the data.

	A's Transcripts		**H's Transcripts**	
Marker	**Correct**	**Omission**	**Correct**	**Omission**
Accusative	127 (86%)	21 (15%)	54 (89%)	7 (11%)
Dative	318 (95%)	16 (5%)	43 (100%)	-
Genitive	380 (98%)	7 (2%)	129 (99%)	1 (1%)
Locative	290 (99%)	2 (1%)	113 (100%)	-
Plural *-maar*	32 (97%)	1 (5%)	3 (50 %)	3 (50%)

Accusative is the most frequently omitted case marker in both the children's speech. The intermittent omission of the marker is visible throughout the data. There are 127 correct and 21 omission instances in A's productions whereas in H, there are seven instances of omission errors

and 54 correct instances.

(185) ***kiɭi_** kaaɳ-cc-~~i~~ ʈay-aam (A, 2;1.1)
bird-*0ACC see-TR-PTCP give-MOD
kiɭiy-e kaaɳ-icc-~~i~~ ʈaɾ-aam [expected form]
bird-ACC see-TR-PTCP give-MOD
'Will show (you) the bird.'

(186) ***elii_** ɳookk-uɳɳu (H, 2;7.17)
mouse-*0ACC look-PRS
[puucca] **eliy-e** ɳookk-uɳɳu [expected form]
[cat] mouse-ACC look-PRS
'The cat is looking at the mouse.'

In addition to the omission errors, there are a few instances where the accusative is assigned in lieu of the nominative, once at 2;0.16, and thrice at 2;5.16 in A's speech.

(187) baag-il~~i~~ ***kokkin-e** oɳʈ-oo? (A, 2;5.16)
bag-LOC crane-*ACC have-Q
baag-il **kokk~~i~~** oɳʈ-oo? [expected form]
bag-LOC crane.NOM have-Q
'Is there a crane inside the bag?'

There is also one instance of overgeneralising the accusative marker to [−ANIMATE] objects each in A's and H's transcripts.

(188) ***booɭin-e** mee-cc~~i~~ ʈaɾ-aɳam (A, 2;5.16)
ball-*ACC buy-PTCP give-MOD
booɭ~~i~~ meeʈi-cc~~i~~ ʈaɾ-aɳam [expected form]
ball-øACC buy-PTCP give-MOD
'(You) should buy (me) the ball.'

(189) ***ʈei-ne** oɳʈaakk-i ʈaa (H, 2;11.18)
train-*ACC make-PTCP give.IMP
ʈrein oɳʈaakk-i ʈaa [expected form]
train-øACC make-PTCP give.IMP
'Make me the train.'

To summarise, we see omission errors 21 times, incorrect assignment four times and one instance of overgeneralisation in A's transcripts, in comparison with the 127 correct instances, while there are seven instances of omissions and one instance of overgeneralisation in H's productions compared to the 54 correct instances.

Omission of the accusative marker is also attested in the cross-sectional data in older children who are past four years of age. Nine out of the sixteen children who were recorded are seen to omit the marker in their speech, as shown in Table 7.15. The presence of omission errors signal that the acquisition of the marker is not yet complete even in children as old as four years and eight months of age. Recall that the accusative is only marked on a subset of object nouns and a bulk of them will remain unmarked, grammatically so.

Table 7.15 Accusative omission in cross-sectional data.

Participants (Age)	**Omissions**	**Correct Instances**
AV (4;6)	1 (9%)	10 (91%)
AD (4;7)	3 (33%)	6 (67%)
NP (4;5)	3 (10%)	19 (90%)
AB (4;8)	1 (6%)	17 (94%)
VS (4;7)	1 (6%)	15 (94%)
AA (4;6)	1 (5%)	20 (95%)
VV (4;8)	1 (4 %)	23 (96%)
ND (4;4)	1 (5%)	18 (95%)
AS (4;8)	2 (9%)	20 (91%)

Similarly, omission in obligatory contexts (190) and overextension of the marker to [−ANIMATE] objects (191) are seen in children acquiring Malayalam and English simultaneously (Raghunathan, Unpublished data).

(190) eniy-kkɨ ***cɑ:lɪ_** kaaɳ-aɳam (A, 2;10.22)
I-DAT charlie*0ACC see-MOD
eniy-kkɨ **caarliy-e** kaaɳ-aɳam [expected form]
I-DAT charlie-ACC see-MOD
'I want to see Charlie.'

(191) ***bukki-ne** ʋa:yikkɨ (A, 2;1.25)
book-*ACC read.IMP
bukkɨ ʋa:yikkɨ [expected form]
book read.IMP
'Read the book.'

We may recall that overt accusative case marking is determined by the lexical feature [±ANIMATE]. We have already demonstrated in Chapters 5 and 6 that the children do not show difficulty in acquiring inflections that are phonologically determined. Specifically, with past tense marking we saw that the children acquired easily and relatively early the allomorphs for underived stems in particular. In fact, children show a distinct preference for such inflections over the allomorphs whose selection is dependent on lexical features such as [±TRANSITIVE].

In order to assess whether the *mode of selection* has an effect on the acquisition process, a Fisher's exact test was run separately on each of the children's speech samples by analysing the object nouns that are [+ANIMATE] and [−ANIMATE]. Tables 7.16 and 7.17 provide the details of the errors in A's and H's speech, respectively. We can see that for [+ANIMATE] nouns where accusative marking is obligatory, A and H have produced 25 and seven incorrect forms and 127 and 54 correct ones, respectively. This is not the case for [−ANIMATE] nouns, where the accusative marking is obligatorily left unmarked. In other words, the nouns do not require overt marking and can be left bare as a default. Consequently, the error rate also is very low and the difference is statistically significant.

Table 7.16 A's accusative productions.

	+ANIMATE	−ANIMATE
Incorrect	25	1
Correct	127	197

Table 7.17 H's accusative productions.

	+ANIMATE	−ANIMATE
Incorrect	7	1
Correct	54	129

Fisher's exact test for this data gives p-values of 4.3×10^{-9} and 1.6×10^{-3} for A and H, respectively, which are statistically significant. The pattern of errors suggests that the delay in acquisition could be because of the complexity of the rule which involves lexical features as part of the selectional restriction.

The genitive marker is omitted once in H's speech and is assigned correctly 129 times while there are seven instances of omission in A's speech compared to the 380 correct instances.

(192) mom: aaɾ-uʈe ʈalay-aa?
who-GEN head-be.PRS
'Whose head is it?'
H: ***accu*** (H, 2;4.14)
achu.*0GEN
accu-nte [expected form]
'Achu's.'

(193) ***ammu*** paaʊ-e appi-aay-i (A, 1;11.16)
ammu.*0GEN doll-LOC dirt-be-PST
ammuun-te paaʊay-il appi-aay-i [expected form]
ammu-GEN doll-LOC dirt-be-PST
'Ammu's doll has become dirty.'

(194) ***kuɲɲi-ʋaaʋa*** peeɾɨ eɳʈ-aa-ɳɳɨ cooyiykk-aam (A, 2;4.18)
small-baby-*0GEN name what-be-COMP ask-MOD
kuɲɲi-ʋaaʋay-uʈe peeɾɨ eɳʈ-aa-ɳɳɨ cooɖiykk-aam [expected form]
small-baby-GEN name what-be-COMP ask-MOD
'Let's ask what the baby's name is.'

(195) ***baɖɖu*** ʋiiʈʈ-ii poo-ɳam (A, 1;11.1)
bhadra-*0GEN house-LOC go-MOD
bʱaɖɾay-uʈe ʋiiʈʈ-il pook-aɳam [expected form]
bhadra-GEN house-LOC go-MOD
'(I) want to go to Bhadra's house.'

In A's transcripts, the locative *–il* is omitted only once at 2;2.2 and never once in H's.

(196) *ammuun-ɨ **ooʈʈoo** pooy-i (A, 2;2.2)
ammu-*DAT auto-*0LOC go-PST
ammu **ooʈʈooy-il** pooy-i [expected form]
ammu.NOM auto-LOC go-PST
'Ammu went in an auto.'

Table 7.14 shows that locative and genitive are the most correctly marked case inflections in the data. A uses the genitive accurately 98% of the time and the locative, 99% of the time, H has 99% correct instances for the genitive and 100% for the locative. Both affixes apply

across the board exceptionlessly and we should expect greater success with these affixes. The few instances of incorrect genitive and locative marking probably arise from an overlap with the semantic environments of dative marking which we discuss in the next chapter.

With respect to number, in both the children's data, the only plural marker that is incorrectly omitted is *-maar*. While A has only one recorded instance of omission of this marker, in H's transcripts three instances of omissions are attested.

(197) G: iṯ-okke aaɾ-aa?
this-all who-be
'Who are all these?'
A: ***ceeʈʈ-an_** (2;1.1)
elder brother-M.*0PL
ceeʈʈ-an-maar [expected form]
elder brother-M-PL
'Elder brothers.'

(198) muuɳɳi ***ceeʈʈ-an_** (H, 2;4.14)
three brother-M.*PL
muuɳɳi **ceeʈʈ-an-maar** [expected form]
three brother-M-PL
'Three brothers.'

(199) ɳaali ***meecc-an_**-e ʋekk-aɳam (H, 2;10.2)
four man-M.*0PL-ACC put-MOD
ɳaali **manuʂy-an-maaɾ**-e ʋekk-aɳam [expected form]
four man-M-PL-ACC put-MOD
'We should put four men.'

We have discussed earlier in Section 7.1. that the features [±HUMAN] and [±ANIMATE] of the nouns decide the degree of obligatoriness or optionality in marking plurality. The marker *-maar* is assigned depending on whether the nominal bears the feature [±HUMAN]. Thus, the plural marker selection is also not a phonologically driven one. This could be one of the reasons for the delay in the acquisition of plural markers as well. Although we do not have enough data to statistically test the validity of this assumption for plural markers, we already had statistically significant results for the accusative production suggesting that children have issues in overtly

marking the accusative, possibly because of the delay in the acquisition of the [+ANIMATE] assignment rule.

We see that both the accusative and the plural markers, whose assignments are determined by the lexical features [±ANIMATE], and both [±HUMAN] and [±ANIMATE], respectively, are the most error-prone compared to the rest of the nominal inflections. Both these inflections are mostly productive - case marking being completely and plural marking being largely exceptionless. However, the formal complexity of plural and case marking on the nominal as determined by one (accusative) or two (plural) inherent lexical-semantic features appears to cause the delay - most for plurals and somewhat less so for the accusative.

Acquisition of German plural formation also shows a similar pattern. Plural formation in German is a morphologically complex phenomenon. German has six suffixes to indicate plurality: *-n*, *-en*, *-e*, *-s*, *-er*, and ø (zero or no affix), among which ø, *-e*, and *-er* can also be combined with umlaut (fronting of the stem vowel), resulting in nine different ways to signal plurality. Further, the realisation of most of these suffixes depends on the gender feature of the noun stem. For example, 70% of the masculine and neuter nouns select *-e* as their regular plural suffix and 96% of feminine nouns choose *-n* or *-en*, while the assignment of the marker *-er* is restricted to neuter nouns and the marker occurs only with a lexicalised class of monosyllabic nouns. Finally, pure umlaut in German is synchronically non-productive[23] and thus its occurrence is lexicalised forcing the children to separately learn these items (Kauschke, Kurth, & Domahs, 2011).

Similar to the syllabic constraint that drives the past tense marking in Malayalam underived verbs (see Section 6.2), German plural formation is prosodically driven, that is, there is a prosodic constraint that requires a plural word-form to end in an unstressed syllable with either *-e* (schwa) or a syllabic consonant, resulting in bisyllabic structures, according to Wiese (1996, as cited in Kauschke et al., 2011). This makes the German plural system an ideal case to investigate the prosodic influences on the acquisition of morphological systems. Thus, the acquisition of German plural formation is an ideal case to check our *mode of selection* proposal. Since children are shown to have mastery over the phonotactic and prosodic constraints of their target language, we expect the children to have greater success in the prosodically driven plural formation compared to the ones that require the learning of lexical features.

Kauschke et al. (2011) have performed a plural elicitation task which comprised words

[23] This is similar to the synchronic non-productivity of past tense formation in Malayalam derived verbs.

covering all German plural forms, with 60 typically developing German-speaking children between three and six years of age. Children are shown to achieve complete mastery over the plural formation patterns relatively late by 5-6 years of age given the complexity of the plural marking system. Kauschke et al. report that children's plural productions rarely violate the prosodic constraint. In fact, children's productions are shown to meet the prosodic constraint for German noun plurals when the subjects produced trochees with an unstressed final syllable by choosing the marker *-e* (schwa), which according to the authors is the most unmarked way of plural formation. Children have also shown particular difficulty in mastering the pure umlaut forms. This is not surprising because of the synchronic non-productivity of the plural umlaut rule. What is most interesting is that children's productions demonstrate sensitivity to the prosodic constraint. Their substitution errors are prosodically driven which reveal that prosodic shape of the word stem is a decisive factor in the selection of the plural allomorph. For example, *-s* was mainly substituted to trochaic stems instead of zero-plurals (e.g. **vaters* instead of *väter* 'fathers'), thus preserving the trochaic structure which ends in an unstressed syllable, whereas *-e* (schwa) was substituted for *-en* in monosyllabic noun stems (e.g. **bäre* instead of *bären* 'bears'), thus making it a trochaic structure.

Thus, the acquisition of German plural formation shows that prosodically driven inflectional assignment is relatively easier for the children to acquire compared to the other kind of assignments that require mastery of referential and lexical features. This is congruent to the acquisition patterns observed in Malayalam in both verbal inflections (past tense formation) and nominal inflections (accusative and plural marking). This shows that *mode of selection* is not limited to Malayalam inflectional acquisition, but it has a role to play in the acquisition of inflections in general.

7.5 Summary

We find that among the nominal inflections, case markers are in place early and are interspersed with omission and commission errors. Among the case markers, the accusative case that is predicated on the lexical feature of [±HUMAN] is the most difficult while the more uniformly assigned locative and genitive are successfully and completely learned. Compared to the case markers, gender, and number inflections emerge quite late and are not as frequent. The total number of unique stems that are marked for plurality is severely restricted to 14 and seven types

in A's and H's speech respectively. Similarly, the gender marked unique stems do not exceed 14 for A and 12 for H. Gender marking and number marking rely on lexical features of the noun and in combination pose a challenge to the children. Across all inflectional material, we find that *mode of selection* turns out to be a complicating factor in acquisition. In the next chapter, we focus on the dative case and in particular, the use of dative subjects.

Chapter 8

Dative Subject Marking in Malayalam

8.1 Introduction

Non-nominative subjects are a cross-linguistic phenomenon. Most South-Asian languages as well as languages like German, Icelandic, Finnish, Faroese etc. (Verma & Mohanan, 1990; Butt, 2006) feature dative (and sometimes other cases as well) nominals in what is the *subject* position of a sentence, structurally defined. The status of these non-canonically case-marked subjects has been a topic of discussion for over three decades (Belletti & Rizzi, 1988; Landau, 2002; Butt, 2006). The verbs that permit such subjects are variously known as *experiencer* or *quirky-case* marking or *psych* predicates.

Belletti and Rizzi (1988) argue that *psych-verbs* like *precoccupare* 'to worry' (200) and *piacere* 'to please' (201) pattern differently from other predicates like *temere* (202) 'to fear', which pattern with canonical agentive predicates in having non-derived subjects, by analysing their syntactic behaviour with respect to anaphoric cliticisation, arbitrary *pro*, embedding under causative verbs, and passivisation. They argue that the *psych-verbs* in Italian differ minimally in the selection of inherent cases which is a lexical parameter. They propose that each verb has a specified case-grid and a theta-grid. The former determines the selection of inherent cases and subsequently, each of the selected inherent cases is linked to specific roles in the theta-grid. The resulting verb entries are further mapped to syntactic projections. Thus, while *temere* has an external argument which is the experiencer and has a simple transitive structure, *preoccupare* and *piacere* have no external arguments. The latter predicates' internal *experiencer* arguments get the inherent accusative case and the inherent dative case, respectively.

(200) Questo preoccupa gianni Underlying structure: [_[[preoccupa questo]$_{V'}$ Gianni]$_{VP}$]$_{TP}$
This worries Gianni

(201) A Gianni piace questo Underlying structure: [_[[piace questo]$_{V'}$ a Gianni]$_{VP}$]$_{TP}$
To Gianni pleases this

(202) Gianni teme questo Underlying structure: [Gianni [teme questo]$_{VP}$]$_{TP}$
Gianni fears this

(Belletti & Rizzi, 1988, p. 291)

Verma and Mohanan (1990) present studies on the experiencer subject construction in various South-Asian languages. A range of languages such as Marathi, Bhojpuri, Sinhalese, Marwari, Oriya, Punjabi, Bengali, Kalasha, Gujarati, Nepali, Maithili, and Malayalam are covered and Verma and Mohanan argue for an approach focused on the semantics of dative nominals. We return in particular to K. P. Mohanan and Mohanan's account of Malayalam dative nominals (in Verma & Mohanan, 1990) in the following.

Landau (2002) follows the same classification of experiencer predicates as adopted by Belletti and Rizzi (1988): Class I involves a nominative experiencer and accusative theme as in *John loves Mary*, Class II involves a nominative theme, accusative experiencer as in *The show amused Bill* and Class III involves a nominative theme, dative experiencer as in *The idea appealed to Julie*. He proposes that all experiencers are mental locations and therefore locatives. Following Belletti and Rizzi (1988), Landau assumes that the object experiencers are assigned inherent case and that such inherent cases are assigned by (phonologically null) prepositions. That is, just like the non-subject locatives that are generally introduced by a preposition, the object experiencers (which are also locatives of mental locations) are introduced by a preposition which is null in bare object experiencers. He argues that all object experiencers are either oblique or dative. Just like the quirky experiencer subjects that move to the subject position, the object experiencers also undergo 'locative inversion'. That is, they undergo raising to the subject position at LF.

However, treating Malayalam experiencer subjects as locatives (mental locations) does not successfully explain their distribution and behaviour. We follow Butt's (2006) analysis where she considers dative experiencers as agents lacking volition. As we will demonstrate shortly, the nominative experiencers in Malayalam pattern with canonical agentive verbs whereas dative experiencers pattern differently showing complete lack of volition.

Butt (2006) argues that the analysis of case, based on the notion of structural/lexical cannot fully explain the significance of case-marking in a language. According to her, languages employ case alternations on arguments in order to express semantic contrasts. The selection of a particular case marker to signal a particular semantic contrast is determined by a language's case system as a whole. Butt (2006) proposes that both the ergative and the dative in the Indo-Aryan languages come from a common ancestor with the former signalling *control* (over the event/action) and the latter *goalness*. She further explains that dative can be interpreted not just as a goal, but in languages that employ the marker to indicate semantic contrasts, it can be construed as an agent with reduced control over the action as in Urdu.

(203) nadya=ko zu ja-na hE
Nadya.F.SG=DAT zoo.M.SG.LOC go-INF.M.SG be.PRS.3SG
'Nadya has/wants to go to the zoo.'

(204) nadya=ne zu ja-na hE
Nadya.F.SG=ERG zoo.M.SG.LOC go-INF.M.SG be.PRS.3SG
'Nadya wants to go to the zoo.'

(Butt, 2006, p. 22)

In the examples above, the ergative subject signals greater control over the event and the dative subject signals obligation or desire. According to Butt, the semantics of case plays a crucial role in understanding its full import and therefore, the analysis of case has to be done both in terms of a language's entire case system and the semantic contrasts the language employs with particular reference to space and control/agency.

A major debate around these dative marked nominals is whether they really are subjects (Ura, 2000; Davison, 2003). Ura (2000) proposes the splitting of grammatical functions (GFs). He notes that a GF results from a certain feature checking relation with T. Thus, the properties of control into adjuncts and binding of a (purely) subject-oriented reflexive result from *phi* and EPP feature checking relations with T, respectively. Subject-verb agreement involves phi-feature (gender, number, and person) checking with T. He demonstrates using these feature checking properties that the dative nominals occurring in subject positions in Japanese qualify as actual 'subjects'.

Davison (2003) classifies the transitive verbs of Hindi/Urdu into four classes (Table 8.1) and in (205) - (208) we see examples of constructions involving predicates that belong to these four classes.

Table 8.1 Classification of transitive verbs in Hindi/Urdu.

	Case of Subject	**Case of Direct Object**	**Case of Indirect Object**
Class A	Obligatorily ergative	Nominative or dative	Dative (lexical)
Class B	Optionally ergative	Nominative or dative	*
Class C	Dative (lexical)	Nominative	*
Class D	Nominative	Lexical postposition	*

(Davison, 2003, p. 201)

(205) Class A: Obligatorily ergative subject

bhaaluu-nee apnee daaNtooN-see baccooN-koo Dar-aa-yaa
bear.M.SG-ERG self's teeth-from children.M.PL-DAT fear-CAUS-PRF.M.SG
'The bear frightened the children with its teeth.'
'(The bear caused the children to be afraid of its teeth)'

(Davison, 2003, p. 201)

(206) Class B: Optionally ergative subject

a. jab maiN-nee maasTar-jii-see sawaal samajh-aa, too
when I-ERG teacher-HON-from question understand-PRF then
maiN-nee usee dubaaraa apnee aap hal kar-kee deekh-aa
I-ERG 3SG.DAT again self's self solution do-PTCP see-PRF
'When I understood the question from the teacher, then I saw it again solved.'

(Nespital, 1997 as cited in Davison, 2003, p. 201)

b. maiN yah baat pahlee hii samajh-aa [ki
I.NOM this matter first only understand-PRF [that
raakee apnii zid-par dṛh hai]
Rakesh self's obstinacy-on fixed is]
'I understood from the first that Rakesh had become fixed on his own obstinacy.'

(Nespital, 1997 as cited in Davison, 2003, p. 202)

(207) Class C: Dative (lexical) subject

mujhee eek upay suujh-aa
I.DAT one means.M.SG.NOM see-PRF.M.SG
'A solution came to my mind/ I saw a solution.'

(Bahri, 1992 as cited in Davison, 2003, p. 202)

(208) Class D: Nominative subject

baccee	bhaaluu-see	Dar-tee	haiN /	Dar ga-yee
children-PL.NOM	bear-M.SG-from	fear-IMPF.M.PL	be.PRS.3PL/	fear go-PRF.M.PL

'The children are afraid of the bear/became afraid of the bear.'

