Production of verbal morphology in heritage speakers of Tamil

by

Shankhalika Srikanth

B.A., University of Toronto, 2021

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Xʷsepsəm/Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day.

Supervisory Committee

Production of verbal morphology in heritage speakers of Tamil

by

Shankhalika Srikanth

B.A., University of Toronto, 2021

Supervisory Committee

Dr. Sonya Bird (School of Languages, Linguistics, and Cultures)
Supervisor

Dr. John Archibald (School of Languages, Linguistics, and Cultures)
Departmental Member

Abstract

This study investigates the differences in the verbal morphology of heritage Tamil speakers and L1 Tamil speakers with the ultimate goal of providing applicable pedagogical insights for heritage Tamil teachers and learners. This work is centred within my community of heritage Tamil speakers. Tamil is a Dravidian language that is dominant primarily in South Asia. Tamil heritage speakers, like heritage speakers of other languages, often face issues of social stigma, shame, and feelings of exclusion when speaking their heritage language, in addition to linguistic barriers because of the lack of pedagogical materials and research on heritage Tamil acquisition. Nevertheless, many adult Tamil heritage speakers have a strong motivation to learn their heritage language. The long-term goal of my research is to identify and implement strategies to support heritage Tamil learners who want to (further) develop their oral proficiency in the language, as well as to raise awareness for heritage Tamil being a legitimate variety of Tamil in its own right. For the purposes of this thesis, I limited my scope to verbal morphology, and the question I investigated was how heritage Tamil speakers and fluent L1 Tamil speakers differ in their production of verbal morphology. I used a combination of games and stories to elicit verb forms with a range of tense and person-number-gender (PNG) affixes, making sure to use methods that could be reapplied in the classroom as teaching tools. The heritage speakers I worked with used a combination of target-like forms as well as different strategies of overgeneralization, notably overmarking, to simplify irregular morphological paradigms that are present in L1 tense and PNG morphology. The findings from this study demonstrate that heritage Tamil speakers have a deep awareness of Tamil's complex verbal morphology, while also highlighting clear points of divergence between L1 Tamil and heritage Tamil grammars that parallel those seen in other heritage languages cross-linguistically and which can be useful for heritage language teachers in determining what types of forms to focus on in the classroom.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

/	in morpheme gloss, indicates the morpheme may be glossed as either of forms on either side of the slash
()	material within is optional
[]	material within is a morphological unit
-	morpheme boundary
—	hesitation
1	1st person
2	2nd person
3	3rd person
ADV	adverbial
COMP	completive
COND	conditional
DAT	dative
EMPH	emphatic
EUPH	euphonic
EXC	exclusive
F	feminine
FUT	future
HON	honorific
INC	inclusive
INF	infinitive
INTERR	interrogative
LOC	locative
M	masculine
MID	epenthetic material taken by middle stems
N	neuter
NTARG	non-target-like allomorph of a tense morpheme (where “target” refers to the baseline of an L1 speaker)
PL	plural
PNG	person-number-gender marker (unspecified)
PRS	present
PST	past
S	singular
SOC	sociative
STR	epenthetic material taken by strong stems
TARG	target-like allomorph of a tense morpheme (where “target” refers to the baseline of an L1 speaker)
TRANSCAUS	transitive-causative
UNC	unclear
VBLP	verbal participle

Acknowledgements

This thesis would not have been possible without the support of many people. First and foremost, thank you to the speakers I worked with: RB, AB, AS, and DS. Thank you for sharing your time, your insights, and your language with me.

Thank you to Erin Hashimoto, Akshaya Radhakrishnan, Sadhana Srinivasan, Dr Conor Quinn, and Anoushka Agrawal for your feedback at various stages in the research and writing process.

Thank you to Vasunthara Srinivasan, Vasumathi Srikanth, and Dr Ilamparithi Thirumarai Chelvan at the MACA West-End and VITS Tamil schools for giving me the opportunity to teach alongside you, and thank you to my students that I am continuously learning from.

Thank you to my family for putting up with all of my Tamil questions at the most random of times. Special thanks to my Appa – I know I've hijacked many dinner-time conversations to talk about Tamil grammar! Thank you to my Akka, I'm so grateful you came into my life when you did – so I could bug you with Tamil questions.

I could not have finished this thesis without the support and encouragement of my friends. Special thanks to Joanna Mo and Sabrina Yu for the seminars and always keeping me accountable. I can't begin to express my gratitude.

Thank you to my initial committee, Dr Megan Lukaniec and Dr Ewa Czaykowska-Higgins, for understanding what I wanted to do when I didn't understand it myself and making sure I built a solid foundation for my work. To my supervisory committee, Dr Sonya Bird and Dr John Archibald, who stepped in and supported me to the finish line: thank you for helping me make sense of how to organize my ideas and figure out how they fit into the bigger picture. Thank you to Dr Chandan Narayan for your insightful questions and feedback.

Finally, thank you to my Amma. All of this is possible only because of you.

Dedications

This thesis is dedicated to my community of heritage Tamil learners and speakers; and to my Amma, my first language teacher.

1 Introduction

This thesis set out to investigate the differences in the verbal morphology of heritage Tamil speakers and L1 Tamil speakers with the ultimate goal of providing applicable pedagogical insights for heritage Tamil teachers and learners. Heritage languages are languages that are taught to children by their family and tend to only be spoken in the home. For many heritage speakers, speaking their heritage language is key to their sense of cultural and familial identity. Unfortunately, many heritage speakers lose proficiency in their heritage language once they start attending school in the majority language (Polinsky, 2018b; Polinsky & Kagan, 2007).

This work is centred within my community of heritage Tamil speakers. Tamil is a Dravidian language that is dominant primarily in South Asia. Tamil heritage speakers, like heritage speakers of other languages, often face issues of social stigma, shame, and feelings of exclusion when speaking their heritage language, especially with fluent L1 speakers who can sometimes view heritage varieties as inferior. Heritage Tamil speakers also face linguistic barriers because of the lack of pedagogical materials and research on heritage Tamil acquisition. Nevertheless, many adult Tamil heritage speakers have a strong motivation to learn their heritage language.