(Davison, 2003, p. 202)

Classes A and B differ from C and D in having a complex vP shell with a light verb v while C and D are said to have only a simple VP projection. Davison proposes that the dative NP and the nominative NP in Class C are equidistant from [Spec, TP] and thus, either of them can raise there and enter into a feature-checking relation with T resulting in different GFs. When the dative DP is raised to [Spec, TP], it acquires subjecthood properties such as binding the reflexive. If the nominative DP raises, then, the dative remains in situ within the VP, and in such cases, the nominative is the subject. Then, the dative shows properties of non-subjects such as binding a pronoun.

Malayalam has two variants for the dative marker: *-nɨ* and *-kkɨ*. The selection of these variants is phonologically conditioned as we saw in Chapter 7. As may be expected, the distribution of dative marked nominals subsumes subject and non-subject argument positions. In non-subject positions, the dative nominals function as indirect objects (of a ditransitive verb) or as adjuncts signalling *goal*, *purpose*, *price*/*value*, and *location* (see Section 7.1). In this chapter, we focus on dative subjects as they occur in a range of constructions in Malayalam.

Dative subject predicates in Malayalam can be broadly divided into three classes based on their distribution in the language (Nizar, 2010). Class I includes dative nominals that are lexically marked by certain light verbs like *ʈoonn-* 'feel', *var-* 'come', *patt-* 'be able' that are mostly N + V constructions (209) and experiencer predicates (210) like *viɛakk-* 'be hungry', *ɖaahikk-* 'be thirsty' etc. It has to be noted that Class I predicates assign the dative case only to nouns bearing the feature [+ANIMATE].

(209) ɾaama-nɨ koopam ʋa-nnu

rama-DAT anger come-PST

'Rama became angry.'

(210) ɾaama-nɨ ʋiɛa-nnu

rama-DAT be hungry-PST

'Rama became hungry.'

Class II predicates are the bound modals *-aam* and *-aŋam*[1] (cliticised form of *veeŋam* 'want') that take infinitival TPs as their complements. These modal predicates lexically case mark the nominals like the Class I predicates.

(211) [kuʈʈi-kaɭ-kkɨ$_i$ [PRO$_i$ kaɭiykk-]aŋam]
child-PL-DAT [PRO play]-MOD
'The children want to play.'

Class III predicates include verbs that occur with the copula *uɳʈɨ* and indicate *possession*. Unlike Class I predicates, Class II and III predicates are not restricted by the animacy feature.

(212) ʂiiʈay-kkɨ paŋam uɳʈɨ
sita-DAT money have.PRS
'Sita has money.'

(213) kaʂeeɾay-kkɨ kaal uɳʈɨ
chair-DAT leg have.PRS
'The chair has leg(s).'

It must be noted that not all experiencer verbs assign the dative case to their subjects. For example, experiencer verbs like *ʂaɲʈooʂiykk-* 'be happy', *ɖuk^{h}iykk-* 'be sad' etc. assign nominative case to their subjects as in (214) and (215).

(214) avaɭ ʂaɲʈooʂi-ccu
she.NOM happy-PST
'She was happy.'

(215) ɾaaman ɖuk^{h}i-ccu
Ram.NOM sad-PST
'Ram was sad.'

Jayaseelan (2004) tries to establish a demarcation between verbs that correspond to physical experiences and those that correspond to mental experiences. He argues that verbs pertaining to physical experiences such as *viɕakk-* 'be hungry' take a dative subject in both the simple (216)

[1]When the modal occurs without the infinitival complement, it surfaces as *veŋam* 'want' and can assign the accusative case to its object.
e.g. ʂiiʈay-kkɨ kiɭiy-e ʋeeŋam
sita-DAT bird-ACC want
'Sita wants the bird.'

and the complex constructions involving light verbs (217), whereas verbs associated with mental experiences like *ʂaɳʈooɻikk-* 'be happy' take a nominative subject for the simple predicates (218) and a dative subject for the complex predicates (219).

(216) avaɭ-kkɨ vɪɛa-ɳɳu (physical experiencer simple)
she-DAT be hungry-PST
'She became hungry.'

(217) avaɭ-kkɨ vɪɛappɨ va-ɳɳu (physical experiencer complex)
she-DAT hunger come-PST
'Hunger came to her.'

(218) avaɭ ʂaɳʈooɻi-ccu (mental experiencer simple)
she.NOM be happy-PST
'She is happy.'

(219) avaɭ-kkɨ ʂaɳʈooɻam ʈooɳɳ-i (mental experiencer complex)
she-DAT happiness feel-PST
'She felt happiness.'

However, his theory cannot account for the behaviour of experiencer predicates like *iɻʈappeʈ-* 'to like' and *ɖaahiykk-* 'be thirsty' which permit both nominative and dative subjects with differences in interpretation. In a footnote, Jayaseelan does identify the verb *iɻʈappeʈ-* and calls its behaviour exceptional. He does not mention other such predicates.

We argue that the essential difference between nominative experiencer predicates and dative experiencer predicates is that the latter express absence of volition/agency of the *subject*/*experiencer* compared to the former. This can be clearly demonstrated with verbs like *iɻʈappeʈ-* and *ɖaahiykk-* where both structures are possible. Jayaseelan also says that his mental experiencer verbs (the ones with the nominative case subjects) encode greater volition. The subject in the nominative is interpreted as having greater volition/control over the action than in the dative subject alternative, as shown in the examples (220) - (223). This is similar to the ergative/nominative difference in Urdu ((203) - (204)) discussed in Butt (2006) where the ergative subject shows more control/agency over the event and the nominative shows reduced control/agency. Malayalam is a non-ergative language and the nominative/dative case is used to indicate the two semantic states.

(220) eni-kkɨ avaɭ-e iʂṭappe-ṭṭu [non-volition]
I-DAT she-ACC like-PST
'I liked her (despite myself).'

(221) ɲaan avaɭ-e iʂṭappe-ṭṭu [volition]
I.NOM she-ACC like-PST
'I (willingly) liked her.'

(222) eni-kkɨ ḍaahi-ccu [non-volition]
I-DAT thirst-PST
'I am thirsty.'

(223) ɲaan ḍaahi-ccu [volition]
I.NOM thirst-PST
'I thirsted (for something).'

Unlike Malayalam, in Tamil, the experiencer verbs that take the nominative case are morphologically different from those that take the dative case. For example, in Malayalam, the nominative experiencer verb *ʂaṇṭooʂiykk-* 'be happy' and the dative experiencer verb *viɕakk-* 'be hungry' are morphologically identical. In Tamil, the nominative experiencer verb *ʂaṇṭooʂappeṭa-* 'be happy' is a complex predicate comprising a noun and a light verb and morphologically different from the dative experiencer verb *paʂikka-* 'to be hungry'.

Landau (2002, p. 7) demonstrates that adverbs like *deliberately* cannot occur with stative verbs by virtue of their semantics since they lack agency. Similarly, in Malayalam, adverbs that qualify the predicate can pick up on the available semantic property of volition/agency of the subject and can be used as a diagnostic. For example, only the agentives (224) and nominative experiencers (225) can occur with the adverb *ariyaaṭe* 'unknowingly' but not the dative experiencers (226).

(224) avaɭ ariyaaṭe praarṯʰi-ccu
she.NOM unknowingly pray-PST
'She prayed unknowingly.'

(225) avaɭ ariyaaṭe ʂaṇṭooʂi-ccu
she.NOM unknowingly be happy-PST
'She became happy unknowingly.'

(226) *avaɭ-kkɨ ariyaat̪e viɕa-n̪n̪u
she.DAT unknowingly be hungry-PST
'She became hungry unknowingly.'

The adverb *ariɲɲukoɳʈɨ* 'knowingly' also has the same effect as *ariyaat̪e* 'unknowingly'.

(227) avaɭ ariɲɲukoɳʈɨ praart̪ʰi-ccu
she.NOM knowingly pray-PST
'She prayed knowingly.'

(228) avaɭ ariɲɲukoɳʈɨ s̪an̪t̪oos̪i-ccu
she.NOM knowingly be happy-PST
'She became happy knowingly.'

(229) *avaɭ-kkɨ ariɲɲukoɳʈɨ s̪an̪t̪oos̪am va-n̪n̪u
she.DAT knowingly happiness come-PST
'Happiness came to her knowingly.'

(230) *avaɭ-kkɨ ariɲɲukoɳʈɨ viɕa-n̪n̪u
she.DAT knowingly be hungry-PST
'She knowingly became hungry.'

Likewise, the dative experiencers cannot occur with the adverb *manappoorvam* 'deliberately'.

(231) avaɭ manappoorvam praart̪ʰi-ccu
she.NOM deliberately pray-PST
'She deliberately prayed.'

(232) avaɭ manappoorvam s̪an̪t̪oos̪i-ccu
she.NOM deliberately be happy-PST
'She deliberately became happy.'

(233) *avaɭ-kkɨ manappoorvam s̪an̪t̪oos̪am va-n̪n̪u
she.DAT deliberately happiness come-PST
'Happiness deliberately came to her.'

(234) *avaɭ-kkɨ manappoorvam viɕa-n̪n̪u
she.DAT deliberately be hungry-PST
'She deliberately became hungry.'

It can be seen that the agentive verb *prarʈʰiykk-* 'pray' and the nominative experiencer *ʂaɳɳtooʂiykk-* 'be happy' pattern together and differently from the dative experiencer *viɛakk-* 'be hungry'. With respect to case-marking and adverb usage, experiencers like *ʂaɳɳtooʂiykk-* 'be happy' are similar to agentive predicates showing their semantic differences from dative experiencer predicates.

The modal suffix *-aam* that select infinitival TP complements also shows alternation between dative and nominative case-marked subjects with different semantic interpretations. With a nominative subject, *-aam* indicates *possibility* (235) and with dative subjects, it shows *potential*/*ability* or *permission* as in (236) and (237), respectively.

(235) kiɭi parakk-aam
Bird fly-MOD
'The bird may fly away.'

(236) kiɭiy-kkɨ parakk-aam
bird-DAT fly-MOD
'The bird can fly.'

(237) ʋiɖʱyaarʈʰi-kaɭ-kkɨ pook-aam
student-PL-DAT go-MOD
'Students may (are permitted to) leave.'

The modal *-aɳam* which is a cliticised form of the verb *veeɳam* 'want' employs the dative to show internal need, wish, desire or compulsion (238) and the nominative subject to indicate external demand, compulsion or requirement (239) (K. P. Mohanan & Mohanan, 1990).

(238) ammay-kkɨ kuʈʈiy-e aʈiykk-aɳam
mom-DAT child-ACC eat-MOD
'The mother wants to beat the child.'

(239) amma kuʈʈiy-e aʈiykk-aɳam
mom.NOM child-ACC eat-MOD
'The mother must beat the child.'

(K. P. Mohanan & Mohanan, 1990, p. 45)

However, unlike *-aam*, which is a cliticised form of *aakum* 'will be' which is the future form of *aak-* 'be' (Rajaraja Varma, 1896), *veeɳam* is not an alternating predicate. The nominative subjects indicating obligation as in (239) always have an underlying subject which can be

a discourse antecedent. The underlying subject of (239), which is an impersonal construction meaning 'It is necessary for the mother to beat the child', is an arbitrary *pro* as shown in (240). Here the arbitrary *pro* corresponds to a dative subject. The presence of an underlying dative subject is revealed in constructions (241)-(243). This becomes clearer especially in constructions where dative NPs surface as overt subjects as in (242) and (244).

(240) [pro_{arb} [amma kuʈʈiy-e aʈiykk-]$_{TP}$ aɳam]$_{TP}$

(241) ɲaan aiʂkriim ʋaaŋŋ-aɳam
I.NOM icecream buy-MOD
[pro_{dat} [ɲaan aiʂkriim ʋaaŋŋiykk-]$_{TP}$ aɳam]$_{TP}$
'I should buy ice cream.'

The import of (241) is that 'Someone wants me to buy ice cream'. This is shown in (242) where only dative subjects are permitted as the external argument. The sentences (243) and (244) further clarify this distinction.

(242) [aʋa-ni̵ [ɲaan aiʂkriim ʋaaŋŋ-]$_{TP}$ aɳam]$_{TP}$
he-DAT I.NOM ice cream buy MOD
'He wants me to buy ice cream.'

(243) ɾaaman kat^{h}a paray-aɳam
rama.NOM story say-MOD
[pro_{dat} [ɾaaman kat^{h}a paray-]$_{TP}$ aɳam]$_{TP}$
'Rama should tell a story.'

(244) ʂiiʈay-kki̵ ɾaaman ʈann-ooʈi̵ kat^{h}a paray-aɳam
sita-DAT rama.NOM self-SOC story say-MOD
[ʂiiʈay-kki̵$_{i}$ [ɾaaman$_{j}$ ʈann-ooʈi̵$_{i/*j}$ kat^{h}a paray-]$_{TP}$ aɳam]$_{TP}$
'Sita wants Rama to tell her a story.'

Class I dative nominals assign the accusative case to the objects (245). However, Class II predicates (when they are clausal complement predicates) and Class III predicates cannot assign the accusative case to objects. In Class II clausal complement constructions that have an apparent accusative marked object, the accusative case is actually assigned by the embedded infinitival verb (246). Class II predicates' inability to assign the accusative is demonstrated in (247). Class III predicates assign the nominative case to the objects (248). We assume that while the copula assigns the dative case to the external argument which is the *possessor*, the *possessed*

nominal bears a structural case (nominative). In Tamil, this nominal also triggers phi-agreement on the verb, though in Malayalam this is not visible because of the lack of such subject-verb agreement. The licensing of features in these constructions will be discussed in Section 8.2.

(245) ɾaama-nɨ **ʂiiʈay-e** iʂʈam-aaɳɨ
rama-DAT sita-ACC like-be.PRS
'Rama likes Sita.'

(246) ʂiiʈay-kkɨ kiɭiy-e piʈiykk-aɳam;
[ʂiiʈa-kkɨ$_i$ [PRO$_i$ kiɭiy-e piʈiykk-]$_{TP}$ aɳam]$_{TP}$
sita-DAT bird-ACC catch-MOD
'Sita wants to catch the bird.'

(247) ʂiiʈay-kkɨ kiɭi uraŋŋ-aɳam;
[ʂiiʈay-kkɨ [kiɭi uraŋŋ-]$_{TP}$ aɳam]$_{TP}$
sita-DAT bird.NOM sleep-MOD
'Sita wants the bird to sleep.'

(248) ʂiiʈay-kkɨ ɳaalɨ paɕu-kkaɭ uɳʈɨ
sita-DAT four cow-PL.NOM be.PRS
'Sita has four cows.'

We have seen the properties of predicates that have dative subjects. In the following, we will critically analyse the various accounts in the literature of the Malayalam dative subject phenomenon. We also establish that the dative nominals, such as those we have discussed above, are indeed *subjects* by analysing their syntactic behaviour. We will use Ura's (2000) Agrless Checking theory to account for the subject properties of dative nominals and to show how these nominals enter into various syntactic relations such as EPP, phi-feature checking etc. The overt dative case assignment is aligned to the theta-grid of individual predicates and instantiates inherent case selection. Finally, we employ the acquisition data to a) substantiate the analysis, and b) demonstrate the various effects predicted by such an analysis on language development itself, i.e., on the observable acquisition patterns of dative subjects.

K. P. Mohanan and Mohanan (1990) attempt to provide a unified account of the subject and non-subject dative nominals by subsuming their occurrence within the ambit of the two semantic roles of *goal* and *possession*. They demonstrate this by using two instances involving the verb *var-* 'to come' which can function as either a light verb or a main verb.

(249) baala-nɨ ɖukʰam / ʂaɳʈooʂam ʋa-ɳɳu
boy-DAT grief / happiness come-PST
'The boy became sad/happy.'
(Lit: To the boy came happiness or sadness.)
(K. P. Mohanan & Mohanan, 1990, p. 47)

(250) baalan ʂkuuɭ-il-eekkɨ ʋa-ɳɳu
boy.NOM school-L-DAT come-PST
'The boy came to school.'
(K. P. Mohanan & Mohanan, 1990, p. 48)

K. P. Mohanan and Mohanan analyse the dative nominals in both the examples above as targets of movement and as bearers of the same semantic category *goal* which can combine with other semantic categories such as *experiencer* and *location* forming categories like *experiencer goals*. They extend this idea to dative experiencer verbs by making a distinction between dative and nominative experiencers. They treat the former as signalling the advent of a state *to* an individual (thereby assigning the *goal* theta role to the individual) and the latter as the change of state *in* an individual. They treat the dative experiencer predicates like *viɛakk-* 'be hungry', *ɖaahikk-* 'be thirsty', *uʂɳikk-* 'feel hot' etc. as *experiencer goals*.

As Nizar (2010) points out, this analysis fails to account for the dative assigning modals like *-aam* and *-aɳam*. Despite acknowledging the behaviour of these modals, K. P. Mohanan and Mohanan do not explain how the dative case assignment can account for modal constructions under the semantics of *goal* or *possession*. For example, the dative nominal marked by *-aɳam* signals an internal need (251) and can only be labelled *experiencer*, but the dative assigned by *-aam* signals *potential*/*compulsion* (252). The semantics of *goal* cannot be associated with these dative nominals.

(251) aʋa-nɨ ʂkuuɭ-il ʋaɾ-aɳam
he-DAT school-LOC come-MOD
'He wants to come to school.'

(252) aʋa-nɨ ʂkuuɭ-il ʋaɾ-aam
he-DAT school-LOC come-MOD
'He can come to school.'

Moreover, not all dative subject constructions can be treated as goals.

(253) aʋa-nɨ t̪alaʋeed̪ana ʋa-n̪n̪u
he-DAT headache come-PST
'He got a headache.' (Headache came to him)

(254) aʋa-nɨ t̪alaʋeed̪ana aaɳɨ
he-DAT headache be.PRS
'His head aches.' (He is experiencing a headache)

(255) aʋa-nɨ t̪alaʋeed̪ana uɳʈɨ
he-DAT headache have.PRS
'He has a headache.'

Native speakers agree that there is a clear semantic difference between (253) and (254). While the dative nominal in (253) may be seen as a *goal* (movement towards), the dative nominal in (254) is by no means a *goal*, nor does it signal *possession* since there is a semantic difference between (254) and (255). While the subject in (255) is a *possessor*, the one in (254) is straightforwardly an *experiencer* (dative and not nominative) with the light verb signalling the state of the individual. Hence, their attempt at generalising the *goal-ness* and *possessed-ness* fails in these cases.

In another analysis of Malayalam dative subjects, Jayaseelan (1999, 2004) argues against the treatment of the dative NP as a subject at all. In many languages, the binding of subject-oriented anaphors by a c-commanding antecedent (typically long-distance) is taken as a diagnostic of subjecthood and as an identifier of the structural position of the dative nominal. It is seen that indirect objects cannot serve as binders, unlike the *experiencer* argument[2]. In his account of possessive-experiencer datives, Jayaseelan uses anaphor binding facts to argue against the use of this test altogether. His argument runs as follows: since the anaphor *t̪aan* can also be bound by a possessor NP (256) and a direct object (257), it does not need a c-commanding antecedent and, hence, the test itself is empirically unreliable.

(256) ɟooɳ-inte$_i$ ʋicaaɾam [meeɾi$_j$ t̪ann-e$_{i/*j}$ s̪neehikk-un̪n̪-illa en̪n̪ɨ] aaɳɨ
john-GEN thinking.N mary.NOM self-ACC love-PRS-NEG COMP be.PRS
'John's thinking (impression) is that Mary does not love him.'

[2]The dative experiencer also occurs linearly ahead of objects. The canonical word order being SOV, prima facie it seems to be in the subject position.

(257) [ʈan-te$_i$ makaɭ-uʈe ʋiʋaaha-kaaryam] ɟooɳi-ne$_i$ alaʈʈ-i
self-GEN daughter-GEN marriage-matter john-ACC bother-PST
'(The thought of) his daughter's marriage bothered John.'
(Jayaseelan, 2004, p. 236)

Malayalam *ʈaan* is a simple, long-distance, pronominal-anaphor (like Dutch *zich*) and is subject-oriented in all cases. It must be free in its governing category (like other pronouns) but it is phi-deficient for gender and therefore, is preferentially bound by any c-commanding subject with which it will share the gender feature. Other nominals like a matrix object or an indirect object cannot bind *ʈaan* as can be seen in (258). *ʈaan* can also be bound by a discourse antecedent.

(258) ɟooɳ$_j$ ʈan-te$_{i/*j}$ peena meeɾiy-kki$_j$ koʈu-ttu
john self-GEN pen mary-DAT give-PST
'John gave Mary his pen.'

Jayaseelan's argument throws away the bulk of the data that deals with the binding of *ʈaan* in the quest to get experiencer dative nominals to pattern with indirect objects or other oblique nominals. In (256) which is a copular construction, the possessor NP is within the subject phrase and the anaphor is within the complement, which is a CP. Example (257) has an unaccusative verb *alaʈʈ-* and the subject is a derived one (Pesetsky, 1995). Neither the derived subject nor the possessor can bind *ʈaan*. However, they can be co-referential with it. In languages with long distance anaphors, the overall *point of view* also matters. *ʈaan* is acceptable in these sentences only when the point of view coheres with the referent of the possessor or the object. Otherwise, the regular pronoun *avan* [he.NOM] is appropriate and *ʈaan* is ungrammatical. In other words, *ʈaan* has a logophoric function as well which cannot be conflated with its general pattern of use. So these examples cannot be used as counter-evidence for the subject-orientation of *ʈaan* in general. Dative nominals can bind *ʈaan* as can be seen in (259) and (260) where it occurs in the object position and as the possessor of the object, respectively. In the former cases, it requires a long distance antecedent whereas, in the latter, the antecedent can be a local one.

(259) ɟooɳi-ni meeɾi ʈan-ne ka-ɳʈu eɳɳi ariy-aam
john-DAT mary.NOM self-ACC see-PST COMP know-MOD
[ɟooɳi-ni$_i$ [PRO$_i$ [meeɾi$_j$ ʈan-ne$_{i/*j}$ kaɳʈu eɳɳi]$_{CP}$ ariy-]$_{TP}$ aam]$_{TP}$
'John knows that Mary saw him.'

(260) ɟooɳi-nɨ t̪a-nte kaar uɳʈɨ
john.DAT self-GEN car have.PRS
[ɟooɳinɨ [t̪a-nte kaar]$_{DP}$ uɳʈɨ]$_{TP}$
'John has his (own) car.'

The evidence for both the position of the dative nominals as the highest in the structure and their subject-hood, since they are able to bind the pronominal anaphor, appears incontrovertible.

A second argument that Jayaseelan (2004) advances against the subjecthood of dative nominals is control. The ability of a dative nominal to control the subject of the embedded infinitival clause is taken as empirical evidence of its subjecthood. Jayaseelan believes that since object control is also possible, the dative NP's control properties do not underscore or confirm their subjecthood. A better test for the subjecthood of dative nominals in his opinion would be sentences where the dative NP is a null subject (PRO) in an infinitival complement clause (i.e., an infinitival experiencer predicate). In other words, the ability of a potential experiencer argument to be controlled by a matrix subject would establish its subjecthood. He states unequivocally that such a construction is impossible in Malayalam, However, Nizar (2010) points out that such a construction is indeed possible and perfectly grammatical (261).

(261) [aʋaɭ$_i$ [PRO$_i$ s̪an̪t̪oos̪am t̪oon̪n̪-aan]$_{TP}$ s̪rami-ccu]$_{TP}$
she happiness feel-INF try-PST
'She tried to feel happy.'

(Nizar, 2010, p. 31)

It must be noted that in the above example, an overt dative marked nominal (but not a nominative NP) is permitted as the subject of the infinitival clause instead of PRO (262). Inherently case marked and theta marked arguments can appear in the subject position of infinitivals.

(262) [aʋaɭ$_i$ [t̪aniy-kkɨ$_i$/*t̪aan s̪an̪t̪oos̪am t̪oon̪n̪-aan]$_{TP}$ s̪rami-ccu]$_{TP}$
she self-DAT/*self.NOM happiness feel-INF try-PST
'She tried to feel happiness.'

The fact that PRO alternates with a dative marked nominal underscores the point that dative NPs are subjects as shown in (263) and (264).

(263) avа-nɨ pook-aan ariy-aɳam
he-DAT go-INF know-MOD
[avа-nɨ$_i$ [PRO$_i$ [PRO$_i$ pook-aan]$_{TP}$ ariy-]$_{TP}$ aɳam]$_{TP}$
'He wants to know how to go (there).'

(264) avа-nɨ ṯaniy-kkɨ pook-aan ariy-aɳam
he-DAT self-DAT go-INF know-MOD
[avа-nɨ$_i$ [**ṯaniykkɨ**$_i$ [PRO$_i$ pook-aan]$_{TP}$ ariy-]$_{TP}$ aɳam]$_{TP}$
'He wants to know how to go (there).'

Jayaseelan's third argument is that the lack of subject-verb agreement with the dative does not favour the subjecthood of dative nominals. This argument is irrelevant because Malayalam lacks subject-verb agreement altogether and nominative subjects are no more privileged than dative subjects. In languages like Tamil, Kannada etc. with phi-agreement, the main verb bears default agreement but that does not argue against treating the dative NP as a subject. Inherently case-marked nominals in subject positions do not show phi-agreement with their predicates in most of the world's languages. We return to this aspect in Section 8.2 to discuss the subject properties of the dative nominals.

Using the above as empirical arguments, Jayaseelan proposes that all of the so-called dative subject constructions, in fact, contain a nominative NP, which can either be overt (as in Class III predicates with a DAT-NOM combination) or covert (as in Class I and II predicates). In support, he offers the following two sentences where he says that the dative nominal (subject) (265) can freely alternate with other oblique nominals (266) which shows that dative is an oblique case.

(265) eniy-kkɨ kaɻiy-illa ṉinn-e ṉookk-aan
I-DAT be able-NEG you-ACC look after-INF
'I cannot look after you.'

(266) enn-ekkoɳʈɨ kaɻiy-illa ṉin-ne ṉookk-aan
I-INSTR be able-NEG you-ACC look after-INF
'I cannot look after you.'

(267) eniy-kkɨ / *enn-ekkoɳʈɨ ṉin-ne iṣṭam illa
I-DAT/ I-INS you-ACC liking NEG
'I don't like you.'

(Jayaseelan, 2004, p. 240)

However, in (266), *koɳʈɨ* is not an instrumental case marker but a participial selecting an accusative case object with a null subject as enumerated in (268). Jayaseelan's example sentence (266) does not provide the correct parse of the morphemes or the underlying structure. This construction certainly does not provide evidence that the dative NP is an oblique argument that freely alternates with other obliques and this is further supported by (267) where such alternation is impossible. The underlying structure of (266) with correct morpheme parses are given in (268). Further, (269) shows that dative NP is the underlying subject of (268) which clearly argues for the subjecthood of dative NPs.

(268) [pro$_i$ [enn-e koɳʈɨ] [PRO$_i$ ɳinn-e ɳookk-aan] kaɻiy-illa]
I-ACC V.PTCP you-ACC look after-INF be able-NEG
'I cannot look after you.'

(269) [eniy-kki [enn-e koɳʈɨ] [PRO$_i$ ɳinn-e ɳookk-aan] kaɻiy-illa]
I.DAT I-ACC V.PTCP you-ACC look after-INF be able-NEG
'I cannot look after you.'

In constructions like (265), the dative nominal can in fact bind *ʈaan* in the object position of the embedded clause (via PRO) (see (270)). Given this, the dative nominal must be the argument of the matrix verb since the embedded verb is a regular, transitive verb, in the infinitival form and will not be able to case mark its subject. There is no covert nominal as a potential binder for the anaphor as Jayaseelan says must be the case.

(270) avɒa-ni$_i$ [PRO$_i$ ʈanneʈann-e$_i$ ɳookk-aan] kaɻiy-illa
he-DAT self-ACC look-INF be able-NEG
'He cannot look after himself.'

Finally, if there is a null/pleonastic, nominative pronoun in all the dative subject constructions that lack a nominative NP, as Jayaseelan says they do, then it would mean that the valency of all the predicates in question is increased by one, since a null pronoun is an argument position regardless of its theta properties. He fails to take into account this potential across-the-board valency change. There is no other evidence which might indicate that these verbs show an increase in valency and their semantics does not support it either. This move appears to be uneconomical and runs counter to all the facts at hand.

In contrast, Nizar (2010) argues for the subjecthood of dative NPs (especially the Class I experiencer subjects) based on their syntactic behaviour. She applies several syntactic tests

to demonstrate that the dative NPs and their nominative counterparts in the subject position behave alike in a variety of structures. She demonstrates that in constructions involving sequential events with a matrix clause and an embedded clause, where the embedded verb is a finite participial, the subject NPs are coreferential with the null subject (271).

(271) avaɬ-kkɨ$_i$ ɖaahicc-iʈʈɨ [pro$_i$ veɭɭam kuʈi-ccu]
she-DAT be thirsty-PFV water drink-PST
'She became thirsty and drank water.'

(Nizar, 2010, p. 19)

(272) *avaɬ-kkɨ ɖaahicc-iʈʈɨ [avan veɭɭam kuʈi-ccu]
she-DAT be thirsty-PFV he.NOM water drink-PST
'She became thirsty and he drank water.'

The ungrammaticality of (272) shows that the matrix and the embedded subjects have to be coreferential. Nizar does not include these, but (271) has other variants as in (273)[3].

(273) a. [avaɬ-kkɨ$_i$ ɖaahicc-iʈʈɨ [**avaɬ**$_i$ veɭɭam kuʈi-ccu]]
she-DAT drink-PFV she.NOM water drink-PST

b. [**avaɬ**$_i$ [**pro**$_i$ ɖaahicc-iʈʈɨ] veɭɭam kuʈi-ccu]]
she.NOM drink-PFV water drink-PST

c. avaɬ$_i$ [ʈaniy-kkɨ$_i$ ɖaahicc-iʈʈɨ] veɭɭam kuʈi-ccu
she.NOM self-DAT drink-PFV water drink-PST

d. *avaɬ-kkɨ$_i$ ɖahicc-iʈʈɨ ʈaan$_i$ veɭɭam kuʈi-ccu
she-DAT drink-PFV self water drink-PST
ʈaan$_i$ veɭɭam [avaɬ-kkɨ$_i$ ɖaahicc-iʈʈɨ] kuʈi-ccu (underlying structure)

That is, in both subject NP positions, the full-NP can be realised (273a) or either one can be left null (271 vs 273b). The point is that there is referential identity since the subjects involved in the two events are the same though their case-marker can vary depending on the predicate. (273c) shows that an anaphor in the embedded dative subject can be bound by the matrix subject but not vice-versa. The ungrammaticality of (273d) arises from the fact that *ʈaan*

[3]These are probably VP conjoined sentences. There is no overt conjunction but the subject is the same for both verbs and in fact, must be the same. When both predicates are of the nominative-accusative type, only one subject is visible. When the predicates are a mix of nominative-accusative and dative subject, the dative nominal can be overtly present since it is generated VP internally. But we do not pursue this here except to show that dative nominals are really subjects.

cannot be bound by *avaḷkk-* and further, cannot take a discourse antecedent since these subjects have to co-refer.

Nizar demonstrates that an indirect object cannot share the reference with such a null-subject (274). This underscores the difference between dative subjects and dative objects.

(274) [avaḷ$_j$ kuṭṭiy-kk-i$_j$ paiṣa koṭuṯṯ-iṭṭ-i] [pro$_{i/*j}$ kaṭay-il pooy-i]
she.NOM child-DAT money give-PTCP shop-LOC go-PST
'She gave the child money and went to the store.'

(Nizar, 2010, p. 18)

The above string can also be assigned the structure [*avaḷ*$_i$ [pro$_{i/*j}$ *kuṭṭi-kk-*$_j$ *paiṣa koṭuṯṯ-iṭṭi*] *kaṭay-il pooy-i*]. However, in neither structure can the indirect object co-refer with the embedded subject NP, whether overt or null. The dative NP in (271) and the nominative NP in (275) behave alike in sharing the coreference with the null-subject.

(275) [avaḷ$_j$ iruṇṇ-iṭṭ-i [pro$_i$ veḷḷam kuṭi-ccu]]
she.NOM sit-PFV water drink-PST
[pro$_i$ iruṇṇ-iṭṭ-i [avaḷ$_j$ veḷḷam kuṭi-ccu]]
'She sat and drank water.'

In control structures, Nizar (2010) shows that PRO can be controlled by either a nominative or a dative subject in the matrix clause (276). Conversely, a null subject of an embedded dative predicate can be controlled by the matrix subject (277). Further, (278) shows an instance of a dative controller with an embedded dative predicate.

(276) nominative controller with embedded dative
[avaḷ$_j$ [PRO$_i$ saṇṭooṣam ṭooṇṇ-aan]$_{TP}$ sṛami-ccu]$_{TP}$
she.NOM happiness feel-INF try-PST
'She tried to feel happy.'

(Nizar, 2010, p. 31)

(277) dative controller with embedded nominative
[avaḷ-kk-i$_i$ [PRO$_i$ ameerikkay-il pook-aan]$_{TP}$ aagraham illa]$_{TP}$
she-DAT america-LOC go-INF wish NEG
'She does not wish to go to America.'

(Nizar, 2010, p. 21)

(278) dative controller with embedded dative

[aʋaɭ-kkɨ$_{i}$ [PRO$_{i}$ s̪an̪t̪oos̪am t̪oon̪n̪-aan]$_{TP}$ s̪aad̪ʱiykk-um]$_{TP}$

she-DAT happiness feel-INF be able-FUT

'She can feel happy.'

Having established the subject properties of the dative subject, Nizar (2010) concludes by saying that these tests are only reliable for dative experiencer predicates but not for modal verbs belonging to Class II. Nizar cannot account for the control properties of the Class II predicates because she treats the modals *-aam* and *-aɳam* as common suffixes, and not as bound predicates that take infinitival TPs as their complements. The fact that these are bound modals that take infinitivial TPs as their complements automatically shows that these are control constructions. We show that, just as dative experiencer subjects can co-refer with the null-subject of a finite participial embedded clause, the subjects of dative modals can also co-refer with the null-subject of an embedded clause.

(279) [aʋaɭ-kkɨ$_{i}$ [PRO$_{i}$ uraŋŋ-]$_{TP}$ aɳam]$_{TP}$

she-DAT sleep.INF-MOD

'She wants to sleep.'

(280) [aʋaɭ-kkɨ$_{i}$ [PRO$_{i}$ uraŋŋ-]$_{TP}$ aam]$_{TP}$

she-DAT sleep.INF-MOD

'She can sleep.'

(281) [aʋaɭ-kkɨ$_{i}$ [PRO$_{i}$ kiɭi-ye ʋaɭart̪t̪-aan]$_{TP}$ ʋeeɳam]$_{TP}$

she-DAT bird-ACC raise-INF want

'She wants to raise the bird.'

Finally, Nizar uses causativisation to analyse dative subject constructions. Morphological causativisation is highly productive in Malayalam as we have seen in Chapter 6. In such causative constructions, both dative and nominative NPs behave alike. The subject of the transitive verb in (282) is changed to the object of the predicate with the introduction of the causer as the new subject (valency increase by one) in (283). In the dative construction in (284), the dative nominal is converted to a direct object in (285). In other words, the valency increase of the derived predicate instantiates a new argument/case/theta structure.

(282) paɕu-kkaɭ pullɨ t̪i-n̪n̪u
cow-PL.NOM grass.ØACC eat-PST
'The cows ate grass.'

(283) ɲaan paɕu-kkaɭ-e pullɨ t̪ii-tt-i
I.NOM cow-PL-ACC grass eat-CAUS-PST
'I fed the cows grass.'

(Nizar, 2010, p. 24)

(284) aʋaɭ-kkɨ ʋeed̪ani-ccu
she-DAT hurt-PST
'She was hurt.'

(285) aʋan aʋaɭ-e ʋeed̪ani-pp-iccu
he.NOM she-ACC to pain-CAUS-PST
'He hurt her.'

Causativisation in Malayalam shows that when valency change is effected, dative and nominative subjects are both transformed into accusative objects which in turn shows that both the NPs are same kind of syntactic beings[4].

It can now be clearly seen that the *goal* and *possession* role attributions (K. P. Mohanan & Mohanan, 1990) do not provide a proper account of the behaviour of Malayalam dative subjects. However, K. P. Mohanan and Mohanan's (1990) initial idea that the dative case assignment is directed by specific semantic roles appears to be a move in the right direction. As we saw earlier in this chapter, there are dative-nominative alternants of the same verbs (including modals) and this difference is centred on the semantic differences of the two predicates and their argument selection properties (case and theta role). Jayaseelan's (1999, 2004) argument that the dative nominals are oblique and that there is always a nominative NP (which is either overt or covert)

[4]This makes an interesting contrast in Tamil which has periphrastic causatives. Here the causative verb typically changes the erstwhile embedded nominative subject into an accusative object. Dative subjects are not modified; compare 1 and 2:

1. raaman [siitaav-ai saapiDa] veitaan
   rama.NOM sita-ACC eat.INF make.3SG.M

2. raaman [sitaa-kku pasikka] veitaan
   rama.NOM sita-DAT hunger.INF make.3SG.M

The syntax of structural case-assignment is different from the morphology of valency change and its impact on the s- and c- selection properties of the derived causative.

in all such constructions is clearly problematic, not the least because of the resulting valency change which is unaccounted for, but also because there are constructions that are grammatical in Malayalam which he predicts do not exist. None of the binding or control facts run counter to the treatment of the dative nominal as a subject. Nizar (2010) has also demonstrated that the behaviour of experiencer predicates in different syntactic contexts underscores the subject properties of dative nominals. We have shown that the property of control can be extended to modals as well. In the next section, we will analyse the subject properties of dative nominals in greater detail.

8.2 Structure of Dative Subject Constructions

In order to establish the subjecthood of the dative nominals with Class I, II and III predicates that we saw earlier, we can apply a number of empirical tests. These tests are intended to compare the behaviour of dative nominals occurring in subject positions to the nominative marked nominals, which are considered to be the canonical subjects, since we need to establish the position and grammatical status of these dative arguments. In the previous section, while arguing against certain extant proposals of Malayalam dative subjects, we have already touched upon several of these properties. Here we make a more systematic case for the subjecthood of these dative arguments. We show that the typical *subject* properties of anaphor binding, control, agreement, nominalisation, as well as verbal properties of tense and aspect marking, case assignment and theta selection collectively confirm the subjecthood of these atypical subjects.

Like the other Dravidian languages, the third person pronoun *ʈaan* is phi-deficient for gender and requires a suitable, c-commanding antecedent which must be the subject (and not other non-subject c-commanding nominals). The subject of the local clause can serve as an antecedent only in reflexive constructions. Otherwise, any matrix subject with the appropriate phi-features can bind the pronoun (subject to principle B). As can be seen in (286) - (288), a dative subject can serve as a suitable antecedent like a nominative subject (289), a property that we discussed in the last section as well[5].

(286) [ɾaama-ni$_i$ [ʈa-nte$_i$ maaʈaa-piʈaa-kka[-e]$_{DP}$ oorma ʋa-nnu]$_{TP}$
rama-DAT self-GEN mother-father-PL-ACC memory-N come-PST
'Rama remembered his parents.'

[5]We pointed out earlier that *ʈaan* can also be logophor.

(287) [ɾaama-ni$_i$ [t̪aan$_i$ paʈʰikk-]$_{TP}$ aɳam]$_{TP}$
rama-DAT self study.INF-MOD
'Ram wants to study.'

(288) [ʋa-ni$_i$ [t̪aniy-kki$_i$ ʋiɕakk-]$_{TP}$ aɳam]$_{TP}$
he-DAT self.DAT be hungry-MOD
'He wants to be hungry.'

(289) [ɾaaman$_i$ [t̪an-te$_i$ kuʈʈi-kaɭ-e]$_{DP}$ ura-kk-i]$_{TP}$
rama.NOM self-GEN child-PL-ACC sleep-CAUS-PST
'Rama put his children to sleep.'

The understood subject (PRO) of an embedded, non-finite predicate is determined by a suitable controller (subject or object) in the matrix clause. PRO is distributionally restricted to external, theta position without case and picks up referential features form the controller /antecedent. Being controlled by the subject of the matrix clause and being able to control the subject of a subordinate clause are characteristic features of subjecthood. Nominative NPs which are the canonical subjects show these properties and we have seen that dative subjects do so too. In (290), PRO is in the subject position of a dative subject predicate and in (291) that of a nominative subject predicate. In (292) the PRO can be controlled as well as *t̪aan* could be bound by either the nominative or the dative subject which show that both are equally potential subject cases.

(290) [ʋaɭ$_j$ [PRO$_i$ s̪an̪t̪ooʂam t̪oon̪n̪-aan]$_{TP}$ s̪rami-ccu]$_{TP}$
she.NOM happiness feel-INF try-PST
'She tried to feel happy.'

(Nizar, 2010, p. 31)

(291) [ʋaɭ-kki$_i$ [PRO$_i$ pook-aan]$_{TP}$ kaɻiy-um]$_{TP}$
she-DAT go-INF be able-FUT
'She can go.'

(292) ɾaaman ʂiiʈay-kkɨ ʈa-nte puʂʈakam ʋaayikk-aɳam
rama.NOM sita-DAT self-GEN book read.INF-MOD
eɳɳɨ para-ɲɲu
COMP say-PST
[ɾaaman$_i$ [ʂiiʈay-kkɨ$_j$ [PRO$_{i/j}$ ʈa-nte$_{i/j}$ puʂʈakam ʋaayikk-]$_{TP}$ aɳam
eɳɳɨ]$_{CP}$ para-ɲɲu]$_{TP}$
'Rama said Sita wants to read his/her book /
Rama said Sita wants Rama to read his/her book.'

A third property that signals subjecthood is the behaviour of dative NPs in equative constructions. Both the dative and the nominative NPs can be derived as the subject of an equative construction with a null copula, indicating that the nominals occupy the same *subject position* and continue to be the subjects when the predicates are nominalised to form the equative constructions.

(293) ɾaaman bʱakʂaɳam kaɻi-ccu
rama.NOM food eat-PST
'Rama had food.'
bʱakʂaɳam kaɻiccavan ɾaaman 'the one who ate' = Rama

(294) ɾaama-nɨ ʂiiʈay-e iʂʈappeʈʈu
rama-DAT sita-ACC like.PST
'Rama liked Sita.'
ʂiiʈaye iʂʈappeʈʈavan ɾaaman 'the one who liked Sita' = Rama

(295) ɾaama-nɨ ʂiiʈay-e ɳaʂʈappeʈʈu
rama-DAT sita-ACC lose.PST
'Rama lost Sita.'
ʂiiʈaye ɳaʂʈappeʈʈavan ɾaaman 'the one who lost Sita' = Rama

As might be expected, the nominative subject predicates and dative subject predicates of all the three classes occur in different tense-aspect combinations. These predicates are finite and not deficient in any sense with respect to the TAM markers.

(296) ʂiiʈay-kkɨ ɾaama-ne iʂʈam-aaɳi/aayiɾuɳɳu/aakum (Class I)
sita-DAT rama-ACC like-be.PRS/PST.PFV/FUT
'Sita likes/had liked/will like Rama.'

(297) ʂiiʈay-kkɨ ɾaama-ne iʂʈam-aak-aɳam/ aɳamaayiɾuɳɳu (Class II)
sita-DAT rama-ACC like-be-MOD/MOD.PST.PFV
'Sita should/should have liked Rama.'

(298) ʂiiʈay-kkɨ ɾanʈɨ makkaɭ unʈi/unʈaayiɾuɳɳu/unʈaakum (Class III)
sita-DAT two children have.PRS/have.PST/have.FUT
'Sita has/had/will have two children.'

The dative NPs are in a direct theta marking relation with the predicate like their nominative counterparts. As we suggested earlier, theta marking differences yield the interpretive differences of volition/no volition (or control over the event and lack of control) between the two nominals. In the minimal pair in (299) and (300), we see a single predicate alternating between the non-volitional, experiencer (with the dative case marking) and the true agent-like reading/a robust volitional reading (with the nominative case). The interpretive differences are triggered by the predicate via the theta-role and the inherent/structural case assigned on the subject. This also shows us that the two nominals are of equivalent structural status as subjects.

(299) eniy-kkɨ avaɭ-e iʂʈappeʈʈ-u [non-volition]
I-DAT she-ACC like-PST
'I liked her despite myself.'

(300) ɲaan avaɭ-e iʂʈappeʈʈ-u [volition]
I-NOM she-ACC like-PST
'I liked her.'

Butt (2006) argues that languages like Urdu employ case-alternations to show the difference in semantics and that datives can be interpreted as an agent with reduced control over the action in such languages(see Section 8.1). Malayalam employs the nominative and dative cases to mark such semantic differences.

In a pattern that diverges from most languages with dative subject predicates, (Sridhar, 1979; Ura, 2000; Davison, 2003, etc.) Malayalam (and Tamil) permit accusative objects with the Class I experiencer predicates[6] and Class II *veeɳam*.

(301) ʂiiʈay-kkɨ kiɭiy-e veeɳam
sita-DAT bird-ACC want
'Sita wants the bird.'

[6]The accusative case is overtly assigned only on [+ANIMATE] nouns as seen in Chapter 7.