The long-term goal of my research was to identify and implement strategies to support heritage Tamil learners who want to (further) develop their oral proficiency in the language, as well as validate heritage Tamil as a legitimate variety of the language. For the purposes of this thesis, I limited my scope to verbal morphology, and the question I investigated was how heritage Tamil speakers and fluent L1 Tamil speakers differ in their production of verbal morphology. I begin chapter 1 with a brief overview of the language community context as well as my own positionality. Chapter 2 then provides background on the types of variation

present in Tamil (2.1), Tamil verbal morphology (2.2), heritage language acquisition (2.3), and Tamil language acquisition (2.4) before introducing the research question (2.5). Chapter 3 describes the methodology and methods I used in this work. Chapter 4 details the results of my research, and chapter 5 discusses the implications of these results in the context of heritage language grammars cross-linguistically, as well as specifically the possibilities it opens up for heritage language learning and teaching. Heritage Tamil educators who are specifically interested in the implications of this work for their classrooms can skip to 5.4.

1.1 Tamil language context

Tamil is a Dravidian language primarily spoken in India and Sri Lanka, with large populations also in Malaysia, Singapore, and Mauritius. In addition to these homeland varieties of Tamil, there exist large Tamil diaspora populations in places such as Europe, North America, other parts of Africa, and Australia. Several of these diaspora populations can be classified as heritage speakers of Tamil; that is, they were raised in a home where Tamil was/is spoken, but the dominant language of society is not Tamil.

Heritage speakers can face stigma from first language speakers from the homeland, with their language being deemed “incomplete”, “incorrect”, or “ungrammatical”, and heritage speakers are often shamed by more fluent speakers in the community because they don’t speak “as fluently” as them. This judgement that comes from within a heritage speaker’s own community is often the biggest barrier to actually speaking and (re)learning the language. In this work I aim to encourage moving away from such beliefs. In analysing the language of heritage Tamil speakers, I do not endeavour to pass judgement on a supposedly “inferior” variety by comparing it to a hypothetical, idealized, “standard” Tamil. An important part of this is acknowledging the innumerable regional, social, and register-based varieties of Tamil that exist (discussed in 2.1).

While it is important to make sure not to pass value judgements on heritage speakers’ language, it is also important to respect the desire many heritage learners have to

improve their language skills, i.e. through developing literacy skills or increasing comfort and proficiency with oral communication. In my experience teaching both youth and adult learners, learners always express the desire to be able to communicate with fluent first language speakers. For some, they don't feel the need to *sound* like first language speakers; they just want to understand first language speakers and be understood in turn. For others, they want to sound like the first language speakers in their community, like the first language speakers who spoke to them when they were growing up. Crucially: both of these desires are valid, and both of these groups of learners have the right to be supported in reaching their goals. Learners who prioritize communicability over all else should not be forced to speak the more "standard" homeland varieties of the language, and learners who want to acquire these varieties should be given the tools to do so (Correa, 2011). Privileging homeland dialects over heritage varieties is prescriptivist; refusing to support learners who want to learn these varieties, out of a misplaced desire on the part of the teacher or researcher to validate learners' speech, is limiting and condescending.

In my study, I worked with Tamil speakers living in the diaspora (specifically, in the Greater Toronto Area). My focus was on colloquial Tamil (also known as spoken Tamil, or *koDuntamil;* the latter term is defined in 2.1.2).

1.2 Self-location

I am a first-generation settler of Indian Tamil origin and heritage. I was raised in the city of Mississauga on the traditional and present-day territory of the Haudenosaunee, the Mississaugas of the Credit, the Anishinaabek, the Métis, and the Wendat nations. It is treaty land covered by treaties 13-A, 14, 22, and 23 made between the Mississaugas and the Crown.

I grew up being a part of the immigrant Tamil community in this territory, known also as the Greater Toronto Area (GTA). At home, my family spoke to me in both Tamil and English, and so in my early years I was acquiring both as a first language. After starting

school, where English was the dominant language, my acquisition of Tamil dropped off and my fluency declined. Today I consider myself a heritage speaker of the language. My comprehension is much stronger than my production; I can understand spoken colloquial Tamil with little difficulty. I can communicate with Tamil speakers in an informal context, but will rely heavily on English to fill in vocabulary gaps. There are several complex morphological structures that I can comprehend but lack the confidence to produce. In these instances I am able to use circumlocution to get my point across, though my phrasing may sound clunky to the "native" ear.

I attended a Sunday language and culture school for a few years as a pre-teen, but I don't feel that this significantly improved my oral proficiency. I am now a teaching assistant at this school, alongside two other volunteer instructors, one of whom is my mother. The goal of the class is to expose the children to Tamil culture and language, with specific focuses on conversational skills and basic literacy. I am trying to draw on my own experiences as a former student of the school to improve how the class is run now. Volunteering at this school has made me much more aware of the range of proficiency seen in heritage speakers.

Volunteering with the school has also made me cognizant of the specific variety of Tamil that I speak, and how this relates to my identity within the larger Tamil community. Since I have only been exposed to the variety spoken by my parents, this is the variety that I have acquired. Unlike individuals growing up in Tamil societies, I was not exposed to other dialects and registers. As such, similar to other heritage speakers of Tamil, I gained little to no sociolinguistic awareness of these other varieties, or even an understanding of what aspects of my speech function as overt sociolinguistic markers of my background (Renganathan, 2015).

My parents speak a variety of Tamil that is largely marked by region (North Arcot/Chennai) and caste, and this is the variety I have acquired. I cannot claim as complete

an understanding as I would like of casteism and the ways in which it surfaces through language. Nevertheless, I am aware of the privilege I was born with in being born to parents with caste-privilege and of how my Tamil reflects this. I discuss casteism and in particular its relationship to language in more detail in 2.1.1.