(302) ɾaama-nɨ ɲiity-e iɻʈam-aaɳɨ
ram-DAT sita-ACC like-be.PRS
'Ram likes Sita.'

In these constructions, the dative nominal is the experiencer and the accusative object is the theme. However, constructions like (303) and (304) cannot be passivised, unlike the equivalent nominative-accusative constructions as shown in (305). The nominative case subject signalling greater volition/agency with an accusative object can be passivised (306) whereas the dative version signalling lack of volition cannot be passivised (307).

(303) a. ɲiiɲay-kkɨ kiɭiy-e ʋeeɳam [active]
sita-DAT bird-ACC want
'Sita wants the bird.'

b. *kiɭi ɲiiɲay-aal ʋeeɳʈa-ppeʈʈu [passive]
bird.NOM sita-INS want-PASS
'The bird was wanted by Sita.'

(304) a. ɾaama-nɨ ɲiiɲay-e iɻʈam aaɳɨ [active]
rama-DAT sita-ACC like.N be.PRS
'Rama likes Sita.'

b. *ɲiiɲa ɾaaman-aal iɻʈam-aaka-ppeʈʈu [passive]
Sita.NOM ram-INS like-be-PASS
'Sita was liked by Rama.'

(305) a. ɾaaman ɲiiɲa-ye ɲneehi-ccu [active]
ram.NOM sita-ACC love-PST
'Ram loved Sita.'

b. ɲiiɲa ɾaaman-aal ɲneehiykka-ppeʈʈu [passive]
sita.NOM ram-INS love-PASS
'Sita was loved by Ram.'

(306) a. ɾaaman ɲiitay-e iɻʈappeʈʈu [active]
rama.NOM sita-ACC like.PST
'Rama liked Sita.'

b. ɲiiɳa ɾaaman-aal iʂʈappeʈa-ppeʈʈu [passive]
sita.NOM rama-INS like-PASS
'Sita was liked by Rama.'

(307) a. ɾaama-nɨ ɲiitay-e iʂʈappeʈʈu [active]
rama-DAT sita-ACC like.PST
'Rama liked Sita.'

b. *ɲiiɳa ɾaaman-aal iʂʈappeʈa-ppeʈʈu [passive]
sita.NOM rama-INS like-PASS
'Sita was liked by Rama.'

The inability to create passives (despite the availability of the accusative case marked object) is because the external argument is not an *agent* argument and the event characterised by the predicate does not encode agency. In many sentences in English too, the lack of agentive force prevents passivisation of what are overtly transitive constructions (e.g. *marry*, *meet*, *fail* etc.) (de Mattia-Viviès, 2009).

One major test suggested by Belletti and Rizzi (1988, p. 325) to identify the canonical objects (structurally accusative marked) is to check whether they allow extraction of material. They show that in Italian, the accusative marked object of the verb *preoccupare* (309) does not allow extraction of material unlike the verb *temere* 'to fear' (308), and thus, the former assigns the morphological accusative to its object whereas the latter assigns the structural accusative.

(308) La ragazza di cui Gianni teme il padre.
The girl of whom Gianni fears the father.

(309) *La ragazza di cui Gianni preoccupa il padre.
The girl of whom Gianni worries the father.
(Belletti & Rizzi, 1988, p. 325)

Similarly, this test can be used to check whether the accusative case assigned by the dative experiencer predicates is structural or inherent. The adjective *ɲiramuɭɭa* 'coloured' in (310) can be extracted out of the NP as shown in (311).

(310) puuccay-kkɨ ɲiiɳay-uʈe [ɲiramuɭɭa kiɭiy-e]$_{DP}$ ʋeeɳam
cat-DAT sita-GEN coloured bird-ACC want.PRS
'The cat wants Sita's coloured bird.'

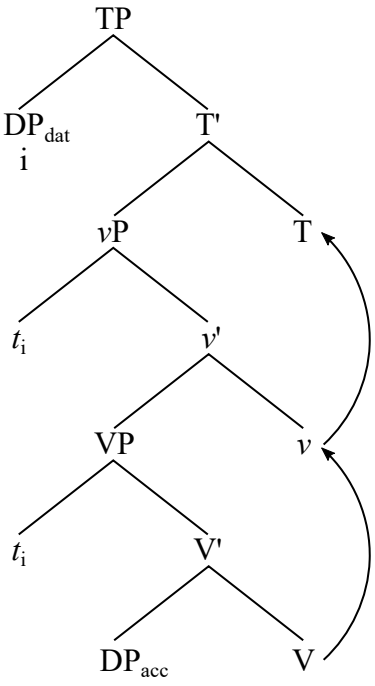
(311) ɲiramuɭɭa$_i$ ʂiitay-uʈe [t$_i$ kiɭiy-e] ʋeeɳʈa puucca
coloured sita-GEN bird-ACC want.ADJ cat
'The cat that wants Sita's coloured bird.'

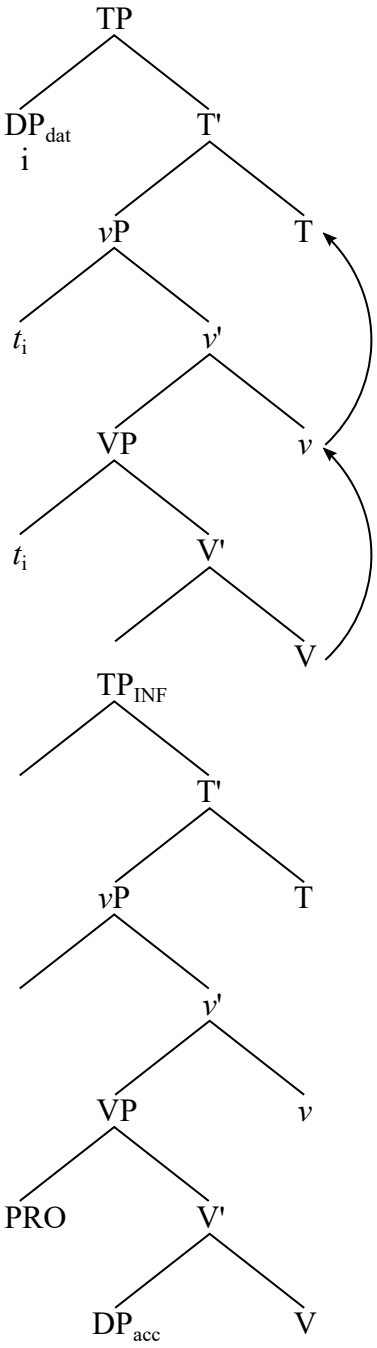
The fact that these objects allow extraction of material demonstrates that they are bearers of the structural accusative case and not an inherent one. Belletti and Rizzi (1988, p. 344) propose that a verb is a structural case assigner iff it has an external argument. In that case, the converse should also be true. Therefore, if a verb assigns a structural objective case, it has an external argument, which in our case is the dative NP. In other words, the dative NP is a true subject as much as the nominative NP.

We have demonstrated that indirect objects can never bind *ʈaan* which differentiates Class I-III dative NPs from dative indirect objects. Another point to be taken into account is the ungrammaticality of object control in dative subject constructions. The fact that object control is not permissible implies that dative subject NPs are treated on par with subjects and never with objects. Their selection is determined lexically by the predicate and they manifest the inherent case marking in the overt syntax. We demonstrated this through the use of empirical tests such as control and binding and the analysis of several constructions such as causatives, alternating predicates, accusative marking etc. Earlier accounts of the phenomenon attempted to treat the dative nominal as either not the subject (Jayaseelan, 2004) or as having the theta roles *goal* and *possession* uniformly making them similar to indirect objects (K. P. Mohanan & Mohanan, 1990). We have shown that these treatments a) make the wrong predictions and b) are not comprehensive. We look at how the dative subjects belonging to each class is licensed in the following.

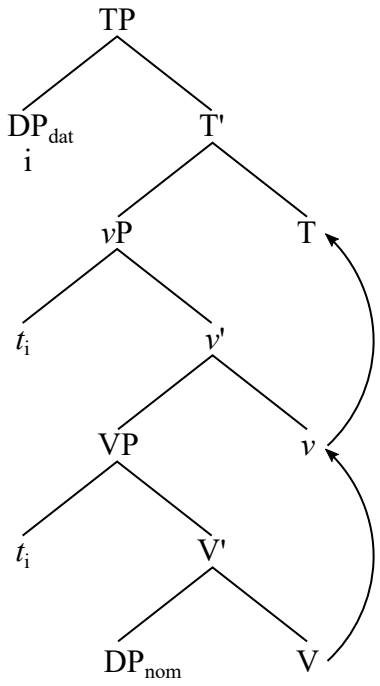
The subjecthood tests show that the dative NP behaves like a subject rather than an object. We adopt the split function framework proposed by Chomsky (1995) and refined by Ura (2000) to account for the structures of these sentences. We propose the following tree structures for the different types of dative subject constructions present in Malayalam.

(312) a. Class I dative predicates b. Class II dative predicates (modals *-aam* and *-aɳam*)





c. Class III dative nominals



As per our analysis, only Class I predicates can assign the accusative case to the object if transitive. In the above structure, intransitive dative experiencers like *vicakk-* 'be hungry' will not have an object.

The object DP is optional depending on whether the embedded verb is transitive or intransitive in Class II as in *piʈiykk-* 'catch' [*ʂiiɳay-kki*$_{(dat)}$ [PRO *kiʃi-ye*$_{(acc)}$ *piʈiykk*]$_{TP}$ *aɳam*]$_{TP}$ 'Sita

wants to catch the bird.' versus *uraŋŋ-* 'sleep', [*ʂiiʈay-kkɨ*$_{(dat)}$ [PRO *uraŋŋ*]$_{TP}$ *aɳam*]$_{TP}$ 'Sita wants to sleep.'

The Class III copular dative predicates cannot assign the accusative case to their objects. As a result, the object NP undergoes covert feature movement to T to get the nominative case. We assume that T and its *nominative* feature is weak and the checking of features takes place at the LF in Malayalam. The dative which is inherent and the structural accusative are assigned by V for the Class I constructions. By the time the υP shell is formed and T is introduced by *merge*, both the NPs are assigned cases and thus there are no more active probes in the derivation to check the nominative feature of T. It is the dative NP that raises to satisfy T's EPP feature. We argue that since the *nominative* feature of T is weak, it can be deleted at the LF if the case-filter is fulfilled. The EPP feature of T, however, is strong. In Class II constructions, the EPP is satisfied by the dative NP in the matrix clause and the nominative feature of the matrix T is deleted at the LF as in the Class I constructions since the case-filter conditions are already met. However, when there are constructions like (313), the nominative feature of the embedded infinitive clause is checked off by the matrix T, since it has a nominative feature left to be checked (which otherwise gets deleted if the case-filter is fulfilled).

(313) ʂiiʈay-kkɨ ɾaaman kiɭiy-e urakk-aɳam
sita-DAT rama.NOM bird-ACC sleep.INF-MOD
[ʂiiʈay-kkɨ [ɾaaman kiɭiy-e urakk-]$_{TP}$ aɳam]$_{TP}$
'Sita wants Rama to put the bird to sleep.'

In Class III constructions where the copula cannot assign the accusative case, the nominative object checks off the T's nominative feature at the LF with only feature movement, while the dative subject raises to satisfy T's EPP feature.

Ura (2000) proposes that the properties of control into adjuncts and binding of a (purely) subject-oriented reflexive result from *phi* and EPP feature checking relations with T, respectively. The agreement between a subject and its verb also involves *phi*-feature checking with T. Here, we account for the feature checking properties of Malayalam dative subjects NPs.

The dative NPs belonging to all the three classes can control into adjuncts, as shown in the examples below.

(314) [ʂiiʈay-kkɨ$_i$ [PRO$_i$ kiɭiy-e υaɭar-ʈʈ-aan]$_{TP}$ ʈoonn-i]$_{TP}$
sita-DAT bird-ACC grow-CAUS-INF feel-PST
'Sita felt like raising the bird.'

(315) [kuʈʈiy-kk$_{i}$ [PRO$_{i}$ paʈ^{h}ikk-aan]$_{TP}$ s̪kuuɭ-il pook-aɳam]$_{TP}$
child-DAT study-INF school-LOC go-MOD
'The child wants to go to school to study.'

(316) [kuʈʈiy-kk$_{i}$ [PRO$_{i}$ paʈ^{h}ikk-aan]$_{TP}$ pus̪t̪akam uɳʈ$_{i}$]$_{TP}$
child-DAT study-INF book have.PRS
'The child has books to study.'

However, it has to be noted that despite being able to control into adjuncts, the Class III dative nominal in (316) does not enter into phi-feature checking with T. It is the nominative NP that enters into direct phi-feature checking with T. This is not visible in Malayalam because of the absence of subject-verb agreement. Nonetheless, this is evident in Tamil where constructions involving Class III predicates are seen to agree with the nominative NP instead of the dative NP. Hence, if there is a nominative NP present in a dative subject construction, the phi-feature checking with T is carried out by the nominative NP and not the dative NP because of the former's need to check T's *nominative* feature and meet the case requirement.

Similarly, the dative NPs belonging to all the three classes can bind a purely subject-oriented reflexive *t̪aan* (317) – (319).

(317) ɾaama-n$_{i}$ t̪an-te$_{i}$ maat̪aa-pit̪aa-kkaɭ-e oorma-va-nnu
rama-DAT self-GEN mother-father-PL-ACC memory-come-PST
'Ram remembered his parents.'

(318) ɾaama-n$_{i}$ t̪an-te$_{i}$ maat̪aa-pit̪aa-kkaɭ-e kaaɳ-aɳam
rama-DAT self-GEN mother-father-PL-ACC see-want.MOD
'Ram wants to see his parents.'

(319) ɾaama-n$_{i}$ t̪an-te$_{i}$ pus̪t̪akam uɳʈ$_{i}$
ram-DAT self-GEN book have.PRS
'Ram has his own book.'

The ability to have control into adjuncts and bind a purely subject-oriented reflexive demonstrate the phi-feature checking relations between the dative subject and T, and consequently, establishes their subject properties. So, if we analyse the subjecthood properties of the three classes including accusativity, we can see that it is the Class I and II dative nominals that exhibit all the properties of a nominative subject. Class III dative nominals show subjecthood properties of binding and control but they cannot co-occur with accusative objects. Thus, we

conclude that Class III dative NPs are less-subject like compared to Class I and II dative NPs. In the following section, we bring in additional evidence from developmental data to substantiate our analysis and this difference between the predicates.

8.3 Acquisition of Dative Subjects

Acquisition studies have shown that children are aware of their target language's parameter settings quite early as with, for example, the **head directionality parameter** and their utterances are like the adult forms. Early Italian and English utterances are seen to be head-initial while early Japanese, Turkish, and so on are seen to be head-final (Guasti, 2002).

Hirsh-Pasek and Golinkoff. (1996) conducted a preferential looking paradigm study on 17-month-old children who were at the one-word stage in their language development to see if they relied on word order to comprehend active reversible sentences such as *Big Bird is washing Cookie Monster*. When presented with two scenes featuring the same characters but with the agent and patients reversed, the children preferred to look at the scene that matched with the heard utterance. Since these are reversible sentences, practical knowledge of the world could not have driven the children's preference. Children's preference for looking at the scene that matched the utterance heard, even before beginning to produce sentences, shows that children rely on word order to understand the grammatical roles.

Guasti (2002) proposes that children's sensitivity to word order and their correct assignment of head-direction show that they are sensitive to the distributional regularities of the ambient language. Additionally, children's knowledge of the target word order also indicates their tacit understanding of the grammatical roles such as subject and object. If children can successfully distinguish between the arguments in an inflectionally-deficient language like English, where they have to rely solely on word order, languages that have systematic case systems should enable the acquisition of the argument structure.

Guasti (1993) looks at the syntax of verbs in Italian child grammar by analysing the transcripts of three monolingual Italian children between the ages 1;8 to 2;7 from the CHILDES database and the data contributed by Ciprani et. al. (1989, as cited in, Guasti, 1993). Guasti's study shows that children employ both *agent* and *experiencer* subjects in their utterances. She reports that verb agreement with experiencer subjects is accurate in the majority of cases and that agreement errors are not confined to experiencer subjects alone. Further, in languages like Tamil

with dative subject constructions, where the dative occurs with an accusative object, the agreement is default while the dative subject shows subjecthood properties of binding the pronominal anaphor *ṯaan* and construing PRO in embedded clauses despite not agreeing with the verb. Thus, a child acquiring Tamil has to understand both the split properties of the subjecthood of dative nominals and the difference between real and default agreement in Tamil. Sarma (2014) demonstrates that children can distinguish between real and default agreement in Tamil, as discussed earlier in Chapter 5 (see Section 5.2).

There are widely different case marking systems such as ergative/split ergative, accusative, and split case-marking. Therefore, the task for children in acquiring these systems is a reasonably challenging one. Pinker (1984) proposes that children learn argument structure by relying directly on semantic notions such as agency. In fact, *agentivity bias* is hypothesised to help the child find the *subject* of a sentence and determine the requisite case marker. However, in languages like Malayalam and Tamil, where there are non-canonical subjects which are not agents, whether children make use of the *agentivity bias* to figure out the subject has to be checked. The acquisition of ergative languages can also help verify his hypothesis. If the idea of agentivity bias is correct, overgeneralisation errors involving the ergative marker ought to be expected in children acquiring ergative systems.

Narasimhan (2005) studies the acquisition of ergative case marking in Hindi by looking at the longitudinal data from three children between the ages 1;7 to 3;9. Ergative case marking in Hindi is associated with agentive participants of transitive *perfective* actions. Narasimhan reports that children are never seen to overextend the ergative marker *-ne* to agentive participants of imperfective or intransitive actions. In fact, children are seen to adhere to the tense-aspect restrictions involved in the assignment of the ergative marker *-ne*. This shows that Hindi acquiring children are not operating with an agentivity bias. However, there are omission errors involving the marker in obligatory contexts. Narasimhan proposes that the acquisition pattern of *-ne* can be attributed to the shaping of the appropriate construal of agency required for the ergative assignment by the distributional patterns in the input.

There are only a few studies on the acquisition of Dravidian languages. We have already discussed the acquisition of Tamil verbal inflections by Raghavendra and Leonard (1989) and Lakshmanan (2006), and the demonstration of full-competence in early Tamil by Sarma (1999, 2014), in Chapter 5. Lakshmi Bai (2004) and Usha Rani and Sailaja (2004) study the acquisition of Tamil and Telugu dative subjects, respectively, and demonstrate that children treat dative

subjects differently from dative marked non-subjects.

Lakshmi Bai (2004) studies the acquisition of dative subjects in Tamil by analysing the longitudinal data of three children Ramya (R), Deepa (D), and Chetan (C). The author notes that once the children begin to use the dative case in their productions in Tamil, there is a clear tendency on their part to differentiate dative marked subjects from other non-subject dative nominals. To maintain the difference, they appear to resort to different strategies. One such strategy has been to selectively use the dative marker to encode certain semantic roles preferentially over other semantic roles. For example, of the 25 tokens of dative forms in D's transcripts between the ages 1;5.3 and 1;7.3, 23 forms encoded the *benefactive* role (320) and only two the *directional* role (321).

(320) Mother: laalipaap vaanka-laam.
lollypop buy-HORT
'Let us buy lollypop.'
D: taataaya-ki.
grand.father-DAT
'For grandfather?'

(321) D: neene allaa iNTi-ki.
I Sarala house-DAT
'I want to go to Sarla's house.'

(Lakshmi Bai, 2004, p. 258)

Another strategy has been to mark only the subject datives and leaving the non-subject dative positions unmarked during different stages of their development. For example, the child C's early dative nominals were exclusively the experiencer subject.

(322) naa-ku pikka-tu.
I-DAT like-3N.SG
'I like it.'

(323) idi veenum naa-ku.
this wanted I-DAT
'I want this.'

(Lakshmi Bai, 2004, p. 256)

During the period when the child exclusively assigned the dative marker to the *experi-*

encer subjects, he did not overtly express nominals encoding the *recipient* (324) or *benefactive* (325) roles in his utterances.

(324) adi kaaTTu maa [missing nominal: enakku, I-DAT]
that show-IMP mother
'Mother, show that!'

(325) adi taa maa [missing nominal: enakku, I-DAT]
that give-IMP mother
'Mother, give me that!'

(Lakshmi Bai, 2004, p. 257)

Further, Lakshmi Bai reports that the children were seen to overextend the dative marker to other subject positions as well. According to her, the child's utterances, (326) and (327), are obligatory situations requiring a nominative subject signalled by the modal *-Num*.

(326) *ena-kku kraap veTT-i-kka-Num. (R, 2;6)
I-DAT crop cut-PTCP-REF-MOD
'I have to have a crew cut.'

(327) *paapaa-kku goregaon poo-ka-Num. (R, 2;6.15)
baby-DAT Goregaon go-PRS-MOD
'The baby has to go to Goregaon.'

(Lakshmi Bai, 2004, p. 253)

However, it has to be noted that these could very well be perfectly grammatical sentences where the child is expressing her wish to get these things done by employing the dative subject with the modal *-Num*. For example, (326) and (327) could mean 'I want to have a crew cut' and 'The baby wants to go to Goregaon', respectively.

Usha Rani and Sailaja (2004) look at both spontaneous speech and elicited speech data to analyse the acquisition of dative subjects in Telugu. The authors analyse the spontaneous speech samples of six children between the ages 2;0 and 2;9 that were taken from Sailaja (1989, as cited in, Usha Rani & Sailaja, 2004). They find that in the spontaneous data, children distinguish between indirect objects and dative subjects. Usha Rani and Sailaja have also performed imitation tasks with 36 children between the ages 2;0 and 5;0 to compare the productions with the longitudinal data. Children have been asked to imitate sentences including both dative and nominative subject constructions. The elicitation tasks performed by them show that only the

nominals in the subject position (both dative and nominative) were dropped and not objects, irrespective of the agreement; i.e., even when the nominative object phi-agrees with the verb (and presumably recoverable from context), it is the dative subject that is dropped. They argue that children treat dative and nominative subjects alike and pro-drop applies uniformly.

(328) naa-ku muuDu iLLu unn-aa-yi [target form]
I-DAT three houses be-PST-3PL.NH
Ø muuDu iLLu unn-aa-yi [child form]
three houses be-PST-3PL.NH
'I have three houses.'

(329) aame naa-ku uttaram raas-in-di [target form]
she I-DAT letter write-PST-3SG.NM
Ø naa-ku uttaram raas-in-di [child form]
I-DAT letter write
'She wrote me a letter.'

(Usha Rani & Sailaja, 2004, p. 217)

The works on the acquisition of Tamil and Telugu demonstrate that children understand non-canonical subject case assignment in their languages. These studies also offer additional evidence that subjecthood is privileged in certain ways and that the dative subject is treated in the same way as the nominative subject, offering further evidence of its subject status. Analyses such as Jayaseelan's (2004) or K. P. Mohanan and Mohanan's (1990) cannot account for the acquisition patterns of dative nominals occurring in the subject position. The Malayalam acquisition data we discuss will also demonstrate that children exhibit higher grammatical competence in assigning the non-canonical subject case which differentiates between the agentive volitional and non-volitional subjects. Dative subject assignment in Malayalam also helps in understanding whether the agentivity bias is operational in children acquiring Malayalam.

8.3.1 Dative Marking in the Data

We may recall that in the previous section we offered syntactic evidence for a) the lexical properties of certain predicates (Class, I, II and III predicates) that assign a non-canonical dative case to the subject argument and, b) we used empirical tests and analyses of various constructions to show that these are indeed subjects. When we analyse the production data of children, we

find that the patterns of use bear out these conclusions. The three different classes of dative predicates are reproduced in Table 8.2. The list of various dative assigning predicates used by the children are given in Appendices E and F.

Table 8.2 Classification of Malayalam dative subject predicates.

Classification	Predicate types	Examples
Class I	light verbs of N+V constructions	ɾaamani **koopam vaṉṉu** 'Rama became angry'
	experiencer predicates	ɾaamani **vicaṉṉu** 'Rama became hungry'
Class II	modals *-aam* and *-aɳam*	ɾaamani **kaɭikkaɳam** 'Rama wants to play'
		ɾaamani **kaɭikkaam** 'Rama can play'
Class III	copula *uɳɽi*	ɾaamani paɳam **uɳɽi** 'Rama has money'

Before we discuss dative subjects, we need an aside on children's use of the first person singular pronouns. The first person nominative pronoun is *ɲaan* while the accusative, dative and genitive are built on a suppletive stem *en-* with the identifiable accusative, dative and genitive cases (*enne*, *eniykki* and *ente*, respectively). Therefore, these suppletive forms have to be learned by the children. None of the other pronouns shows suppletion[7]. Subject A never uses first person pronouns and instead uses her own name when she refers to herself. On the other hand, H typically uses the first person pronoun form but instead of the suppletive stem *enn-e*, *eniy-kki* and *en-te* he uses *ɲaa-ne*, *ɲaa-ni*, and *ɲaa-nte* for the accusative, dative, and genitive, respectively, using *ɲaan* as the base stem.

In Table 8.3 we provide the numbers for the use of non-subject datives in the transcripts. In addition to the 93 tokens of indirect objects, A uses the dative in a variety of thematic roles such as *location*, *time*, and *purpose*, though the use is not prolific. H's use of non-subject datives is very limited in comparison. Both the children show competence in the use of the dative case across a variety of predicates and across a variety of argument positions as appropriate to the predicate as can be seen in the following.

Both children appear to be competent in the use of the dative subject with the different classes of predicates. In A's speech, subject datives are found in the data, from the very first recording, at the age of 1;9.14, although she omits the predicate *oɻiykkaɳam* 'want to pee' in that utterance.

[7]In Tamil, suppletion in pronouns is across-the-board, whereas in Malayalam, it is restricted to a few stems.

Table 8.3 Dative marking in non-subject positions in children's transcripts.

Grammatical roles	**A**	**H**
Indirect objects	93	7
Location	4	-
Purpose	1	1
Time	2	1

(330) **ammuu-ni̶** muuṯṯam (A, 1;9.14)
ammu-DAT pee
ammuu-ni̶ muuṯṯam-oɻiykk-aɳam [adult form]
ammu-DAT pee-pour-MOD
'Ammu wants to pee.'

As we might expect, the dative case is not limited to subjects. A uses the dative case marker in non-subject positions for *indirect objects* (331), *location* (332), *time* (333), and *purpose* (334), like adults do.

(331) **indirect object**
ʂuraaji-ni̶ ʊikk-aam baḍḍeecci-ne (A, 2;1.1)
suraaj-DAT sell-MOD bhadra-ACC
[ṉamu-kki̶] bʰaḍreecciy-e **ʂuraaji-ni̶** ʊilkk-aam [adult form]
[we-DAT] bhadra-ACC suraaj-DAT sell-MOD
'Let's sell Bhadra to Suraaj.'

(332) **location**
aikriim meeʈiykk-aan oɾu kaʈeen-te **ʈiɳɳay-kki̶** oɾu aana
icecream buy-INF one shop-GEN floor-DAT one elephant.NOM
ʊa-ṉṉ-eṉṉi̶ (A, 2;1.1)
come-PST-COMP
oɾu aana oɾu kaʈayu-ʈe ʈiɳɳay-kki̶ aiʂkriim meeʈiykk-aan
one elephant.NOM one shop-GEN floor-DAT icecream buy-INF
ʊa-ṉṉ-eṉṉi̶ [adult form]
come-PST-COMP
'An elephant came to the shop floor to buy ice cream.'

(333) **time**

ɲaaɭe **uccay-kki** ʋaɾ-um (A, 2;8.1)

tomorrow afternoon-DAT come-FUT

'will come tomorrow afternoon.'

(334) **purpose**

acca **ɟooliy-kki** pooy-i (A, 2;6.16)

dad.NOM job-DAT go-PST

'Dad went for work.'

H uses dative subjects as well, but somewhat less prolifically and his non-subject dative productions, are restricted to *indirect objects* (336) with just one instance signalling *time* (337) and *purpose* (338) each. The first instance of a dative subject attested in his transcripts is at the age of 2;5.2 (335).

(335) naa-ni kiʈʈ-iy-ee (H, 2;5.2)

I-DAT got-PST-EMPH

eni-kki [aʈi] kiʈʈ-iy-ee [adult form]

I-DAT that got-PST-EMPH

'I got it.'

(336) **indirect object**

iʈi pa-cci ʈaa **naa-ni** (H, 2;11.1)

this pluck-PTCP give.IMP I-DAT

eniy-kki iʈi pari-cci ʈaa [adult form]

I-DAT this pluck-PTCP give.IMP

'Pluck and give it to me.'

(337) **time**

Mother: ceeʈʈan eppo ʋaɾ-um?

brother.NOM when come-FUT

'When will brother come?'

H: **uccay-kki** (2;4.14)

afternoon-DAT

'In the afternoon.'

(338) **purpose**

ɟooliy-kki pooy-i appaaccan		(H, 2;8.2)
work-DAT go-PST grandpa.NOM		
appaaccan **ɟooliy-kki** pooy-i		[adult form]
grandpa.NOM work-DAT go-PST		
'Grandpa went to work.'		

The details of dative subject marking as they occur in the data are given in Table 8.4. Sample utterances of dative subject marking are shown in (339) - (341). A has greater number of dative marking instances compared to H. In general, H has been less loquacious than A throughout the period of the study. Both the children have fewer number of instances of Class III predicates.

Table 8.4 Children's dative subject production.

	A's Transcripts		**H's Transcripts**	
	Total Overt	**Topic-Drop**	**Total Overt**	**Topic-Drop**
Class I	43	32	9	20
Class II	59	203	7	55
Class III	4	2	3	4

(339) Class I

a. ammuu-ni ɖaayikk-uɳɳu (A, 2;1.1)
ammu-DAT be thirsty-PRS
ammuu-ni ɖaahiykk-uɳɳu [adult form]
ammu-DAT be thirsty-PRS
'Ammu is thirsty.'

b. naa-ni peeʈiy-aa (H, 2;9.1)
I-DAT fear-be.PRS
eniy-kki peeʈiy-aaɳi [adult form]
I-DAT fear-be.PRS
'I am scared.'

(340) Class II

a. ammuu-ni̵ booɭi̵ kaɭiykk-aɳam (A, 2;1.1)
ammu-DAT ball play-want.MOD
'Ammu wants to play ball.'

b. ɲaa-ni̵ miṯṯaayi meeɳam (H, 2;5.18)
I-DAT candy want.MOD
eniy-kki̵ miṯṯaayi ʋeeɳam [adult form]
I-DAT candy want.MOD
'I want candy.'

c. n̪ama-kki̵ ammumme-eṭe ʋiiṯṯ-ii pook-aam (A, 2;9.3)
we-DAT grandma-GEN house-LOC go-MOD
'Let's go to grandma's place.'

d. n̪ama-kki̵ iṯi̵ kaɭiykk-aam (H, 2;6.1)
we-DAT this play-MOD
'Let's play this.'

(341) Class III

a. aanay-kki̵ kaɳɳi̵ oɳʃi̵ (H, 2;11.18)
elephant-DAT eye have.PRS
'The elephant has eyes.'

b. kariy-kki̵ eɾiʋi̵ oɳʃi̵ (H, 2;5.18)
curry-DAT spiciness have.PRS
'The curry is spicy.' ('The curry has spiciness')

c. cambuu-ni̵ uʈuppi̵ oɳʃi̵ (A, 2;6.15)
shambhu-DAT dress have.PRS
'Shambhu has a dress.'

Class III dative predicates that mark *possession* are fewer when compared to Class I and II in both the children's speech production. A and H have only four and three such instances, respectively. This could be because of the lack of biuniqueness in the assignment of the dative marker since the genitive case markers *-nte* and *-uṭe* also signal *possession* in Malayalam. In Chapter 7, we saw that children are prolific in the use of the genitive. In A's transcripts, there are 375 instances of genitive marking while in H's transcripts, there are 227 instances. This could

delay the acquisition of the dative assignment rule to signal *possession*, in early Malayalam learners.

Malayalam is a topic-drop language, as we saw in Chapter 1. Consequently, arguments (both subjects and objects) may be dropped in sentences, given the discourse context. Not unexpectedly, both children drop arguments. It is well understood that children set the *pro*-drop and *topic*-drop parameter very early and this is visible in their productions very early. In pro-drop languages like Italian and Tamil, where the subject can be recovered from the agreement on the verb, and in topic-drop languages like Chinese and Japanese, where subjects and objects can be dropped when they can be recovered from the discourse, children's productions do not deviate from that of the adult native speakers (Wang et al., 1992; Rizzi, 1994). This in fact tells us that children's grammatical competence is adult-like (Guasti, 2002). Similarly, A's and H's speech productions show both overt and covert[8] instances of various arguments as summarised in Table 8.5.

Table 8.5 Overt and topic-drop instances of arguments in the data.

	A's Transcripts		**H's Transcripts**	
Argument	**Total Overt**	**Topic-Drop**	**Total Overt**	**Topic-Drop**
Nominative subject	469	419	217	234
Dative Subject	106	237	19	79
Object	447	179	192	155

A's topic drop instances exceed the number of overt dative subjects primarily for Class II

[8]Both the children's dative subject drop instances are appropriate to the context of utterance, as can be seen in the following examples. However, we are only focussing on the overt instances in this chapter.

1. A: (pointing to the picture of a ball)
   Ø ii booḷɨ ʋeeŋaʋ-eɲɲɨ ṯooɲɲ-uɲɲu (2;7.2)
   [ammuu-nɨ] ii booḷɨ ʋeeŋam-eɲɲɨ ṯooɲɲ-uɲɲu [adult form]
   [ammu-DAT] this ball want-COMP feel-PRS
   'Ammu wants that ball.'

2. Mother: koṯuki-ne ɲamu-kkɨ eɲṯɨ ceyy-aŋam?
   mosquito-ACC we-DAT what do.INF-MOD
   'What should we do to the mosquito?'
   H: Ø Ø a-ccɨ koll-aam (2;10.7)
   [ɲamu-kkɨ] [koṯuki-ne] aṭi-ccɨ koll-aam [adult form]
   [we-DAT] [mosquito-ACC] beat-PTCP kill.INF-MOD
   'Let's swat the mosquito.'

dative subject predicates (which includes the modals *-aam* and *-aŋam*) with 35 unique predicates. H topic-drops the dative subjects preferentially across all predicate classes. In comparison, A has 469 overt nominative subjects and 419 instances of dropped nominative subjects. H produces 217 overt nominative subjects and drops it 234 times. Accusative objects are also dropped frequently. Topic-drop of dative nominals in the subject position cannot be an indicator of a differential competence in the use of dative vs. nominative subjects, since the children apply topic drop across the board. The net numbers of dative subjects are fewer than the nominative subjects but this also follows form the fact that there are far fewer such predicates in the lexicon than the regular transitive or intransitive predicates.

Children's productions also reveal that they understand the subject properties of the dative nominals. Their productions can be used to demonstrate the following: a) binding the reflexive pronominal *ʇaan*; b) occurrence with accusative objects; c) subject control; and d) case-alternations with the same predicate to encode different theta roles. We discussed these properties as empirical findings that mark the *subjecthood* of these nominals. The presence of these properties in the children's productions tells us that children can differentiate between the dative subjects and dative non-subjects.

Sentences like (342) show us that dative subjects act as an antecedent to the phi-deficient pronominal anaphor *ʇaan* resulting in the EPP feature checking with T and also that the null subject of the embedded infinitival is controlled by the same dative nominal.

(342) ammuu-ni$_i$ ʇanne$_i$[9] kuɭiykk-aan patt-aṯṯ-illa (A, 2;10.3)
ammu-DAT self bathe-INF be able-FUT-NEG
[ammuu-ni$_i$ [PRO$_i$ ʇanne$_i$ kuɭiykk-aan]$_{TP}$ patt-aṯṯ-illa]$_{TP}$
'Ammu can't bathe by herself.'

The matrix subject *ammu* binds the pronominal *ʇaan* through PRO and the phi-features of the pronominal *ʇaan* are filled. This shows the child's tacit understanding of the underlying grammatical principles such as binding that licenses pronominal anaphors and the requisite feature checking relationship with T which in turn determines the *subjecthood* properties of the dative nominal. There are also dative subjects with accusative objects marking the former as the legitimate syntactic subject of the utterances as explained in Section 8.2.

[9] *ʇanne* is used as the short form for the reflexive *ʇanne-ʇaan* in colloquial speech, both in adult and children's speech.

(343) ammuu-n**ɨ** kaʈuʋay-e kaaɳ-uɲɲ-ill.alloo (A, 2;3.1)
ammu-DAT tiger-ACC see-PRS-NEG.EMPH
'The tiger is not visible to Ammu.'

(344) naa-n**ɨ** peeʈiy-aa ammeen-e (H, 2;8.16)
I-DAT fear-be.PRS mom-ACC
eniy-kk**ɨ** ammay-e peeʈiy-aaɳ**ɨ** [adult form]
I-DAT mom-ACC fear-be.PRS
'I am scared of mom.'

Children's production data also exemplify the control properties of dative subjects. In the sentences (345) and (346), the matrix subject, which is assigned the dative case, controls the null-subject of the adjunct clause and shows *subject-control*. In (345), the subject *ammu* which is lexically assigned dative by the predicate *veeɳam* 'want' (Class II) controls the null subject of the non-finite embedded verb *pook-* 'go'. In (346), the subject first person pronoun is dative-marked by the predicate *ariy-* 'know' and controls the null subject of the embedded predicate *caacc-* 'sleep'. By employing subject control into adjuncts, children reveal that they know it is the dative nominal that enters into phi-feature checking relationship with T. These two examples also demonstrate the children's understanding of the control properties of different predicates.

(345) ammuu-n**ɨ** kuuɭ-ii pooʋ-aan**ɨ** baag**ɨ** ʋeeɳam (A, 2;5.16)
ammu-DAT school-LOC go-INF bag want
[ammuu-n$ɨ_i$ [PRO$_i$ ʂkuuɭ-il pook-aan]$_{TP}$ baag**ɨ** ʋeeɳam]$_{TP}$ [adult form]
ammu-DAT school-LOC go-INF bag want
'Ammu wants a bag to go to school.'

(346) naan-**ɨ** caacc-aan ariy-aam (H, 2;10.2)
I-DAT sleep-INF know-MOD
[eniy-kk$ɨ_i$ [PRO$_i$ caacc-aan]$_{TP}$ ariy-aam]$_{TP}$ [adult form]
I-DAT sleep-INF know-MOD
'I know how to sleep.'

A also employs case-alternations that correspond to different semantic roles with the same predicate as in (347) and (348). Here, *kaaɳ-* can mean 'see' and can be used transitively with a nominative subject and accusative object (348). Alternatively, it can mean 'appear' or 'be visible' (347) and occurs with a dative subject and accusative object (Class I) .

(347) ammuu-**ni** kaʈuʋay-e **kaaɳ**-uɳɳ-ill.alloo (A, 2;3.1)
Ammu-**DAT** tiger-ACC **see**-PRS-NEG.EMPH
'Ammu is not able to see the tiger.'

(348) oɾu koɾaŋŋan caaʈi~caaʈ-i poo-y-appam oɾu aanay-e (A, 2;5.16)
one monkey.**NOM** jump~jump-PTCP go.PST-ADV one elephant
ka-ɳʇu
see-PST
'When the monkey jumped and went he saw an elephant.'

The utterances demonstrate the assignment of dative and nominative cases to the subjects by the same verb *kaaɳ-* 'see' given the degree of volition/ agentitvity/control of the subject. This shows a) the child's understanding of the semantic differences resulting from case and theta alternations and b) that she treats the dative and nominative nominals as both appropriate syntactic subjects.

We can see that the children show a high degree of competence with the use of dative subjects in accordance with the lexical properties of the particular predicates used. There is a high degree of correlation between the adult and child grammars. However, the production data is not without errors and these errors offer interesting insights into both the acquisition process and the syntactic properties of Malayalam, as we shall see in the next section.

8.4 Errors in the Production of Datives

The dative case inflection lacks biuniqueness since the case to theta role (*goal*, *location*, *purpose*, *experiencer*, *potential*, *possession* etc.) and case to grammatical role (*subject*, *indirect object*, *adjunct*) map is one-to-many. It is a lexical property of the predicates to select dative subjects (which must be learned) and the nominals that bear the dative case also vary in their lexical features across the predicate classes (Classes I requires [+ANIMATE] nouns while Classes II and III have no such restrictions). The formal complexity is high and thus, we should expect errors in the data.

No omission errors are attested in H's speech but H's dative productions themselves are far fewer when compared to A. The lack of prolificity could be attributed to the formal complexity of the inflectional morpheme. The unique dative marking error in H's transcripts is given in (349).

(349) ***ɲaa-ni*** ɳana-ɲɲ-ee (H, 2;8.16)
I-DAT get wet.PST-EMPH
ɲaan ɳana-ɲɲ-ee [expected form]
I.NOM get wet.PST-EMPH

The predicate is unaccusative and has a nominative subject. The child's production has a dative subject. This may be induced by the semantics of the verb where the subject is not an *agent* and is more experiencer-like. However, without further data it is not possible to aver that this is indeed the case.

A's productions have omission errors involving both subject and non-subject datives, as can be seen in Table 8.6. A omits the required/expected dative marker twice on an indirect object (with the same verb and the same noun but in different transcripts ((350) -(351)). Her overall accuracy in the use of the dative for indirect objects is 98%.

Table 8.6 A's omission errors vs correct uses.

Grammatical Role	**Omission**	**Correct Use**	**Total**
Indirect object	2	93 (98%)	95
Subject Class I	2	43 (96%)	45
Subject Class II	-	59 (100%)	59
Subject Class III	1	4 (80%)	5

(350) Mother: baɖɽeecciy-kki koʈukk-aɳ-oo?
bhadra-F-DAT give-MOD-Q
A: ***baɖɖeecc-i***_[10] kokk-aɳam (2;5.16)
bhadra-F.*0DAT give-MOD
bʱaɖɽaceecci-kki koʈukk-aɳam [expected form]
bhadra-F-DAT give-MOD
'Should give this to Bhadra.'

(351) ***baɖɖeecc-i***_ koolimiʈʈayi kokk-aɳam (A, 2;5.2)
bhadra-F.*0DAT lollipop give-MOD
bʱaɖɽaceecc-iy-kki koolimiʈʈayi koʈukk-aɳam [expected form]
bhadra-F-DAT lollipop give-MOD
'Should give the lollipop to Bhadra.'

A does not mark the dative on the subject on three occasions, twice for Class I predicates (352), (353) and once for a Class III predicate (354). A produces very few Class III datives (only 4 correct instances) compared to Classes I and II where her accuracy is very high[11].

(352) ***ammu_** paniy-aa (A, 2;2.2)
ammu.*0DAT fever-be.PRS
ammuu-n~~i~~ paniy-aa [expected form]
ammu-DAT fever-be.PRS
'Ammu has fever.'

(353) i̪~~i~~ ***ceecc-i_** iṯ-aan patt-uʊ-oo? (A, 2;10.3)
this.Ø sister-F.*0DAT put-INF be able-FUT-Q
ceecc-iy-kk~~i~~ i̪~~i~~ iṯ-aan patt-uʊ-oo? [expected form]
sister-F-DAT this.Ø put-INF be able-FUT-Q
'Can sister put this on?'

(354) ***acca_** kaʈuʊay-e peeʈi illa (A, 2;5.16)
dad.*0DAT tiger-ACC fear have.NEG.PRS
acca-n~~i~~ kaʈuʊay-e peeʈi illa [expected form]
dad-DAT tiger-ACC fear have.NEG.PRS
'Dad is not scared of tigers.'

In one instance, A incorrectly substitutes a dative nominal for an expected, possessive nominal (355).

(355) Mother: aaɾ-uṯe kaar-aa?
who-GEN car-be.PRS
'Whose car?'
A: ammuu-n~~i~~* (2;2.2)
ammu-*DAT
ammuu-nte [expected form]
ammu-GEN
'Ammu's'

[10]This could be a grammatical instance if A meant 'I want Bhadra to give..'. However, the context shows that it is not the case. Plus, she uses the dative marker correctly in every other instance involving the verb *koṭukk-* 'give'.

[11]To estimate error rates, we only count the overt instances of the dative subjects used with the error.

We may recall that Malayalam does have dative-subject possessives (Class III) as well. This may be an instance of overgeneralisation. There are a few omission errors involving the genitive case though the overall accuracy of genitive marking is 98%, as we saw in the previous chapter where the low error rate in production was attributed to the uniformity of the genitive marker.

This overextension of the dative marker to indicate possession is similar to the dative substitution for the locative in Tamil where both the markers can signal *location* (*goal* which is encoded by the dative is also an indicator of *location*). Lakshmi Bai (2004) reports that the child C's speech showed an overextension of dative to the locative. Examples (356)- (358) are from Lakshmi Bai's paper.

(356) (Expected: Dative)

ammaa danni viiTTu-*la poolam (C: 2;1)

mother Dhanni house-LOC go.HORT

'Mother, let us go to Dhanni's house.'

(357) (Expected: Locative)

bablu viiTT-*ukku aaNTi illa (C, 2;1.12)

Bablu house-DAT aunty is.not

'Aunty is not there in Bablu's house.'