In the context of heritage language teaching, it is common for learners to come from a variety of language backgrounds. Because – like me – most learners are only exposed to the variety spoken at home, it is common to not have a sociolinguistic awareness of other varieties, as in my own case. However, it is important for the teacher to have this awareness and be inclusive of the varieties spoken by their students, rather than imposing their variety as the standard (Boon & Polinsky, 2015; Ducar, 2008; Duff et al., 2017; Polinsky & Kagan, 2007; Sanchez-Munoz, 2016). In the context of my work, this principle is doubly important, as my dialect of Tamil reflects my caste and thereby my caste-privilege, and so can potentially be an assertion of a power imbalance. It was important to me that I didn't unintentionally perpetuate casteism in my work and cause harm in this way. I discuss how I tried to do this in more detail in 3.1.

2 Background

2.1 Tamil language varieties

Teaching heritage language learners always presents a challenge because heritage language learners are a hugely diverse population, with their individual language proficiencies and competencies largely influenced by their unique language background**.** In the specific context of Tamil, one of the biggest challenges heritage learners face is knowing how to identify and navigate between commonly used varieties of Tamil, which occur along three axes: region, religion/caste, and diglossia/register (Annamalai, 2019). In Tamil linguistics/literature, these variations are described using the terms *vattara vazhakku* and *pothu tamil* to describe the regional and social varieties, and *koDuntamil* and *centamil* to describe the different diglossic registers. We first define and compare *pothu tamil* and *vattara vazhakku*.

2.1.1 *Pothu tamil* and *vattara vazhakku*

The native term for a regional and/or social dialect is *vattara vazhakku*, loosely translated as "local dialect". Its counterpart is *pothu tamil,*[1] "general Tamil", which is the variety of Tamil that emerges after a speaker has removed all regional and social markers from their speech. Where such a "neutral" alternative for a word does not exist, the speaker substitutes a *centamil* form (discussed in 2.1.2). As such, "mainstream" Tamil refers not to a specific *vattara vazhakku*, but to a somewhat artificial (and variable) variety that aims to minimize obvious markers of regional or social background (Annamalai, 2019; Schiffman et al., 2009).

The major regional dialects of Tamil include Indian Tamil, Sri Lankan/Eelam Tamil, and Singaporean Tamil. Within the Indian Tamil dialect spoken in the state of Tamil Nadu, there are numerous regional varieties including Chennai Tamil (which I speak), Coimbatore

[1] வட்டார வழக்கு and பொதுத் தமிழ் respectively.

Tamil, and Palakkad Tamil, to name just a handful. Regional variation surfaces in lexical items, phonology, and sometimes morphosyntactic structures.

Social/religious dialects of Tamil are delineated primarily by caste. Caste, in particular Brahminism, is a type of systemic discrimination that continues to inflict harm in Indian society, both in the homeland and in diaspora (Taku, 2020). It is a social hierarchy that is assigned at birth, where individuals are "born" into the caste of their family. Historically it was very common, (and in many communities still is common) to segregate based on caste. Relationships between members of different castes continue to be stigmatized in certain communities, and individuals belonging to lower castes continue to face violence and oppression perpetuated by those belonging to a more privileged caste. Casteism is often discussed in terms of Brahminism. "Brahmin" is the caste that sits at the top of the social hierarchy and holds the most social power (in most regions, including the state of Tamil Nadu), and so in some contexts casteism and Brahminism are interchangeable terms. In many ways the hate and prejudice that fuels this systemic oppression can be compared to white supremacy in the west (though this is definitely a simplification). While many individuals are now trying to move away from caste and will disavow identifying with their own caste or interacting with others based on theirs, the privilege or lack thereof that comes from being born into a certain caste remains deeply entrenched in South Asian communities; indeed, most heritage Tamil speakers will grow up speaking a variety of Tamil that is marked by the caste(s) of their parents, often unknowingly. As such, the reality and impacts of caste, specifically in how this surfaces in heritage Tamil speech, needs to be acknowledged when doing language work.[2]

[2] I am by no means an expert on the history and realities of caste-based violence and discrimination, and so apologize and take ownership of any mistakes I may have made in this work as I continue to try to learn about my own caste privilege. For those who are similarly interested, an essential introduction to anti-caste literature is "Annihilation of Caste" by B.R. Ambedkar. The works of other important anti-caste reformers such as 'Periyar' E. V. Ramasamy (Tamil Nadu), and Savitribai and Jyotirao Phule (Maharashtra) may also be of interest. More specific to the Tamil language context, the final essay in Annamalai (2011a) titled "Political Role" provides a concise but illuminating overview and analysis of the interplay between caste, Brahminism, Dravidian identity, and Tamil language and identity as it has historically played out in the sphere of Tamil Nadu state politics.

Caste is a major sociolinguistic marker in varieties of Tamil. Broadly, the two major varieties have been classified in the literature as the "Brahmin" dialect and the "non-Brahmin" or "common" dialect (Annamalai, 2011a, 2019; Bright, 1990a; Krishnamurti, 2003), though there are subclassifications within these categories as well. Most of the variation is lexical, with some (morpho)phonological variation as well. Brahmin dialects often replace native words with Sanskrit loanwords (and preserve Sanskrit phonology when doing so), and the two dialects vary the shape of some select morphemes in a way that isn't entirely tied to a systematic phonemic correspondence (Bright, 1990, p. 15-16). Note that the Brahmin variety of Tamil is not recognized as the "standard" form of the language, nor is it the variety taught in schools, privileged in public institutions, or common in popular media. In fact, it is a more marked way of speaking and would not be recognized as the norm (Annamalai, 2011a, p. 67). However, speaking this variety does mark the speaker as someone belonging to a caste with added privilege and power.