(Lakshmi Bai, 2004, p. 257)

Lakshmi Bai reports a similar substitution error in R's transcripts as well.

(358) Father: rammiy-ooTa buk enka.

Rammi-GEN book where

'Where is Rammi's book?'

R: haalu-*kku irkku. (2;4.15)

hall-DAT is

'It is to the hall.'

(Lakshmi Bai, 2004, p. 254)

Likewise, A's substitution errors also target the nominative subject NPs. There are five sentences where she substitutes the dative case subject for the nominative subject.

(359) *ii **ceecc-iy-kk~~i~~** maɻay-a**ṯṯ~~i~~** **ṉ**ikk-uʊ-aaɾu**ṉṉ**u (A, 2;5.2)
this sister-F-*DAT rain-LOC stand-PTCP-be.PST
ii **ceecc-i** maɻay-a**ṯṯ~~i~~** **ṉ**ilkk-uʊ-aaɾu**ṉṉ**u [expected form]
this sister-F.NOM rain-LOC stand-PTCP-be.PST
'That girl was standing in the rain.'

(360) ***ammuu-n~~i~~** oo**ṯṯ**ooy-ii pooy-i (A, 2;5.16)
ammu-*DAT auto rickshaw-LOC go-PST
ammu oo**ṯṯ**ooy-il pooy-i [expected form]
ammu.NOM auto rickshaw-LOC go-PST
'Ammu went in an auto.'

(361) *oɾu **kokki-n~~i~~** pa-**ṉṉ~~i~~** pooy-i (A, 2;5.16)
one stork-*DAT fly-PTCP go-PST
oɾu **kokk~~i~~** para-**ṉṉ~~i~~** pooy-i [expected form]
one stork.NOM fly-PTCP go-PST
'A stork flew away.'

(362) ***a~~ṯ~~i-n~~i~~** kaʈiykk-um (A, 2;8.1)
that-*DAT bite-FUT
a~~ṯ~~i kaʈiykk-um [expected form]
that.NOM bite-FUT
'That will bite.'

(363) ***kuɲɲi-n~~i~~** ammuumm-e-em accan-e-m iɾu-**ṯṯ**-um (A, 2;10.3)
baby-*DAT grandma-ACC-CORD dad-ACC-CORD sit-TR-FUT
kuɲɲ~~i~~ ammuumm-e-em accan-e-m iɾu-**ṯṯ**-um [expected form]
baby.NOM grandma-ACC-CORD dad-ACC-CORD sit-TR-FUT
'Baby will make grandma and dad sit.'

Both the omission and the substitution errors continue till the last recording session at the age of 2;10.3. The child appears not to have completed her learning of the different rules involved in the dative assignment. One reason for such substitution errors in both Tamil and Malayalam may be the incomplete learning of the lexical properties of the predicates. As we have seen, there is no one-to-one correspondence between the case markers and the underlying theta roles of the arguments, in Malayalam. It is conceivable that this causes difficulties in

early language development, allowing for misanalysis of the verb meaning, which will also affect the surface (case) properties of the arguments. Of greater interest is the dative subject substitution for nominative subjects in A' speech ((359)-(363)). At the very least, these errors tell us that the child has two competing subject case assignment rules at her disposal. They also show us that the child has not yet completely acquired the dative subject assignment rule across predicates, grammatical roles, and theta roles. A closer look at the constructions involving the over-extension of the dative marker with the nominative assigning predicates reveals that these involve the predicates *aaŋɨ* (359)'to be' and *pook*- 'go' (360) - (361), and the rest in the future tense, marked by the morpheme -*um* (362) - (363). The predicate *aaŋɨ* also licenses the dative subject to signal an experiencer subject as in (364).

(364) **avaɭ-kkɨ** paampukaɭ-e peeʈi-**aaŋɨ**
she-DAT snakes-ACC fear-be.PRS
'She is afraid of snakes.'

Further, the verb *pook*- 'go' involves movement. This is one of the characteristic semantic feature of dative subject predicates when they instantiate the goal theta role that involves movement. Finally, the future tense marker -*um* is phonologically close to the modals -*ŋam* and -*aam*. Further, the predicates involving these substitution errors are not canonical agents. These could all be confounding factors that will increase the difficulty in the acquisition of the dative subject.

From the above discussion, we can see that the children's dative case assignment shows that they make a clear distinction between dative subjects and non-subjects. They use these forms correctly, across different predicates, and also understand the properties of control and binding. The overgeneralisation errors tell us that dative marking rule targets the subject position just like the nominatives and offer evidence for the view that both the nominative and dative NPs are treated on par as potential *subject* cases. Nonetheless, the continuing errors in dative marking reveal that the acquisition of the dative assignment rule is incomplete at the age of 2;10.3 for A, which is the last recording session. There are no omission errors attested in H's transcripts and the only incorrect instance of dative assignment includes an overgeneralisation error. The pattern of errors shows us that the lexical properties can be confounding to the child since the same verbs could use different case-markers and the same semantic properties may surface with different case forms. The formal complexity of the dative rule and the lack of one-to-one correspondence (biuniqueness) between the case-markers and the theta roles that

they encode, are the underlying reasons for the mismatches. The fact remains that for many of the predicates, the learning converges on full competence and there is ample evidence that the non-canonical subject phenomenon has already been successfully built into the developing grammar.

8.5 Conclusion

The analysis of the different classes of dative subject predicates shows that the dative nominals assigned by them behave as subjects and never as objects. Empirical tests included control, binding, causativisation, case-alternation, accusativity etc. Children's dative productions also reveal that they treat dative marked subjects differently from dative marked indirect objects. In fact, their dative productions reveal that they understand the subjecthood properties of dative nominals quite clearly and we offered evidence for the same from the transcripts. The fact that dative subjects are present in their speech as soon as they start combining words together to form sentences and that they over-extend the dative to nominative assigning predicates show that early Malayalam learners are not operating with an agentivity bias. Their constructions provide ample evidence to show that non-canonical subject marking is included in child grammar as it is in Malayalam adult grammar. However, the omission errors can be attributed to the high formal complexity of the marker assignment rule itself.

Chapter 9

Conclusions and Further Research

The main purposes of the thesis were a) to analyse the verbal and nominal inflectional morphology of Malayalam and identify the grammatical processes involved, and b) to study the effects of the grammatical properties on the developing grammars. We found that Malayalam is an agglutinative language where many of the inflections exhibit several key morphological properties that facilitate language learning. These include phonological salience of the morpheme for identification and segmentation, morphological transparency, biuniqueness, and high level of regularity and productivity. The inflectional system (as in many typologically similar languages) has many attributes that make it *acquirable* with relative ease by the young child. As also established by many others (Guasti, 2002; Demuth, 2001; Jusczyk et al., 1993; Mehler et al., 1988; Werker & Tees, 1984, etc.), competence in the phonological system develops early and enables word and syntactic learning. A system that uses the phonology reliably within the other levels of grammar, permits the children to acquire those levels relatively early and readily.

In our analysis of Malayalam TAM inflections, we find that all the inflectional morphemes other than a subset of the past tense allomorphs are across-the-board affixes and there are no omission errors involving any of them. The subset of past tense suffixes that prove the most difficult and where we find errors in the children's productions are exactly those whose use is determined by lexical and grammatical features of the verb rather than its phonology, specifically, its phonotactics. The past tense of denominal/deadjectival verbs and valency changed verbs fall into this class. The verbal inflectional errors attested in the data are overgeneralisation errors, but in the nominal inflectional morphology, we find both overgeneralisation errors and omission errors. Similar error patterns are found in the Malayalam cross-sectional data and the Malayalam-English bilingual data for both verbal and nominal inflections, as was discussed

in the preceding chapters.

Children's productions show phonological accommodations such as cluster simplification and phoneme substitutions, but early Malayalam utterances are not prosodically constrained. This is substantiated by the fact that no verbal inflections are omitted. In fact, during the course of the study, the children have shown a clear understanding of their target phonotactics as was demonstrated by the past tense marking patterns where the selection of the tense marker depended on considerations of what constitutes a legitimate syllable as well as the properties of the stem-final consonants. We do not know whether the children's target syllable structures were prosodically constrained prior to the period of study, though there are no real grounds to suggest that it may have been the case. The prosodic constraints model (Demuth, 1996, 2001) which tries to explain the production (and lack) of early functional categories through phonology cannot account for the current Malayalam data.

The acquisition of morphology has been addressed by two influential models, the *dual-route* model (Pinker, 1991; Prasada & Pinker, 1993; Pinker & Prince, 1994; Marcus, 1995) and the *single-route* model (Rumelhart & McClelland, 1986; Daugherty & Seidenberg, 1994; Elman et al., 1996). In the former, regular inflectional marking is rule-based, compositional, composed online, and not stored in the lexicon (evidenced by regular affixation on nonce-words in various production tasks, most famously with Berko's *wug*-test (Berko, 1958)). In contrast, irregular forms are stored in the lexicon as different clusters of phonological neighbourhoods that share associative links to one another (Ambridge & Lieven, 2011). In the single-route mechanism model, inflected forms are non-compositional, stored in an associative memory network like the irregular forms in the dual-route model, with phonological analogies contributing to the formation of generalisations. The child forms schemata of particular word forms that are phonologically similar as it encounters more and more inflected forms and the strength of each schema is dependent on the input frequency and the number of members within a particular schema.

In Malayalam, although there is no regular-irregular distinction among inflections, the past tense and plural morphology is sensitive to stem classes. We discussed overgeneralised past tense forms in Section 6.6.3. Recapitulating the discussion here, in brief, we saw that each stem class has exceptions. The overgeneralisation error *kolli* 'killed' in the acquisition data (instead of the exceptional *koṉṉu*) appears to be an overextension of the phonologically driven past tense formation rule using *-i*. Equally, *kolli* could also be the outcome of a schema where stem-final lateral-geminates choose the past tense marker *-i*, in analogy with *ṯalli* 'pushed', *colli* 'said' etc.

The overgeneralisation errors involving *-kk* derived, exceptional verbs, that choose *-n̪t̪u* instead of *-t̪u*, are **iriccu* and **n̪iccu*, instead of the expected (adult) *iruṉṉu* 'sat' and *n̪iṉṉu* 'stood', respectively. While **iriccu* can be accounted for by both the models since phonologically similar, high-frequency forms like *viriccu* 'spread' and *t̪iriccu* 'turned' exist in Malayalam, **n̪iccu* can only be explained by the dual-route model since no analogical forms exist. The form **niccu* can only be predicted by an overgeneralisation of the default tense formation rule of inflecting *-kk* derived verbs with *-t̪u*.

A third example of overgeneralisation that we saw involved the bare stem imperative and the sentence is reproduced below (see also Section 5.3.4).

(365) accay-ooʈɨ paɾay-aɳam mee-cc-ooɳʈɨ *ʋaɾɨ-eṉṉɨ (A, 2;5.16)
dad-SOC say-MOD buy-PTCP-with come.*IMP-COMP
acc-an-ooʈɨ paray-aɳam [baagɨ] meeʈi-cc-ooɳʈɨ [expected form]
dad-M-SOC say-MOD [bag] buy-PTCP-with
ʋaa-eṉṉɨ
come.IMP-COMP
'Should tell dad to buy (the bag).'

As we saw in Section 5.1, consonant-final Malayalam verb stems like *uuɾ-* 'remove' and *paaʈ-* 'sing' form the bare-imperatives *uuɾɨ* 'remove!' and *paaʈi* 'sing!', respectively, with the aid of the release vowel [ɨ][1]. The vowel is added in accordance with the phonotactic constraints which do not permit any consonant except [m, n, ɳ, l, ɭ, r, y] at the end of a word. On the other hand, the vowel-final stems do not change in the bare imperative forms (e.g. *poo* 'go!', *iɾi* 'sit!' etc. from *poo-* 'go', and *iɾi-* 'sit', respectively). The stems *vaɾ-* 'come' and *t̪aɾ-* 'give' form the irregular bare-imperatives *vaa* 'come!' and *t̪aa* 'give!' instead of the predicted regular forms **vaɾɨ* and **t̪aɾɨ*, patterning with the vowel-ending stems. The child overgeneralises the rule for sonorant ending CVC stems which are exceptional cases in Malayalam[2]. If the assumptions of the single-route model are correct, then, the child should produce the correct forms *vaa* 'come' and *t̪aa* 'give', especially since they are high-frequency forms which constantly occur in child-directed speech. The child is not expected to produce *vaɾɨ** since this is a form that the child will never hear in the input. Since errors of imperative and past tense formation are predicted

[1]This vowel is not a phoneme in Malayalam.

[2]These are the underived stems that choose the past tense marker *-n̪t̪u* instead of *-i*, which is the default past tense marker for underived stems, or *-t̪u*, which is the marker for the rest of the exceptional CVC stems (see Section 6.2).

by the dual-route model to be over-extensions of the default rules, Malayalam acquisition data offer support to the dual-route model of acquisition of inflectional morphology.

We have also seen that the children have no trouble in using the various inflectional categories. They never omit the tense markers in obligatory contexts and show no difficulties in producing finite sentences as was seen from the many examples discussed in Chapters 5 and 6. Also, there is no use of a non-finite verb where a finite verb is required. Argument structure and cases are also appropriately used. Additionally, unlike other phi-deficient languages, there is no identifiable Root Infinitival stage. Malayalam patterns in this regard with Italian, Tamil etc. In the discussion of the dative subjects, we showed that non-canonical subject marking is built into the child's grammar just as it is in the adult grammar. The children's use of dative subjects was correlated with their ability to handle the binding of the reflexive pronominal *ṯaan*, subject-control, accusative objects, and case alternation with the same predicate. Their productions demonstrate that children's grammatical knowledge includes the subject properties of both the dative and nominative subject nominals. Thus, early Malayalam data support the **Full Competence Hypothesis** which proposes that functional categories are available in the developing grammar from the very beginning (Poeppel & Wexler, 1993).

The dative subject marking acquisition data does not support Pinker's (1984) **agentivity bias** hypothesis since the dative marked experiencer (non-agentive) subjects are present in the children's utterances as soon as they begin to produce sentences, and dative experiencer case marking is over-extended to agentive verbs as well. Children clearly have a tacit understanding of the subject properties and treat the dative marked experiencer subjects on par with nominative marked agentive subjects.

The two main models which attempt to account for the production of early functional categories across languages are the **truncation model** (Rizzi, 1993) and the **ATOM** (Schütze & Wexler, 1996) which we discussed in Chapter 2. As we have already seen, there are no omission errors involving any of the verbal inflections. The subjects always employ the markers on verbs in all the obligatory contexts without exception. Hence, the truncation model which states that children can project and have access to all the functional categories but that they could optionally truncate functional projections, and the ATOM which states that the child can check the uninterpretable feature of either T or Agr can both account for the patterns in Malayalam. The former model predicts that early Malayalam learners will have access to all the functional projections, just like in Italian. Likewise, the ATOM can also account for the acquisition patterns

in the current Malayalam data, where the children's utterances adhere to the *minimise violations* constraint.

The omission errors in Malayalam primarily involve the omission of the accusative case marker, which is also observed in both the cross-sectional and the bilingual data. This cannot be explained by the syntactic models since the accusative is structurally assigned by the V in the lower VP which must obviously be available to the children since their speech productions show the presence of all the higher functional projections. The VP is also required for theta assignment and case-marking of the internal arguments. Therefore, the omission of accusatives most likely follows from a lack of mastery of the morphological rule of accusative assignment which is determined by the lexical feature of *animacy*, [±ANIMATE]. Not all objects are overtly marked and the child needs to learn the exact lexical feature restriction[3].

In the discussion of nominal inflections, we saw that different nominal inflections are subject to different modes of selection. Among the case markers, the two genitive case markers and the locative affix *-il* are uniformly and exceptionlessly applied across all nouns. The sociative *-ooṭi* and the locative *-aṯṯi* are lexically selected; the former by verbs of 'saying' such as *paray-* 'say', *cooḍiykk-* 'ask' etc., and the latter by place names ending in *-am* (e.g. *kooṯṯayam* 'Kottayam') and certain weather-related nouns like *maɻa* 'rain', *veyil* 'sun' etc. The accusative and the dative cases are contingent on the morphological feature of animacy. We observed in Chapter 7 that in both the children's data the accuracy of case-assignment is very high, never falling below 85%. This is not unexpected. However, it can be seen that the accuracy varies for different case-markers (Table 7.14 is reproduced here as Table 9.1).

Table 9.1 Omissions vs correct uses in the data.

	A's Transcripts		**H's Transcripts**	
Marker	**Correct**	**Omission**	**Correct**	**Omission**
Accusative	127 (86%)	21 (15%)	54 (89%)	7 (11%)
Dative[4]	318 (95%)	16 (5%)	43 (100%)	-
Genitive	380 (98%)	7 (2%)	129 (99%)	1 (1%)
Locative	290 (99%)	2 (1%)	113 (100%)	-
Plural *-maar*	32 (97%)	1 (5%)	3 (50 %)	3 (50%)

[3]We speculate that it is not a question of understanding animacy as an ontological category, but a question of making case marking contingent on the lexical feature.

While A uses the sociative 10 times with the same predicate *paray-* 'say', H does not use it at all. The locative marker *-aṯṯi* occurs 14 times in A's transcripts and nine times in H's transcripts. In both the children's data, the accusative marker has the lowest accuracy rate, and as was seen earlier in Section 7.4, its selection is based on the morphological feature of *animacy*, [±ANIMATE]. The Fisher's exact test shows that the difficulty in case marking is observed only when it is obligatory (overt) with animate nouns. As we see from the above, accuracy rates fall when lexical distinctions are necessary for inflectional selection.

Other than the past tense markers, all the verbal inflections in Malayalam are applied across the board, with only phonotactic accommodations at the morpheme boundary. The past tense markers, as we saw, undergo two different kinds of selection: a) phonology based, governed by syllabic weight constraints, and b) lexical feature based, *transitivity* [±TRANSITIVE]. While there are no omission errors in the verbal inflections, children do well when phonology drives the inflectional selection and not the lexical features, and likewise for the nominal inflections. Children produce nominal inflections that are selected by the phonological properties of the stem with greater success while being more error-prone and much less productive with nominal inflections that attend to lexical properties such as *animacy* or *human-ness*.

We already saw that the accusative morpheme is the least correctly marked case inflection and that the Fisher's exact test shows that the odds of occurrence of errors while overtly marking the accusative are higher and the results are statistically significant. The plural marking data corroborate this argument. Plural markers which are selected based on the three-way featural distinction for the features *humanness*, [+HUMAN] and *animacy*, [±ANIMATE] are the least used among the nominal inflections. H's accuracy for *-maar* is only 50%; it is the least correctly used and one of the least productive inflections in the data. Production of the plural markers across classes is restricted to a total of 58 instances involving 14 unique nominal stems in A's transcripts and 12 instances involving five unique nominal stems in H's transcripts. They constitute a mere 1.7% and 1.1% of A's and H's total inflections, respectively. The formal complexity of the rules, accusative case marking as determined by a single lexical feature (animacy) and plural, determined by a combination of two lexical features (humanness and animacy), delays the complete acquisition of the inflectional pieces and lowers frequency of usage to the extent that the acquisition of plural is even more delayed and incomplete compared to the accusative. We also showed that congruent with our data is the data on the acquisition of German plural

[4]Both subjects and indirect objects.

formation which is a complex inflectional process involving various suffixes, stem umlaut that is either pure or accompanies suffixation, knowledge of semantic features of the noun, and an overall consideration of the prosodic structure of the word. Their data also reveal that children have greater success in the production of prosodically driven plural formation compared to the other plural markers which are morphologically complex and require lexicalised learning (Kauschke et al., 2011).

There are different proposals that try to account for the omission errors and the relative ease or difficulty observed in the cross-linguistic acquisition of inflections. Slobin (1985) and Dressler (2012) discuss several factors such as phonological salience, morphological and semantic transparency, biuniqueness, productivity, and the relative positioning of the marker which have a discernible effect on the acquisition of inflections. On the other hand, Hyams (2008) attempts to account for the cross-linguistic patterns in inflectional acquisition by adopting a parameter-setting account rather than using factors that are external to the mental grammar.

Hyams (2008) in her parameter setting account (discussed in Chapter 2) claims that children's omission errors are the result of a *stem parameter*. For example, early English learners produce uninflected forms because such forms are deemed to be well-formed at the morphological level since the parameter is set to allow bare stems as well-formed words in English. In languages like Italian, a bare stem is ill-formed (by an alternate setting for the stem parameter) and results in ungrammaticality. In child Italian, therefore, uninflected verbs are rarely attested. According to Hyams, in English, the child does not have to learn any inflections since the verbal stem constitutes a well-formed word. Consequently, the acquisition of English inflections is delayed. In Italian, the stem parameter setting by disallowing bare stems expedites the inflectional acquisition. The stem parameter theory runs into problems because it cannot explain the co-occurrence of omission errors and correctly inflected forms during the same developmental period. When the child comes across the inflected forms, he/she should be able to figure out the members of an inflectional paradigm. This raises the question of when the stem parameter setting value can be altered to accommodate the inflected forms in a language like English. Hyams also uses cross-linguistic data such as overgeneralisation of overt affixes in place of zero-affixation in Russian and Serbo-Croatian, or the delay in the acquisition of bare imperatives in Turkish as evidence for the stem parameter setting that rules out bare stems. The child is then led to replacing all zero-affixes with overt ones or postponing the acquisition of bare stems altogether.

The Malayalam data discussed here do not show overgeneralisation in the context of zero-affixation despite Malayalam being a language where bare stems are ungrammatical in verbs except in imperatives. In fact, the opposite is observed, where the markers are omitted in obligatory contexts, especially in the case of the accusative marker and the plural markers. Young learners do not seem to have a problem in acquiring the bare imperative forms in Malayalam unlike the children learning Turkish[5]. In actual fact, bare imperatives occur more frequently in both the children's transcripts compared to the inflected imperative forms. While there are only 120 instances in total for the three imperative markers (*-aan*, *-ee*, and *-oo*) together, bare imperatives occur in 99 instances, in A's data, whereas in H's transcripts the token frequency of bare imperative forms (81 instances) exceeds that of the three imperative markers combined together (43 instances). Thus, while variation in acquisition may result from the properties of the grammar itself, the impact of various factors that affects affix-learning cannot be dismissed and seem to have a significant role in determining the course of acquisition.