Regional varieties intersect with varieties marked by caste. For example, the "common" dialect as spoken in Chennai is distinct from the "common" dialect spoken in Thanjavur, which is itself distinct from the "Brahmin" dialect spoken in Thanjavur (Annamalai, 2019; Renganathan, 2015). Nonetheless, there are broad similarities across regional dialects and across caste-based dialects. For the purposes of this thesis, I won't go into detail about how these two sociolinguistic markers interact, instead referring to any socio-regional dialect as a *vattara vazhakku*.

The counterpart to a *vattara vazhakku* is *pothu tamil*. This is what is spoken in the "mainstream" and is a variety of Tamil that deliberately doesn't include any regional or social dialect markers. This is a variety that first language speakers growing up in a Tamil-speaking area (i.e., non-heritage speakers) are exposed to and familiar with. This is because they are likely to hear and use *pothu tamil* in social situations when interacting with speakers from different regional and social backgrounds to their own. A defining

characteristic of heritage speakers, on the other hand, is that they are exposed to a very limited range of language variation: they only hear the varieties spoken by their family in the home. As such, there is unlikely to be much necessity or opportunity to hear and use *pothu tamil*, or indeed any other *vattara vazhakku* than the one(s) spoken by their family.

It is common for heritage speakers to assume that the variety of the language they grew up with is the "default"; after all, what other evidence would there be to the contrary? As a result, it can be harder to understand other varieties, and a heritage speaker's use of their own *vattara vazhakku* in contexts where everyone else is speaking *pothu tamil* may mark them to first language speakers as something other than "native" (Renganathan, 2015). Learning to identify the regional and social markers in one's own speech can be a challenge, and may be something worth teaching explicitly; I have only recently been able to identify regional and social markers to a limited degree with my own speech, and even then it does not feel intuitive.

2.1.2 *Centamil* and *koDuntamil*

The language situation in Tamil is hugely diglossic. *Centamil* refers to the "high", formal register that is reserved for literature, politics, and news media. It is contrasted with colloquial speech, *koDuntamil*, which is what heritage speakers are exposed to. Note that while *centamil* does not intersect with *vattara vazhakku* (that is, *centamil* remains the same regardless of the regional and social background of the speaker), *koDuntamil* describes a category of language varieties that includes both *pothu tamil* and any *vattara vazhakku*.

Diglossia in Tamil is stark, and so the surface forms of high register *centamil* are very different from low register *koDuntamil* surface forms (Annamalai, 2019; Renganathan, 2015). Speakers who have only been exposed to *koDuntamil* struggle to understand and communicate in *centamil* because they haven't been exposed to it, and the connection between the two is not transparent. This makes developing literacy skills especially difficult, as learners who have a high degree of oral proficiency in *koDuntamil* may struggle to read at

a very basic level (Renganathan, 2015). This can be very discouraging; informally, I can say that the majority of heritage Tamil speakers I know (and some first language speakers), many of whom communicate orally in the language very comfortably, feel that they "don't really know Tamil" because they don't know how to read and write. A lot of this feeling has to do with the fact that *centamil* carries implications of "purity", and "correctness", and many hold a strong belief that *koDuntamil* is an inferior, incorrect form of the language (Renganathan, 2015).[3] This may be in large part due to the prestige associated with literature in Tamil culture, though definitely the situation is more complex than I delve into in this work. Readers may find the discussions about linguistic purity in Annamalai (2011a) interesting on this point.

Though the focus of this work is *koDuntamil*, it is informative to understand the relationship between *centamil* and *koDuntamil*. Recall that *koDuntamil* is primarily spoken and *centamil* is primarily written; as such, the two are often referred to as "spoken Tamil" and "written Tamil". While the registers contrast across all linguistics domains, relevant to this work is the fact that written Tamil is generally accepted by Tamil linguists to constitute underlying phonological forms. The process that turns these written forms into their spoken versions is what is considered Tamil phonology. As Annamalai (2019) puts it,

> "Lacking an adequate phonology of modern Tamil, linguists take the transcription of the written language as the underlying phonological representation – simultaneously the output of the syntax and the input to the phonology – and the corresponding spoken form as the surface representation. The rules that convert one into the other are considered to be the content of Tamil phonology. […] this practice offers a good view of Tamil phonology because the transparent, agglutinating morphology of the language inhibits the development of complex morphophonemic patterns." (p. 105)

[3] The terms *centamil* and *koDuntamil* literally translate to 'pure Tamil' and 'harsh Tamil' respectively (Annamalai, 2019).

The “written language” Annamalai refers to is *centamil*, and the “corresponding spoken form” refers to *koDuntamil.* This is why, throughout this thesis I will frequently use the terms ‘underlying forms’ and ‘*centamil* forms’ interchangeably. The rows marked as *UR* for “underlying representation” in examples can also most of the time be understood as *centamil* forms, unless explicitly stated otherwise. In places where I omit the *UR* row, this is either because the underlying phonology is not relevant to the discussion, or more likely, because in that instance the *centamil* and *koDuntamil* forms are equivalent.

As mentioned above, the differences between *centamil* and *koDuntamil* occur across all domains, i.e. lexical, phonological, morphophonological, and morphosyntactic (Annamalai, 2019). Tamil contains very complex morphology with a high degree of agglutination. One of the benefits of *centamil* is that the underlying morphology of forms is generally quite transparent in this register, and so the morphosyntactic patterns for things like verb tense and aspect are easy to see, as seen in the first column of (1). In *koDuntamil*, however, forms undergo massive phonological reduction (discussed more in 2.2.5), and the resultant forms become much more opaque with respect to their underlying structure: observe that though the UR and MG rows are identical in *centamil* and *koDuntamil*, the MB of the *koDuntamil* phrase has a much less transparent relationship with the UR than its *centamil* counterpart does.

Note this isn’t a “definitive” *koDuntamil* version of the phrase in (1); this is a *koDuntamil* variant from my specific *vattara vazhakku* and so is simply one of many acceptable *koDuntamil* variants of this singular *centamil* phrase. Additionally, there aren’t universally recognized writing conventions for *koDuntamil*, and so illustrating differences between the two registers is very tricky in the written medium (Annamalai, 2017).

(1) [4]	*centamil*	*koDuntamil*
Transc.ta	குடித்துக் கொண்டு இருப்பாள்	குடிச்சிண்டிருப்போ
Trans.la[5]	kuDiththu koNDu iruppAL	kuDichchiNDrppo:
MB	**kuDi-ththu ko-NDu iru-pp-AL**	**kuDi-chchi-ND-r-pp-O**
UR	kuDi-thth(u) koL-ND(u) iru-pp-AL	kuDi-thth(u) koL-ND(u) iru-pp-AL
MG	drink-VBLP hold- VBLP be-FUT-3S.F	drink-VBLP hold- VBLP be-FUT-3S.F
Transl.	'she must be drinking'	

The distinction between *centamil* and *koDuntamil* is not a strict binary; rather there is a scale of formality, with *centamil* at the most formal end, and then varieties of *koDuntamil* which can vary significantly in degrees of informality towards the other. As a general rule of thumb, the more phonological deletion a given *koDuntamil* form has undergone in comparison to its *centamil* counterpart, the more informal it is.[6] I discuss this in more depth in 2.2.5.

Heritage speakers who are only exposed to *koDuntamil* often may have acquired complex morphological entities as fossilized forms without understanding their analysable parts (Renganathan, 2015). Because of this, what are relatively productive morphological patterns at play in these complex forms are obscured. Surface level forms seem to be "irregular" rather than products of regular morphological patterns, and it's harder for learners to form connections between morphologically related forms. This in turn makes it hard to apply grammatical patterns and processes to understand and produce new forms.

Some morphemes undergo a regular process of reduction in *koDuntamil*, and so for learners who are only interested in *koDuntamil*, these forms can be taught as such.[7] If

[4] In this and subsequent examples, each line corresponds to the following: *Transc.*= transcription (in *ta* Tamil orthography and *la* romanization), *MB*= morpheme breakdown of transcription, *UR*= underlying representation of the morphemes in *MB*, *MG*= morpheme gloss, *Transl*= English translation.

[5] See Appendix B for a key to the romanization I am using here.

[6] Which is not meant to discount the fact that *centamil* is also significantly different from *koDuntamil* in its morphosyntax (Annamalai, 2019; Schiffman et al., 2009).

[7] Annamalai & Asher (2002) do this in their *Colloquial Tamil* course; for example, they teach the present tense form *-(k)kiR-* in its reduced form *-R-* .

learners are explicitly taught how to relate the *koDuntamil* forms they already know to their *centamil* counterparts (for example, learning to associate the present tense *-R-* with its *centamil –(k)kiR-*), this has the potential to make learning new forms in either register easier and more intuitive (Renganathan, 2015).

2.2 Tamil verb morphology

Example (1) just above illustrates the morphological complexity of Tamil. This section provides an overview of Tamil morphosyntax and morphophonology as it relates to verbs. Tamil verbs can be divided into finite and non-finite verbs. Tense and person-number-gender (PNG) agreement are expressed on finite verbs in Tamil, and aspect and modality are (primarily) expressed through compound verb forms, which are themselves composed of finite and non-finite verbs. Discussion of compound verb forms first requires understanding how these finite and non-finite verbs in Tamil are constructed. This, in turn, requires knowledge of Tamil verb stem morphophonology, as all verbs in Tamil are built around a verb stem. 2.2.1 will define and describe verb stems, 2.2.2 will describe the morphology of finite verbs, and 2.2.3 will describe the morphology of non-finite verbs. In 2.2.4 I will define and describe compound verb forms. 2.2.5 will provide an example of the range of morphophonological processes that occur in *koDuntamil* that distinguish it from *centamil*.

2.2.1 Verb stems

Tamil is an agglutinative language. Derivational and inflectional morphology is expressed primarily through compounding and suffixes. (Annamalai, 2019).

The dictionary citation form of any given Tamil verb is its verb stem.[8] The verb stem is the verb root plus any derivational morphology that has been added to it, and is the form that any inflectional morphology is then added to (Lehmann, 1989). Derivational morphology is optional and encodes either transitivity, voice, or causation, each of which are expressed

[8] Note that I am adopting the term “verb stem” as defined in Lehmann (1989). Krishnamurti (2003) calls this same form a “verb base”.

by morphemes that Krishnamurti (2003) calls formatives. A verb stem does not always include a formative—a verb root on its own can also form its own verb stem (Krishnamurti, 2003). (2) and (3) show an example of a verb root with and without a transitive-causative formative. (2) is a verb root meaning “eat” that doesn’t have a formative and is a valid verb stem of its own. (3) shows this same verb root with the */-TT/* transitive-causative formative attached to it, forming a new verb stem with the meaning of “feed”. Both examples are adapted from Krishnamurti (2003, p.280)

(2)	*Example utterance*
Transc.ta	உண்
Trans.la	uN
MB	uN
UR	uN
MG	*eat*
Transl.	‘eat’

(3)	*Example utterance*
Transc.ta	ஊட்டு
Trans.la	u:TTu
MB	u:-TTu
UR	uN-TT
MG	*feed-TRANSCAUS*
Transl.	‘feed’

Verb stems are classified according to whether they are “weak”, “middle”, or “strong” (Lehmann, 1989). This classification is morphophonologically determined. Stems that take an epenthetic */-kk/-* before an inflectional suffix that starts with a vowel are strong stems. Stems that take an epenthetic */-k-/* in the same environment are middle verbs. Stems that

undergo no epenthesis in this environment are weak stems. This classification will become relevant when discussing the construction of certain non-finite verb forms in 2.2.3., as some of them are formed with a suffix that starts with a vowel.