The role of *mode of selection* (i.e., whether a marker is selected based on phonological or morphological features) in determining the relative ease/difficulty in the acquisition of inflectional markers is most visible in the acquisition data of nominal inflections as we saw. Dressler (2012) observes that markers occurring at the *periphery* of a word are identified earlier than markers that occur word-internally. The early emergence of case markers in agglutinating languages is attributed to this periphery effect (Stephany & Voeikova, 2009). In Malayalam, which is also an agglutinative, topic-drop language, all the case markers (except the sociative *-ooṭi* whose distribution is restricted to a few verbs) are in place early. However, it is the accusative that is the most challenging. Despite its relative positioning (at the periphery), phonological salience, morphological transparency, and high productivity, there is another factor that affects the developing grammar. We have argued with evidence from the acquisition data that *mode of selection* plays a crucial role in the developing grammars. Acquisition of German plurals also reveals that *mode of selection* is not limited to Malayalam, but it has a crucial role to play in the acquisition of inflections in general.

Associated with the mode of selection is the absence of RIs in early Malayalam. The current Malayalam data patterns with pro-drop languages like Italian and Tamil where the RI stage is not discernible. Malayalam is a topic drop language with no subject-verb agreement like Japanese and Chinese. Both Japanese and Chinese attest an RI analogue stage (Murasugi,

[5]Bare imperatives are readily acquired in Tamil as well (Sarma, 1999, 2014).

2015) which is absent in the present data. Murasugi (2015) notes that the Japanese RI analogue stage occurs in children between 1-2 years old and typically between the ages of 1;6 and 1;7. In the current Malayalam data the earliest recording is at 1;9. It is possible that the subjects are past the RI stage during the period of the present study. However, in Chinese and other European languages, the RI stage is seen after 2 years of age, which is not the case in either Malayalam or Japanese. In both the longitudinal and cross-sectional data, children are never seen to omit verbal inflections or produce optional infinitives. Guasti (1993, p. 13) elaborating on the ATOM (Schütze & Wexler, 1996) and the truncation model (Rizzi, 1993) has argued that the cause of the production of optional infinitives is the lack of mastery of the referential system associated with functional categories such as tense and determiners. All the verbal inflections in Malayalam apply either across the board or on the basis of the phonological features of the verb stems (except for the past tense in derived stems). Thus, children's creation of various inflected word-forms does not have to look beyond the phonology. This could be the reason for the lack of omission errors and optional infinitives in Malayalam. Assuming that Guasti is correct, we argue that in Malayalam, the exceptionless and phonology-centred selection of verbal inflections obviates children's difficulties in assigning inflectional markers resulting from a lack of mastery of the referential system associated with functional categories. In fact, whenever the selection is not a phonological one, children show difficulty in correctly and frequently employing the marker.

There are at least three open problems that require further work in the acquisition of Malayalam inflectional morphology. One is to understand the lexical property of animacy and its expression in the developing grammar. Animacy as a lexical feature is the determining factor in the assignment of accusative and dative case markers, gender, number, and pronouns not just in Malayalam but also in other Dravidian languages. Further research is needed to understand how animacy as a feature is built into early and adult grammars and what its cognitive status is for the child. Two, Malayalam syllable structure plays a central role in its morphology and in children's acquisition. While we touched upon aspects of the syllable structure in our discussions of the past tense, this was at best a limited discussion. A more complete discussion of syllable structure would have changed the entire focus of the thesis. In particular, with the heavy overlay of Sanskrit vocabulary which places conflicting and contradictory demands on the native (Dravidian) core phonotactics, it is important to understand how children navigate between the native Dravidian phonology and the Sanskrit phonology during different stages in development.

This may afford greater clarity to the study of acquisition and grammar of inflections. Three, Malayalam morphology is better understood in relation to its sister-language, Tamil. Many gaps in our understanding were actually filled when we considered parallel constructions in Tamil as well as the languages' joint history. A comparative study of the inflectional morphology of both languages in adult and child grammars will enable a more nuanced understanding of the theoretical questions.

With respect to methods, we focused on the longitudinal developmental data since there is very little systematic data available for Malayalam. We do realise that elicitation tasks can help bolster many of the arguments given here and will offer certain definite advantages, since we can pay attention to details of lexical items used, or particular inflectional markers, or the semantics of the predicate.

Finally, there are large areas of acquisition of Malayalam that have been left unexplored. Even in the few sample utterances that we presented as examples, the child is able to scramble arguments, bind anaphors, control, form questions etc. Questions on the acquisition of Malayalam syntax remain completely open for future work.

Appendix A

Linguistic Profiles of the Subjects' Parents

Table A.1 Language profile of A's parents.

	Mother	**Father**
Education	Post-Graduate	Post-Graduate
Monolingual/Bilingual	Monolingual	Monolingual
First language	Malayalam	Malayalam
Language/s spoken at home	Malayalam	Malayalam
Other familiar language/s	English	English

Table A.2 Language profile of H's parents.

	Mother	**Father**
Education	Graduate	Graduate
Monolingual/Bilingual	Monolingual	Monolingual
First language	Malayalam	Malayalam
Language/s spoken at home	Malayalam	Malayalam
Other familiar language/s	English	English

Appendix B

Examples of Discarded Items from the Transcripts

a) unknown word

aṯɨ ʊe-cc-appam **kuuɭin-ɨ** ʊekk-aan patt-aṯ ṯ -illa (A, 2;5.2)

that put-PST-ADV ?-DAT put-INF can-FUT-NEG

'[unknown] can't place it.'

b) filler

iṯɨ **pinne** oʈʈakam (A, 2;7.2)

this then camel

'This is a camel.'

c) repetition

iṯɨ **aaŋ-kocci** **aaŋ-kocci** (A, 2;8.1)

this male-child male-child

'This is a boy'

d) false start

ɲaŋŋ-aʈe ŋaŋŋaɭi-nte* aʋaʈe oru anci-kaɳɳ-an-puli oɳʧɨ (A, 2;10.3)

our-GEN our-GEN there one five-eye-M-leopard have.PRS

'There is a five-eyed leopard at our place.'

Appendix C

Inflectional Paradigms in the Transcripts

Table C.1 Number of inflectional paradigms in H's speech.

Age	2	3	4	5	6	7-10	Total
2;3.28	9	2	0	0	0	0	11
2;4.14	3	2	0	0	0	0	5
2;5.2	5	3	0	0	0	0	8
2;5.18	2	5	0	2	0	0	9
2;6.1	4	4	2	0	2	0	12
2;6.16	4	2	0	0	0	0	6
2;7.2	4	1	0	0	0	0	5
2;7.17	4	5	1	0	0	0	10
2;8.2	6	3	1	0	0	0	10
2;8.16	3	2	1	2	0	0	8
2;9.1	9	5	1	1	0	1	17
2;9.16	5	4	1	2	0	0	12
2;10.2	6	2	1	1	0	1	11
2;10.17	9	5	1	1	0	0	16
2;11.1	9	1	1	0	0	2	13
2;11.18	5	5	2	0	1	1	14
3;0.2	10	2	2	0	0	3	17
3;0.17	5	2	0	0	0	1	8

Table C.2 Number of inflectional paradigms in A's speech.

Age	**2**	**3**	**4**	**5**	**6**	**7-10**	**Total**
1;9.14	6	0	0	0	0	0	6
1;10.2	5	0	0	0	0	0	5
1;10.15	1	0	0	0	0	0	1
1;11.1	6	0	0	0	0	0	6
1;11.16	11	2	1	0	0	0	14
2;0.3	8	2	0	1	0	0	11
2;0.16	4	1	0	0	0	0	5
2;1.1	6	7	0	0	0	0	13
2;1.15	11	4	0	0	0	0	15
2;2.2	6	10	1	1	0	0	18
2;2.15	7	5	1	1	0	0	14
2;3.1	11	8	3	1	1	0	24
2;3.16	10	4	2	1	0	1	18
2;4.2	12	4	3	2	1	1	23
2;4.18	12	10	1	1	0	2	26
2;5.2	10	7	2	1	0	0	20
2;5.16	24	9	3	1	3	2	42
2;6.1	19	12	5	1	2	2	41
2;6.15	11	4	1	0	0	0	16
2;7.2	27	6	7	2	1	2	45
2;7.16	21	9	6	3	1	1	41
2;8.1	14	9	2	0	2	1	28
2;8.16	20	6	2	1	1	0	30
2;9.3	23	8	1	1	0	1	34
2;9.15	17	7	4	0	1	1	30
2;10.3	14	7	2	2	0	1	26

Appendix D

Past Tense Forms in the Transcripts

Table D.1 Past tense forms of the underived class in A's transcripts.

Child Form	Adult Form	Gloss
uuɾ-i	uuɾ-i	[remove-PST]
kiir-i	kiir-i	[tear-PST]
maar-i	maar-i	[move-PST]
pooy-i	pooy-i	[go-PST]
caaʈ-i	caaʈ-i	[jump-PST]
keer-i	keer-i	[enter-PST]
ooʈ-i	ooʈ-i	[run-PST]
ciik-i	ciik-i	[comb-PST]
kay-i	kaɻuk-i	[wash-PST]
eɻuʈ-i	eɻuʈ-i	[write-PST]
maaɲʈ-i	maaɲʈ-i	[scratch-PST]
ʈaʈʈ-i	ʈaʈʈ-i	[hit-PST]
poʈʈ-i	poʈʈ-i	[break-PST]
eɳɳ-i	eɳɳ-i	[count-PST]
oŋŋ-i	uraŋŋ-i	[sleep-PST]
ʈenn-i	ʈenn-i	[slip-PST]
ɲiiŋŋ-i	ɲiiŋŋ-i	[move-PST]
capp-i	capp-i	[suck-PST]

Past tense forms of the underived class in A's transcripts. (Continued from previous page)

Child Form	**Adult Form**	**Gloss**
kuʈʈ-i	kuʈʈ-i	[stab-PST]
ʈuuŋŋ-i	ʈuuŋŋ-i	[hang-PST]
ɳookk-i	ɳookk-i	[look-PST]
pamm-i	pamm-i	[cower-PST]
eʈʈ-i	eʈʈ-i	[arrive-PST]
ʈooɳɳ-i	ʈooɳɳ-i	[feel-PST]
camm-i	camm-i	[be ashamed-PST]
koʈʈ-i	koʈʈ-i	[nip-PST]
maʈaŋŋ-i	maʈaŋŋ-i	[fold-PST]
keʈʈ-i	keʈʈ-i	[tie-PST]
ciitt-i	ciitt-i	[spray-PST]
ʈupp-i	ʈupp-i	[spit-PST]
ca-ʈʈu	ca-ʈʈu	[die-PST]
cey-ʈu	cey-ʈu	[do-PST]
iʈ-ʈu	iʈ-ʈu	[put-PST]
ʋiʈ-ʈu	ʋiʈ-ʈu	[release-PST]
ʈa-ɳɳu	ʈa-ɳɳu	[give-PST]

Table D.2 Past tense forms of the derived class in A's transcripts.

Child Form	**Adult Form**	**Gloss**
kee-ʧu	kee-ʧu	[hear-PST]
koʅu-ʈʈu	koʅu-ʈʈu	[give-PST]
eʅu-ʈʈu	eʅu-ʈʈu	[take-PST]
oʅi-ccu	oʅi-ccu	[break-PST]
oɻi-ccu	oɻi-ccu	[empty-PST]
ayi-ccu	aɻi-ccu	[untie-PST]
ku-ccu	kuʅi-ccu	[drink-PST]
kuʅ-ccu	kuʅi-ccu	[bathe-PST]
maaŋ-ccu	ʋaaŋŋi-ccu	[buy-PST]

Past tense forms of the derived class in A's transcripts. (Continued from previous page)

Child Form	Adult Form	Gloss
ʊe-ccu	ʊe-ccu	[put-PST]
coy-ccu	coo$\underline{ɖ}$i-ccu	[ask-PST]
ci-ccu	ciɾi-ccu	[smile-PST]
kayi-ccu	ka$\underline{ɻ}$i-ccu	[eat-PST]
o$\underline{ɻ}$i-ccu	o$\underline{ɻ}$i-ccu	[hide-PST]
ʊaayi-ccu	ʊaayi-ccu	[read-PST]
muri-ccu	muri-ccu	[cut-PST]
mee$\underline{ɻ}$i-ccu	mee$\underline{ɻ}$i-ccu	[buy-PST]
pi-ccu	pi$\underline{ɻ}$i-ccu	[hold-PST]
pa$\underline{ɻ}$i-ccu	pa$\underline{ɻ}^{h}$i-ccu	[study-PST]
i$\underline{ɻ}$i-ccu	i$\underline{ɻ}$i-ccu	[punch-PST]
on$\underline{ɻ}$aakk-i	on$\underline{ɻ}$aakk-i	[make-PST]
maatt-i	maatt-i	[change-PST]
keett-i	keett-i	[enter-PST]
mukk-i	mukk-i	[sink-PST]
$\underline{ɳ}$uukk-i	$\underline{ɳ}$uukk-i	[hang-PST]
ma$\underline{ɻ}$akk-i	ma$\underline{ɻ}$akk-i	[fold-PST]
karakk-i	karakk-i	[spin-PST]
$\underline{ɳ}$iikk-i	$\underline{ɳ}$iikk-i	[move-PST]
pokk-i	pokk-i	[raise-PST]
pa$\underline{ʈʈ}$-i	para$\underline{ʈʈ}$-i	[fly-PST]
iɾu$\underline{ʈʈ}$-i	iɾu$\underline{ʈʈ}$-i	[seat-PST]
kee$\underline{ʈʈ}$-i	ke$\underline{ɻ}$a$\underline{ʈʈ}$-i	[lay-PST]
ʊa-$\underline{ɳɳ}$u	ʊa-$\underline{ɳɳ}$u	[come-PST]
$\underline{ʈ}$i-$\underline{ɳɳ}$u	$\underline{ʈ}$i-$\underline{ɳɳ}$u	[eat-PST]
ka-ŋ$\underline{ɻ}$u	ka-ŋ$\underline{ɻ}$u	[see-PST]
u-ŋ$\underline{ɻ}$u	u-ŋ$\underline{ɻ}$u	[dine-PST]
ko-ŋ$\underline{ɻ}$u	ko-ŋ$\underline{ɻ}$u	[hit-PST]
ʊii-ɳu	ʊii-ɳu	[fall-PST]

Past tense forms of the derived class in A's transcripts. (Continued from previous page)

Child Form	**Adult Form**	**Gloss**
pa-ɲɲu	para-ɲɲu	[say-PST]
kaɭa-ɲɲu	kaɭa-ɲɲu	[discard-PST]
n̪ana-ɲɲu	n̪ana-ɲɲu	[wet-PST]
kayi-ɲɲu	kaɻi-ɲɲu	[finish-PST]
oɳa-n̪n̪u	oɳa-n̪n̪u	[wake-PST]
iɾu-n̪n̪u	iɾu-n̪n̪u	[sit-PST]
n̪aʈa-n̪n̪u	n̪aʈa-n̪n̪u	[walk-PST]

Table D.3 Past tense forms of the underived class in H's transcripts.

Child Form	**Adult Form**	**Gloss**
pooy-i	pooy-i	[go-PST]
uuɾ-i	uuɾ-i	[remove-PST]
kiir-i	kiir-i	[tear-PST]
caaʈ-i	caaʈ-i	[jump-PST]
keer-i	keer-i	[enter-PST]
ooʈ-i	ooʈ-i	[run-PST]
eʈ̪ʈ̪-i	eʈ̪ʈ̪-i	[arrive-PST]
muʈʈ-i	muʈʈ-i	[bump-PST]
poʈʈ-i	poʈʈ-i	[break-PST]
n̪akk-i	n̪akk-i	[lick-PST]
puuʈʈ-i	puuʈʈ-i	[lock-PST]
eɳɳ-i	eɳɳ-i	[count-PST]
n̪ookk-i	n̪ookk-i	[look-PST]
t̪aʈʈ-i	t̪aʈʈ-i	[hit-PST]
koʈ̪ʈ̪-i	koʈ̪ʈ̪-i	[nip-PST]
poɭɭ-i	poɭɭ-i	[burn-PST]
keʈʈ-i	keʈʈ-i	[tie-PST]
kiʈʈ-i	kiʈʈ-i	[get-PST]
t̪uɭɭ-i	t̪uɭɭ-i	[caper-PST]

Past tense forms of the underived class in H's transcripts. (Continued from previous page)

Child Form	Adult Form	Gloss
n̪iiŋŋ-i	n̪iiŋŋ-i	[move-PST]
oŋŋ-i	uraŋŋ-i	[sleep-PST]
t̪a-n̪n̪u	t̪a-n̪n̪u	[give-PST]
pey-t̪u	pey-t̪u	[pour-PST]
iʈ-ʈu	iʈ-ʈu	[put-PST]
t̪oʈ-ʈu	t̪oʈ-ʈu	[touch-PST]

Table D.4 Past tense forms of the derived class in H's transcripts.

Child Form	Adult Form	Gloss
i-ccɨ	iʈi-ccu	[punch-PST]
ku-ccɨ	kuʈi-ccu	[drink-PST]
a-ccɨ	aʈa-ccu	[close-PST]
ka-ccɨ	kaɻi-ccu	[eat-PST]
oɭi-ccɨ	oɭi-ccu	[hide-PST]
ʋaa-ccɨ	ʋaayi-ccu	[read-PST]
ʋe-ccɨ	ʋe-ccu	[put-PST]
kaʈi-ccɨ	kaʈi-ccu	[bite-PST]
kut̪t̪-i-ʋe-ccɨ[1]	kut̪t̪-i-ʋe-ccu	[stab-PTCP-put-PST]
maaŋ-ccɨ	ʋaaŋŋi-ccu	[buy-PST]
aʈa-ccu	aʈa-ccu	[close-PST]
pa-ccu	pari-ccu	[pluck-PST]
poʈʈi-ccu	poʈʈi-ccu	[break-PST]
t̪i-ccu	t̪iɾ-ccu	[turn-PST]
oʈʈi-ccu	oʈʈi-ccu	[stick-PST]
kaanɻi-ccu	kaanɻi-ccu	[show-PST]
caaʈi-ccu	caaʈi-ccu	[jump-PST]
onʈaakk-i	onʈaakk-i	[make-PST]

[1] 'inject'

Past tense forms of the derived class in H's transcripts. (Continued from previous page)

Child Form	**Adult Form**	**Gloss**
maatt-i	maatt-i	[change-PST]
n̪iikk-i	n̪iikk-i	[move-PST]
keett-i	keett-i	[enter-PST]
aakk-i	aakk-i	[make-PST]
pat̪t̪-i	para t̪t̪-i	[fly-PST]
kama t̪t̪-i	kama ɻt̪t̪-i	[topple-PST]
ʋa-n̪n̪u	ʋa-n̪n̪u	[come-PST]
t̪i-n̪n̪u	t̪i-n̪n̪u	[eat-PST]
ka-ɲɲu	ka ɭa-ɲɲu	[discard-PST]
pa-ɲɲu	para-ɲɲu	[say-PST]
e-ɲɲu	eri-ɲɲu	[throw-PST]

Appendix E

Dative Subject Predicates in the Transcripts

Table E.1 Dative subject predicates employed by the children.

	A's Transcripts		H's Transcripts	
Predicate	**Overt**	**Topic-Drop**	**Overt**	**Topic-Drop**
ariy- 'know'	13	3	4	-
ţoonn- 'feel'	3	3	-	-
ḏaahiykk- 'be thirsty'	1	-	-	-
ʋaɾ- 'come'	5	5	1	3
patt- 'be able'	6	11	1	7
kiʈţ- 'get'	-	1	-	5
coriy- 'to scratch'	-	-	-	1
aaŋi 'be'	15	10	10	3
-aam [MOD]	21	54	1	32
-aŋam [MOD]	38	149	5	23
unʈi 'have'	2	2	3	4

Appendix F

Dative-Assigning Modal Predicates in the Transcripts

Table F.1 Unique predicates occurring with modals in H's transcripts.

Overt Subject Predicates	**Predicates Involving Topic-Drop**	
pariykk-aam [pluck-MOD]	pook-aam [go-MOD]	uuɾ-aɳam [remove-MOD]
kaɭiykk-aam [play-MOD]	pariykk-aam [pluck-MOD]	kaɻiykk-aɳam [eat-MOD]
aaʈ-aɳam [swing-MOD]	keer-aam [enter-MOD]	keer-aɳam [enter-MOD]
kaaɳ-aɳam [see-MOD]	kaɭiykk-aam [play-MOD]	eriy-aɳam [throw-MOD]
koɳʧipook-aɳam [take along-MOD]	ɲikk-aam [stand-MOD]	ɲookk-aɳam [look-MOD]
	iʈ-aam [put-MOD]	eʈukk-aɳam [take-MOD]
	ʋiʈ-aam [release-MOD]	kaaɳ-aɳam [see-MOD]
	uuɾ-aam [remove-MOD]	ʋekk-aɳam [put-MOD]
	oɳʧaakk-aam [make-MOD]	oɳʧaakk-aɳam [make-MOD]
	koll-aam [kill-MOD]	koll-aɳam [kill-MOD]
	kaɭiykk-aɳam [play-MOD]	iɾiykk-aɳam [sit-MOD]
	pook-aɳam [go-MOD]	iʈ-aɳam [put-MOD]
	maatt-aɳam [remove -MOD]	poʈʈiykk-aɳam [break-MOD]

Table F.2 Unique predicates occurring with modals in A's transcripts.

Overt Subject Predicates	**Predicates Involving Topic-Drop**	
kaɭiykk-aam [play-MOD]	kaɭiykk-aam [play-MOD]	ʋaraykk-aɳam [draw-MOD]
ʈinn-aam [eat-MOD]	ɲookk-aam [look-MOD]	iʈ-aɳam [put-MOD]
kaɭiykk-aɳam [play-MOD]	ʋekk-aam [put-MOD]	paʈiykk-aɳam [study-MOD]
maatt-aɳam [remove -MOD]	ʋiɭiykk-aam [call-MOD]	aʈiykk-aɳam [beat-MOD]
ʋiɭiykk-aɳam [call-MOD]	kuʈiykk-aam [drink-MOD]	kaaɳ-aɳam [see-MOD]
maattiʋekk-aɳam [keep aside-MOD]	oɭiccɨkaɭiykk-aam [play hide and seek-MOD]	kaɭaɲɲeekk-aam [discard-MOD]
	ʋekk-aɳam [put-MOD]	ʋiiɛ-aɳam [fan-MOD]
	ɨriykk-aɳam [sit-MOD]	ʈaɾ-aɳam [give-MOD]
	koʈukk-aam [give-MOD]	eʈukk-aɳam [take-MOD]
	ɨriykk-aam [sit-MOD]	kaɻuk-aɳam [wash-MOD]
	ʈorakk-aam [open-MOD]	koʈukk-aɳam [give-MOD]
	pook-aam [go-MOD]	ʈorakk-aɳam [open-MOD]
	kaaɳ-aam [see-MOD]	pook-aɳam [go-MOD]
	meeʈiykk-aam [buy-MOD]	ciɾiykk-aɳam [laugh-MOD]
	aaʈʈ-aam [rock-MOD]	ɲookk-aɳam [look-MOD]
	ʋecceekk-aam [keep put-MOD]	huu-ʋekk-aɳam [hoo-put-MOD]
	eraŋŋ-aɳam [get down-MOD]	muuʈʈam oɻiykk-aɳam [pee-MOD]

References

Aksu-Koç, A., & Ketrez, F. N. (2003). Early Verbal Morphology in Turkish: Emergence of Inflections. In D. Bittner & Kilani-Schoch (Eds.), *Development of Verb Inflection in First Language Acquisition: A Cross-Linguistic Perspective* (p. 27-49).