2.2.2 Finite verbs

Finite verbs are the head of main clauses (Krishnamurti, 2003), typically occur at the end of a sentence, and are inflected in several moods including the indicative, imperative, optative, permissive, and hortative (Annamalai, 2019). Every complete utterance in Tamil must have one finite predicate[9] and these typically occur at the end (Annamalai, 2019). They are productively inflected in the indicative, the imperative, and the optative (Annamalai, 2019; Lehmann, 1989). There are also isolated finite forms such as the permissive *-aTTum* and the hortative *-alAm* which can only be conjugated with 3rd singular forms and the 1st person plural form respectively (Annamalai, 2019). The indicative form is the most unmarked and was the most prevalent in the data in the study reported on here, and as such is the form that is described here.

Finite indicative verbs are composed of a verb stem, followed by a tense marker, and then an agreement marker that fusionally expresses three separate features: person, number, and gender (PNG). The indicative mood is not overtly marked. This section will briefly describe Tamil tense morphology and PNG morphology. Note that Tamil sentences do not need noun phrases, and (4) shows how a single finite indicative verb can serve as its own utterance.

[9] Note that in Tamil predicates can be nominal or verbal, and nominal predicates occupy the same place in the syntax as finite verbal predicates.

(4)	*Example utterance*
Transc.ta	வந்தான்
Trans.la	vandhAn
MB	va-ndh-An
MG	come-PST-3M
Transl.	'he came'

Tense

There are three tenses in Tamil: present, past, and future. Each tense is expressed by a set of tense allomorphs. Verb stems are divided into seven classes, and the verb class that a verb stem falls under determines which set of tense allomorphs it takes (Arden, 1910, pp. 74-76; Lehmann, 1989, pp. 58-64). What class a given verb falls under is largely lexically determined.[10] Classes I – IV consist of weak stems, V has only middle stems, and VI and VII verbs are formed by strong stems. Table 1 shows what tense allomorphs correspond to each verb class (adapted from Lehmann (1989, p. 59).

[10] But see Subrahmanyam (1971) for a more streamlined description that reduces the seven classes to four by deriving the remaining three classes through productive morphophonological processes. As my goal isn't to provide the most economical descriptive account here, I've opted to stick with the 7-class representation for ease of reference.

Verb Class	Present	Past	Future
I	-kiR-	-dh-	-v-
II	-kiR-	-ndh-	-v-
III	-kiR-	-in-	-v-
		-i-	
IV	-kiR-	Doubling (the final consonant)	-v-
V	-kiR-	-T-	-p-
VI	-kkiR-	-thth- -chch- (colloquially after /i/, /e/, /ai/)[11]	-pp-
VII	-kkiR-	-ndh-	-pp-

Table 1: Tamil verb classes

The verb *iru* "to be", a verb from verb class VII, is shown in the (5) present, (6) past, and (7) future tenses.

(5)	*Example utterance*
Transc.ta	இருக்கிறேன்
Trans.la	iru**kkiR**En
MB	iru-**kkiR**-En
MG	be-**PRS**-1S
Transl.	'I am (here)'

[11] Annamalai (2017); Annamalai & Asher (2002); Subrahmanyam (1971)

(6)	*Example utterance*
Transc.ta	இருந்தேன்
Trans.la	iru**ndh**En
MB	iru-**ndh**-En
MG	be-**PST**-1S
Transl.	'I was (there)'

(7)	*Example utterance*
Transc.ta	இருப்பேன்
Trans.la	iru**pp**En
MB	iru-**pp**-En
MG	be-**FUT**-1S
Transl.	'I'll be (there)'

While there are some phonological similarities among verb stems of a given class, verb classification is far from being phonologically predictable. The classification is primarily morphological, rather than phonological. This poses a clear problem to the second language learner, who has to learn and memorize the classification of every verb stem individually. Before starting this study, I assumed that heritage speakers, in contrast to second language learners, would have L1-like mastery of verb classes. As we will see in chapter 4, this is partially true.

Person-Number-Gender

Tamil verbs mark person, number, and gender. These are all marked fusionally using a single verb-final morpheme. The 1st, 2nd, and 3rd persons are marked. Gender is only marked on the 3rd persons. Within gender, there is a primary classification between inanimate/non-human

(also called neuter) and animate subjects; then animate/human subjects are further broken down into female and male genders. There is a singular-plural distinction for all persons. There is a distinction between the 1st person plural inclusive and exclusive,[12] which is only distinguished in the pronominal forms; the verbal person-number-gender morphemes are identical. In the 2nd and 3rd persons, in addition to marking number, plurality is used as an honorific. The 3rd person also has a singular honorific form. 3rd person plural neuter marking is rarely used in colloquial speech, and instead the 3rd person singular neuter is generally used when talking about both singular and plural 3rd person neuter subjects.

The full classification of PNG forms is shown in Table 2 (first two columns reference Annamalai [2019]). The second and third columns differentiate between forms that belong to *centamil* vs. forms belonging to different varieties of *koDuntamil*. Note that the *centamil* forms can also be used in colloquial contexts (and many frequently are).

[12] Note the distinction exists in my dialect and the dialect of my speakers, but doesn't exist in every dialect.

Person-number-gender	*Centamil* form	Common *koDuntamil* forms
1S	*-e:n*	*-ẽ:*
1PL.EXC and 1PL.INC	*-o:m*	*-õ:*
2S	*-a:y*	*-e* *-a*
2PL	*-i:rgaL*	*-e:ngaL* *-e:nga* *-e:ngo:* *-e:L*
3F	*-a:L*	*-a:*
3M	*-a:n*	*-ã:*
3HON	*-a:r*	*-a:*
3SN	*-adhu*	
	(for class III verbs) *-iTTRu*[13] (incorporates past tense)	*-chchu* (incorporates past tense)
	-um (incorporates future tense)	Many compound verb forms (defined in 2.2.4) exist that fuse the 3SN morpheme with preceding auxiliaries.
3PL	*-a :rgaL*	*-a:ngaL* *-a:nga* *-a:*
3PL.N	*-ana*	3SN

Table 2: Person-number-gender morphemes in Tamil

Unlike tense, PNG morphology is consistent across all verbs; that is to say, there is generally no lexically conditioned PNG allomorphy a speaker has to be aware of. There remains variation in the PNG forms used in colloquial speech, as seen in the numerous forms in column three corresponding to each morpheme in column 1; but this variation is dialectically determined,[14] not phonologically or lexically. The one minor exception to this is the 3rd person neuter, discussed briefly in the next section.

[13] -இற்று

[14] I haven't included every possible *koDuntamil* form in this table; instead I've prioritized forms corresponding to the dialect spoken by my participants, which happens to be largely the same as mine.

Person-Number-Gender : 3rd person singular neuter

To reiterate from the previous section, a PNG morpheme always appears verb-finally, and generally surfaces as its own distinct morpheme. In the case of the third person neuter, this morpheme takes the form *-adhu*. There are four broad exceptions to this rule, where instead the PNG morpheme fuses with other preceding verbal morphemes to create a new form.

The first exception is when the third person neuter is the subject of a future tense verb. In this situation, the verb will carry the fused morpheme *-um*, which represents both tense and PNG (8)**.**[15]

(8)	*Example utterance*
Transc.ta	மழைப் பெய்யும்
Trans.la	mazhai peyy**um**
MB	mazhai peyy-**um**
MG	rain rain-**FUT.3SN**
Transl.	'it will rain'

The second exception is with class III verbs in the past tense. Class III past tense verbs can carry two different 3rd person neuter forms. The first form is what we would expect: the normal tense morpheme followed by the PNG morpheme *-adhu* (9)**.** The other form is the fused tense and PNG morpheme *-iTTRu* (10) (Lehmann, 1989). In colloquial speech, the form *-iTTRu* often surfaces instead as the form *–(i)chchu* (11) (Annamalai, 2017).

[15] Note that though this morpheme conveys tense, it does not change form based on the verb class of the verb -- tense allomorphy does not apply here.

(9)	*Example utterance*
Transc.ta	நாய் தூங்கியது
Trans.la	na:y thu:ngiy**adhu**
MB	na:y thu:ng-i-y-**adhu**
MG	dog sleep-PST-EUPH-**3SN**
Transl.	'the dog slept

(10)	*Example utterance*
Transc.ta	நாய் தூங்கிற்று
Trans.la	na:y thu:ng**iTTRu**
MB	na:y thu:ng-**iTTRu**
UR	na:y thu:ng-**iTTRu**
MG	dog sleep-**PST.3SN**
Transl.	'the dog slept'

(11)	*Example utterance - colloquial*
Transc.ta	நாய் தூங்கிச்சு
Trans.la	na:y thu:ng**ichchu**
MB	na:y thu:ng-**ichchu**
UR	na:y thu:ng-**iTTRu**
MG	dog sleep-**PST.3SN**
Transl.	'the dog slept'

The third exception takes place in the context of compound verb forms, when the third singular neuter combines with the completive auxiliary *viDu* 'leave' in the past tense. Compound verb forms are discussed more in 2.2.4, but as a brief overview, their structure is

lexical verb + auxiliary verb + inflectional morphology. The combination of the auxiliary *viDu* plus a past tense morpheme plus a 3rd singular neuter morpheme is underlyingly *viTTadhu*, (12) but is often colloquially realized instead as *(u)rththu* (13) (Annamalai, 2017; Annamalai & Asher, 2002)**.**[16]

(12)	*Example utterance*
Transc.ta	வந்துவிட்டது
Trans.la	vandhu**viTTadhu**
MB	va-ndhu-**viT-T-adhu**
UR	va-ndhu-**viT-T-adhu**
MG	come-VBPL-**leave-PST-3SN**
Transl.	'it has come/arrived'

(13)	*Example utterance - colloquial*
Transc.ta	வந்துர்த்து
Trans.la	vandhu**rththu**
MB	va-ndhu-**rththu**
UR	va-ndhu-[**viT-T-adhu**]
MG	come-VBPL-**leave.PST.3SN**
Transl.	'it has come/arrived'

The final exception is with the verb *iru*. *Iru*, when conjugated in the present tense with a third person neuter subject, generally surfaces without an overt PNG marker and only a partial tense morpheme (14).

[16] In section M14, Annamalai (2017) includes the low register form வந்துருச்சு *vandhuruchchu*. *chch* itself is a lower register phonological alternation with the (relatively) higher register *ththu*. Other examples of this phoneme alternation are shown in p. 106 of Annamalai & Asher (2002).

(14)	*Example utterance - colloquial*
Transc.ta	நல்லா இருக்கு
Trans.la	nalla: iru**kku**
MB	nalla: iru-[**kku**]
UR	nalla: iru-[kkiR-adhu]
MG	good.ADV be-[PRS-3SN]
Transl.	'it's nice/good/well'

2.2.3 Non-finite verbs

Non-finite verbs in Tamil are the head of subordinate clauses and must occur alongside a finite predicate. There are several different types of non-finite verbs, and they divide into two categories that are defined by whether they need to be followed by a verbal predicate (*vinaiyeccam*) or a nominal predicate (*peyareccam*.)[17] The first set is what will be relevant for this project and is comprised of the verbal participle,[18] the infinitive, the conditional, the negative, and the negative conditional (Annamalai, 2019).

The verbal participle is used in the formation of many compound verb constructions (see 2.2.4). It is formed by the verb stem, followed by the verbal participle morpheme, and then a euphonic [u] (Lehmann, 1989). The verbal participle morpheme is homophonous to the past tense morpheme; as such, the verb class the verb stem belongs to determines the shape of this morpheme. The first word in (15) is in verbal participle form. The verb, *paar*, 'to see', is a class VI verb, and as such takes the /-thth-/ past tense allomorph. Since the verbal participle is homophonous to the past tense morpheme, this same form is used to form the

[17] These literally translate to "anticipate a verb" and "anticipate a noun" respectively.
[18] Annamalai (2019) calls this the conjunctive.

verbal participle form of the verb, to create the final form *pAththu* (the rhotic in the root is often deleted colloquially).