Ambridge, B., & Lieven, E. (2011). *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge University Press.

Asher, R. E. (1969). Classification of Tamil and Malayalam Verbs: A Comparative Study. In Agesthialingom & K. Raja (Eds.), *Dravidian Linguistics* (p. 249-268).

Asher, R. E., & Kumari, T. C. (1997). *Malayalam*. Routledge.

Belletti, A., & Rizzi, L. (1988). Psych Verbs and θ-theory. *Natural Language and Linguistic Theory*, *6*, 291-352.

Berko, J. (1958). The Child's Learning of English Morphology. *WORD*, *14*(2-3), 150-177.

Broselow, E., Chen, S.-I., & Huffman, M. (1997). Syllable Weight: Convergence of Phonology and Phonetics. *Phonology*, *14*(1), 47-82.

Brown, R. (1973). *A First Language*. Harvard University Press.

Burrow, T., & Emeneau, M. B. (1984). *A Dravidian Etymological Dictionary*. Oxford: Clarendon Press.

Butt, M. (2006). The Dative-Ergative Connection. In *Empirical Issues in Syntax and Semantics 6* (p. 69-92).

Bybee, J. (2001). *Phonology and Language Use*. Cambridge University Press.

Cazden, C. B. (1968). The Acquisition of Noun and Verb Inflections. *Child Development*, *39*(2), 433-448.

Census of India. (2011).

Chomsky, N. A. (1972). *Language and Mind*. Harcourt Brace Jovanovich.

Chomsky, N. A. (1981). *Lectures on Government and Binding*. Foris Publications.

Chomsky, N. A. (1986). *Knowledge of Language: Its Nature, Origin and Use*. Praeger.

Chomsky, N. A. (1993). A Minimalist Program for Linguistic Theory. *MIT Occasional Papers in Linguistics*, *1*, 1-67.

Chomsky, N. A. (1995). *The Minimalist Program*. MIT Press.

Christophe, A., & Dupoux, E. (1996). Bootstrapping Lexical Acquisition: The Role of Prosodic Structure. *The Linguistic Review*, *13*, 383-412.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler., J. (1994). Do Infants Perceive Word Boundaries? An Empirical Approach to the Bootstrapping Problem for Lexical Acquisition. *Journal of the Acoustical Society of America*, *95*, 1570-1580.

Christophe, A., Nespor, M., Guasti, M. T., & van Ooyen, B. (1997). Reflections on Phonological Bootstrapping: Its Role in Lexical and Syntactic Acquisition. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: A Special Issue of Language and Cognitive Processes.* Lawrence Erlbaum Associates.

Clancy, P. M. (1985). The Acquisition of Japanese. In D. A. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (Vol. 1: The Data, p. 373-524). Lawrence Erlbaum Associates.

Comrie, B. (1976). *Aspect*. Cambridge University Press.

Comrie, B. (1985). *Tense*. Cambridge University Press.

Daugherty, K., & Seidenberg, M. (1994). Beyond Rules and Exceptions: A Connectionist Approach to Inflectional Morphology. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The Reality of Linguistic Rules* (p. 353-388). John Benjamins Publishing Company.

Davison, A. (2003). Structural Case, Lexical Case, and the Verbal Projection. In V. Dayal & A. Mahajan (Eds.), *Clause Structure in South Asian Languages.* Kluwer Academic Publishers.

Davison, A. (2004). Non-Nominative Subjects in Hindi-Urdu: VP Structure and Case Parameters. In P. Bhaskararao & P. A. Subbarao (Eds.), *Non-Nominative Subjects* (Vol. 1, p. 245-268). John Benjamins Publishing Company.

de Mattia-Viviès, M. (2009). The Passive and the Notion of Transitivity. *Review of European Studies*, *1*(2), 94-109.

Demuth, K. (1996). The Prosodic Structure of Early Words. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax.* Lawrence Erlbaum Associates.

Demuth, K. (2001). Prosodic Constraints on Morphological Development. In J. Weissenborn & B. Höhle (Eds.), *Approaches to Bootstrapping: Phonological, Syntactic and Neuro-*

physiological Aspects of Early Language Acquisition (Vol. 2, p. 3-21). John Benjamins Publishing Company.

Demuth, K. (2011). The Acquisition of Phonology. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed., p. 571-595). Wiley-Blackwell.

Dressler, W. U. (2010). A Typological Approach to First Language Acquisition. In M. Kail & M. Hickmann (Eds.), *Language Acquisition Across Linguistic and Cognitive Systems* (p. 109-124). John Benjamins Publishing Company.

Dressler, W. U. (2012). On the Acquisition of Inflectional Morphology: Introduction. *Morphology*, *22*, 1-8.

Dressler, W. U., Kilani-Schoch, M., & Klampfer, S. (2003). How Does a Child Detect Morphology? Evidence from Production. In H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (p. 391-425). Mouton de Gruyter.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. Retrieved from https://wals.info/

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2019). *Ethnologue: Languages of the World* (22nd ed.). SIL International. Retrieved from https://www.ethnologue.com

ELAN (Version 5.2) [Computer software]. (2018). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press.

Fromkin, V., Rodman, R., & Hyams, N. (2011). *An Introduction to Language*. Wadsworth Cengage Learning.

Girija Devi, A. (1972). Acquisition of Language by a Malayali Child. In Subramoniam & Valentine (Eds.), *Proceedings of the First All India Conference of Dravidian Linguistics 1971* (p. 420-423).

Gleitman, L. R. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3-55.

Gleitman, L. R., & Wanner., E. (1982). Language Acquisition: The State of the State of the Art. In E. Wanner & L. R. Gleitman (Eds.), *Language Acquisition: The State of the Art* (p. 3-50). Cambridge University Press.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Guasti, M. T. (1993). Verb Syntax in Italian Child Grammar: Finite and Nonfinite Verbs. *Language Acquisition*, *3*(1), 1-40.

Guasti, M. T. (2002). *First Language Acquisition: The Growth of Grammar*. MIT Press.

Ham, W. (1998). *Phonetic and Phonological Aspects of Geminate Timing* (Unpublished doctoral dissertation). Cornell University.

Hamann, C. (2002). *From Syntax to Discourse: Pronominal Clitics, Null Subjects and Infinitives in Child Language*. Kluwer Academic Publishers.

Haspelmath, M., & Sims, A. D. (2010). *Understanding Morphology* (2nd ed.). Hodder Education.

Haznedar, B., & Ketrez, F. N. (Eds.). (2016). *The Acquisition of Turkish in Childhood*. John Benjamins Publishing Company.

Hirsh-Pasek, K., & Golinkoff., R. M. (1996). *The Origins of Grammar*. MIT Press.

Hyams, N. (2008). The Acquisition of Inflection: A Parameter-Setting Approach. *Language Acquisition*, *15*(3), 192-209.

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., van Valin, R. D., Murphy, B. W., & Khalak, H. G. (1996). A Positron Emission Tomographic Study of Regular and Irregular Verb Morphology in English. *Language*, *72*(3), 451-497.

Jayaseelan, K. A. (1999). The Dative Subject Construction in Parametric Syntax. In K. A. Jayaseelan (Ed.), *Parametric Studies in Malayalam Syntax* (p. 1-13). Allied Publishers Limited.

Jayaseelan, K. A. (2004). The Possessor-Experiencer Dative in Malayalam. In P. Bhaskararao & P. A. Subbarao (Eds.), *Non-Nominative Subjects* (Vol. 1, p. 227-244). John Benjamins Publishing Company.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402-420.

Kager, R. (1999). *Optimality Theory*. Cambridge University Press.

Kauschke, C., Kurth, A., & Domahs, U. (2011). Acquisition of German Noun Plurals in Typically Developing Children and Children with Specific Language Impairment. *Child Development Research*, *2011*, 17 pp. (Article ID 718925)

Ketrez, F. N., & Aksu-Koç, A. (2009). Early Nominal Morphology in Turkish: Emergence of Case and Number. In U. Stephany & M. D. Voeikova (Eds.), *Development of Nomi-*

nal Inflection in First Language Acquisition: A Cross-Linguistic Perspective (p. 15-48). Mouton de Gruyter.

Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge University Press.

Kunjan Pillai, S. (1965). *Kriyakaḷuṭe Ganavibhāgam ('Classification of Verbs')*. (Appendix, separately paginated (105 pp.))

Lakshmanan, U. (2006). Assessing Linguistic Competence: Verbal Inflection in Child Tamil. *Language Assessment Quarterly*, *3*(2), 171-205.

Lakshmi Bai, B. (2004). Acquisition of Dative Subject in Tamil. In P. Bhaskararao & P. A. Subbarao (Eds.), *Non-Nominative Subjects, Volume 1* (p. 245-268). John Benjamins Publishing Company.

Landau, I. (2002). *The Locative Syntax of Experiencers.* (Manuscript, Ben Gurion University)

Marcus, G. F. (1995). Children's Overregularization of English Plurals: A Quantitative Analysis. *Journal of Child Language*, *22*, 447-459.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), 1-178. (Serial No. 228)

Marslen-Wilson, W. D., & Tyler, L. K. (1997). Dissociating Types of Mental Computation. *Nature*, *387*, 592-594.

Mateo Pedro, P. (2015). *The Acquisition of Inflection in Q'anjob'al Maya*. John Benjamins Publishing Company.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A Precursor of Language Acquisition in Young Infants. *Cognition*, *29*(2), 143-178.

Mohanan, K. P. (1982). *Lexical Phonology* (Unpublished doctoral dissertation). MIT.

Mohanan, K. P. (1986). *The Theory of Lexical Phonology*. Reidel.

Mohanan, K. P., & Mohanan, T. (1984). Lexical Phonology of the Consonant System in Malayalam. *Linguistic Inquiry*, *15*, 572-602.

Mohanan, K. P., & Mohanan, T. (1990). Dative Subjects in Malayalam: Semantic Information in Syntax. In M. K. Verma & K. P. Mohanan (Eds.), *Experiencer Subjects in South Asian Languages* (p. 43-57). CSLI Publications.

Mohanan, T. (1989). Syllable Structure in Malayalam. *Linguistic Inquiry*, *20*(4), 589-625.

Murasugi, K. (2015). Root Infinitive Analogues in Child Chinese and Japanese. In A. Li, A. Simpson, & W.-T. D. Tsai (Eds.), *Chinese Syntax in a Cross-Linguistic Perspective*

(p. 375-398). Oxford University Press.

Narasimhan, B. (2005). Splitting the Notion of 'Agent': Case-Marking in Early Child Hindi. *Journal of Child Language*, *32*, 787-803.

Nizar, M. (2010). *Dative Subject Constructions in South-Dravidian Languages* (Unpublished master's thesis). University of California, Berkely.

Padmanabha Pillai, S. (1918). *Sabdatharavali*. National Book Stall.

Penke, M. (2012). *The Acquisition of Inflectional Morphology.* (Submitted in The Handbook of Morphology, Wiley-Blackwell)

Pesetsky, D. M. (1995). *Zero Syntax: Experiencers and Cascades*. MIT Press.

Piaget, J. (1971). *Psychology and Epistemology: Towards a Theory of Knowledge*. Grossman.

Pinker, S. (1984). *Language Learnability and Language Development*. Harvard University Press.

Pinker, S. (1991). Rules of Language. *Science*, *253*(5019), 530-535.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Basic Books.

Pinker, S., & Prince, A. (1994). Regular and Irregular Morphology and the Psychological Status of Rules of Grammar. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The Reality of Linguistic Rules.* John Benjamins Publishing Company.

Poeppel, D., & Wexler, K. (1993). The Full Competence Hypothesis of Clause Structure in Early German. *Language*, *69*(1), 1-33.

Prabodhachandran Nayar, V. R. (1972). *Malayalam Verbal Forms: A Phonetic and Phonological Study Supported by Experimental Findings*. Dravidian Linguistics Association.

Prabodhachandran Nayar, V. R. (1985). Language of the pāṭṭu literature. In K. M. P. Variar (Ed.), *History of Malayalam Language.* University of Madras.

Prasada, S., & Pinker, S. (1993). Generalisation of Regular and Irregular Morphological Patterns. *Language and Cognitive Processes*, *8*(1), 1-56.

Punnoose, R. (2010). *An Auditory and Acoustic Study of Liquids in Malayalam* (Unpublished doctoral dissertation). University of Newcastle upon Tyne.

Radford, A. (1990). *Syntactic Theory and the Acquisition of English syntax: The Nature of Early Child Grammar of English*. Blackwell.

Radford, A. (1997). *Syntax: A Minimalist Introduction*. Cambridge University Press.

Radford, A., Atkinson, M., Britain, D., Clahsen, H., & Spencer, A. (2009). *Linguistics: An Introduction* (2nd ed.). Cambridge University Press.

Raghavendra, P., & Leonard, L. (1989). The Acquisition of Agglutinating Languages: Converging Evidence from Tamil. *Journal of Memory and Language*, *16*, 313-322.

Rajaraja Varma, A. R. (1896). *Kerala Paaniniiyam*. National Book Stall.

Ravid, D., & Farah, R. (2009). Noun Plurals in Early Palestinian Arabic: A Longitudinal Case Study. In U. Stephany & M. D. Voeikova (Eds.), *Development of Nominal Inflection in First Language Acquisition: A Cross-Linguistic Perspective* (p. 411-432). Mouton de Gruyter.

Rizzi, L. (1993). Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives. *Language Acquisition*, *3*(4), 371-393.

Rizzi, L. (1994). Early Null Subjects and Root Null Subjects. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (p. 151-176). John Benjamins Publishing Company.

Rumelhart, D., & McClelland, J. L. (1986). On Learning the Past Tenses of English Verbs. Implicit Rules or Parallel Distributed Processing? In J. McClelland, D. Rumelhart, & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press.

Salustri, M., & Hyams, N. (2003). Is There an Analogue to the RI Stage in the Null Subject Languages? In B. Beachley, A. Brown, & F. Conlin (Eds.), *BUCLD 27: Proceedings of the 27th Annual Boston University Conference on Language Development* (Vol. 2, p. 692-703). Cascadilla Press.

Sarma, V. M. (1999). *Case, Agreement, and Word Order: Issues in the Syntax and Acquisition of Tamil* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Sarma, V. M. (2014). Issues in the Acquisition of Tamil Verb Morphology. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian Psycholinguistics* (p. 110-123). Cambridge University Press.

Schütze, C. T. (1997). *INFL in Child and Adult Language: Agreement, Case and Licensing* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Schütze, C. T., & Wexler, K. (1996). Subject Case Licensing and English Root Infinitives. In *Proceedings of the 20th Annual Boston University Conference on Language Development* (p. 670-81).

Shirai, Y. (1998). The Emergence of Tense-Aspect Morphology in Japanese: Universal Predisposition? *First Language*, *18*(54), 281-309.

Skinner, B. F. (1957). *Verbal Behaviour Action*. Copley Publishing Group.

Slobin, D. (Ed.). (1985). *The Cross-Linguistic Study of Language Acquisition* (Vol. 1: The Data). Lawrence Erlbaum Associates.

Song, J. Y., Sundara, M., & Demuth, K. (2009). Phonological Constraints on Children's Production of English Third Person Singular *-s*. *Journal of Speech, Language, and Hearing Research*, *52*(3), 623-642.

Sridhar, S. N. (1979). Dative Subjects and the Notion of Subject. *Lingua*, *49*, 99-125.

Stephany, U., & Voeikova, M. D. (2009). Introduction. In U. Stephany & M. D. Voeikova (Eds.), *Development of Nominal Inflection in First Language Acquisition: A Cross-Linguistic Perspective* (p. 1-14). Mouton de Gruyter.

Steriade, D. (1981). *On the Structure of Malayalam Onsets.*

Steriade, D. (1999). Alternatives to the Syllabic Interpretation of Consonantal Phonotactics. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Proceedings of LP'98 - Item Order in Language and Speech* (Vol. 1, p. 205-242). The Karolinum Press.

Steriade, D. (2002). The Syllable. In W. Frawley (Ed.), *Oxford Encyclopedia of Linguistics* (1st ed.). Oxford University Press.

Stoll, S. (2015). Inflectional Morphology in Language Acquisition. In M. Baerman (Ed.), *The Oxford Handbook of Inflection.* Oxford University Press.

Subbarao, K. V., & Bhaskararao., P. (2004). Non-Nominative Subjects in Telugu. In P. Bhaskararao & P. A. Subbarao (Eds.), *Non-Nominative Subjects* (Vol. 2, p. 161-196). John Benjamins Publishing Company.

Subrahmanyam, P. S. (2006). Dravidian Languages. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (2nd ed., p. 785-795). Elsevier.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Ura, H. (2000). *Checking Theory and Grammatical Functions in Universal Grammar*. Oxford University Press.

Usha Rani, A., & Sailaja, V. (2004). Acquisition of the Non-Nominative Subject in Telugu. In P. Bhaskararao & P. A. Subbarao (Eds.), *Non-Nominative Subjects* (Vol. 2, p. 209-222). John Benjamins Publishing Company.

Verma, M. K., & Mohanan, K. P. (1990). Introduction to the Experiencer Subject Construction. In M. K. Verma & K. P. Mohanan (Eds.), *Experiencer Subjects in South Asian Languages*

(p. 1-11). CSLI Publications.

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Wang, Q., Lillo-Martin, D., Best, C. T., & Levitt, A. (1992). Null Subject Versus Null Object: Some Evidence From the Acquisition of Chinese and English. *Language Acquisition*, *2*(3), 221-254.

Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development*, *7*(1), 49-63.

Wexler, K. (1994). Optional Infinitives, Head Movement and Economy of Derivation. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (p. 305-350). Cambridge University Press.

Wickremasinghe, D. M. d. Z., & Menon, T. N. (1927). *Malayalam Self-Taught by the Natural Method with Phonetic Pronunciation*. Marlborough.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06).*

Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford University Press.

Acknowledgements

Thanks are due to many people without whom this would not have been possible.

I do not have enough words to thank my supervisor Prof. Vaijayanthi Sarma for her guidance, constant support, patience, faith, time, and all her efforts. I have been really fortunate to have you as my guide and will always be grateful for the training you gave me in order to make me a linguist. I think I am still one in the making and I will always look up to you. I have had many teachers, but you are the one I consider my *guru* in the Indian sense of the term. No one could ask for a better guide and teacher.

My panel members, Prof. Milind Malshe, Prof. Mrinmoyi Kulkarni, and Prof. Pushpak Bhattacharya for their constructive comments and suggestions. Prof. Kiparsky, for lending a patient ear to listen to my work and giving me valuable insights regarding the same during IMM-18. K.V. Biju, for his time and efforts in easing me into the jargon of Malayalam traditional grammar during the initial stages of my PhD. My college teacher, N.P. Rajendran, for introducing me to the field of linguistics and for being the first person to put the idea of pursuing a career in linguistics in my head.

Prof. Priya Hosali, for making me believe in myself and helping me find the resolve to switch my career path. Had I not met you, I would never have made that decision. You are the most graceful human being I have ever seen. Thanks for inspiring me in so very many ways. I treasure each and every one of the conversations we had. I know you are the one person who would have truly loved to read my thesis had you been here in this world. This thesis is for you. I will always miss you.

My school teachers, Seema Nair, Indhu Sanal, and Reshmi Suresh, for all their love, concern, care, and encouragement. I believe that it is your blessings that have brought me all the good things in my life. Thanks for letting me into your lives and showering me with so much love.

Deepa Pankajavalli Amma, G.N. Gireesh, Jeena Shyamala, and P. D. Biju, for trusting

me with their children and allowing me to use their children's speech production data for my research. Ammu and Hari, for being the sweetest tiny tots. I thoroughly enjoyed my recording sessions with you guys.

My parents for their unconditional love, support, and most importantly for giving me wings. I owe almost everything in my life to you guys. My gran for showing her love and support in her own special and deeply affectionate ways. My younger brother for his uncanny ability to bring levity into every tense situation and for always believing in me. You are a terrific person and I love you lots.

Nidhi, for the long years of friendship. Thanks for always being a phone call away. I know I can always count on you.

The trio Pradeep Chellappan, Lathi Pradeep, and Pooja Pradeep for giving me a home away from home in Mumbai. I cannot thank you guys enough for your affection right from the moment I landed in this city as a nervous first-timer. For me, Mumbai has never been the same ever since you guys left.

Nisha, for always helping me in figuring out the nitty-gritty of academic and fellowship procedures. I know I have called you up a lot of times. Thanks for being such a great senior. Sindhu, Nawaf, Mashrur, Antonia, Varsha, and Punnya for being such delightful colleagues.

Rohit, for all the fun game nights, the shared love for talking in Malayalam movie dialogues, and for being a kindred spirit. Dawn and Jisha, for all the fun at the mess table. Thank you for finding me in the remote corner of the mess hall, lost in my own world. Since then, meals at the mess hall were never fun enough if you guys were not there. Arathi, for being my true blue PhD partner and for the unwavering support and love. You have always been there for me whenever I needed reassurance, despite being busy yourself. Thank you for your support in both my academic and personal life. Thanks a million for patiently reading the entire thesis. I am so glad we have had each other during these years. Resh, for being my fairy book satchel, a wonderful friend, and tea-partner. Thanks for all those books you have lent and gifted me. They kept my sanity going. I cherish all those cups of tea, conversations, code-names, and home-cooked dinners. Thanks for always having me over. Your apartment was one of my happy places in IITB. Dan, for all the laughs we shared, for all the fun we had, for all the movie nights, travels and trips, food-love, tea making sessions, and countless other big and small things. You made my return flights to Mumbai bearable and the hostel a place worth staying. I would not have been half as happy during my PhD years, had it not been for you. Most of the best memories

of my PhD life revolve around you. Thanks for giving me such colourful and memorable six years. For you and me, it is indeed the end of an era.

Finally, yet most importantly, to Navaneeth, for everything...

Gayathri G