(15)	*Example utterance*
Transc.ta	பாத்து விளையாடினான்
Trans.la	pa:**ththu** viLaiya:Dina:n
MB	pa:-**ththu** viLaiya:D-in-a:n
MG	see-**VBLP** play-PST-3S.M
Transl.	'he played carefully' *lit., 'he saw and played'*

Also used in the formation of many compound verb forms is the infinitive. It is formed by adding the infinitival suffix *-a* to the verb stem, as shown in (16). If the verb is a strong verb (i.e. falls under class VI or VII), then the suffix is preceded by *kk* and so in essence the infinitive takes the form *-kka,* as shown in (17). If it is a middle verb (i.e. falls under class V), it is preceded by *k*. As shown in (18), morphophonological interactions with the stem then result in the infinitive essentially taking the form *-kka* just as it would with strong stems.

(16)	*Example utterance – weak verb*
Transc.ta	உன் கூட பேசணும்
Trans.la	un kUDa pEs**a**Num
MB	un kUDa pEs-**a**-N-um
UR	un kUDa pEs-**a** vEND-um
MG	you with talk.speak-**INF** want.need.must-FUT.3N
Transl.	'I want to talk with you.'

(17)	*Example utterance – strong verb*
Transc.ta	*கவிதை படிக்க ஆரம்பித்தான்*
Trans.la	kavithai paDi**kka** a:rambiththa:n
MB	kavithai paDi-**kka** a:rambi-thth-a:n
MG	poem read-**INF** start-PST-3S.M
Transl.	'He started to read the poem'

(18)	*Example utterance – middle verb (colloquial)*
Transc.ta	*கேக்க கூடாதா?*
Trans.la	ke:**kka** ku:Da:dha:
MB	ke:**k-ka** ku:Da:dha:
UR	ke:T-**ka** ku:Da:dha:
MG	ask-**INF** not.allowed.to
Transl.	'Can I not ask (you)?'

Finally, the conditional is another form that appears frequently in speech. Like the verbal participle, it is formed by the verb stem followed by the verbal participle morpheme. Then the conditional suffix *-a:l* is added. This is shown below in (19).

(19)	*Example utterance*
Transc.ta	ஜானு வந்தால் ராம் சிரிப்பான்
Trans.la	Jaanu va**ndhAl** Ram sirippAn
MB	Jaanu va-**ndh**-**Al** Ram siri-pp-An
MG	Jaanu come-**VBLP-COND** Ram smile.laugh-FUT-3S.M
Transl.	'If Jaanu comes, Ram will smile/laugh.'

I did not look at negative constructions in this project, and so descriptions of the negative and negative conditional are omitted.

Non-finite verbs, *vinaiyeccam* in particular, are integral to the construction of compound verbs, which are detailed in the next section.

2.2.4 Compound verb forms

A Tamil compound verb construction, sometimes called a complex predicate (Krishnamurti, 2003), is either a lexical compound verb or an auxiliary compound verb; the latter is what is relevant to this work (Annamalai, 2019). Auxiliary compound verb constructions are sometimes also called complex verb forms (Bright & Lindenfeld, 1990), or in the context of *centamil*, periphrastic expressions (Arden, 1942; Bright & Lindenfeld, 1990). In these constructions, the main verb is lexical and all other elements of the verbal predicate are grammatical (Lehmann, 1989). These elements are auxiliary verbs that can express aspect, modality, passive voice, causation, negative polarity, or attitude. Every auxiliary verb is (or

used to be) a lexical verb in its own right;[19] it is grammaticalized only in the construction of a compound verb form.

Auxiliary verbs in a compound verb construction always attach to a non-finite verb base, and the type of auxiliary determines which type of non-finite verb (e.g. infinitive, verbal participle, conditional) this needs to be. I use the term verb base here rather than verb stem or main verb because a compound verb form can contain multiple auxiliaries. Each auxiliary that gets added to the main verb contributes to forming a new "base" for the next auxiliary. Thus the verb base is minimally one verb (the main lexical verb), but can be extended by several auxiliary verbs. There is no restriction on whether the final auxiliary needs to be in finite or non-finite form, so long as there is exactly one finite predicate (nominal or verbal) in the utterance (Annamalai, 2019). If a compound verb form is meant to be the finite predicate of a given utterance, then the last adjoining auxiliary needs to be finite. An example of this is shown later in this section in (22).

As described above, Tamil auxiliary verbs can express aspect, modality, passive voice, causation, negative polarity, and attitude. Only the first two were elicited in this thesis, and so described in the following sections.

Aspectual auxiliaries

The Tamil aspects are the perfect, the perfective/completive, the progressive, and the verb *Aku* (to become) which can function as a completive for certain verbs (Krishnamurti, 2003; Lehmann, 1989). The perfect and progressive are interpreted as would be expected. The perfective has a strong completive implication, and has also been said to describe a temporal or epistemic disjunction (one event ending before the other begins, or the event contrasting with the expectations of the speaker, respectively) (Annamalai, 2019). The aspectual usage of

[19] *koL* "to hold" is an example of a verb that has lexical meaning but has been nearly entirely grammaticalized in modern spoken Tamil to carry a progressive aspectual meaning when combined with the stative *iru*.

Aku essentially has the same interpretation as a perfective but only appears with select verbs.

All aspectual auxiliaries adjoin to a verb base in verbal participle form, the formation of which was described in 2.2.3**.** For example, the perfect aspect is expressed with the auxiliary verb *iru* "to be". By adding the appropriate tense marker after the auxiliary, this can be used to express the past, present, or future perfect. See Appendix C for details on the other aspectual auxiliaries.

(20) shows an example of a simple verb in finite form, and (21) and (22) show examples of the same verb as part of compound verb forms that are in the perfect aspect and progressive aspect respectively. (20) shows the verb 'to cry' in simple finite indicative verb form. The verb base /azhu-/ is followed by a past tense marker and the 1st person singular PNG morpheme *-e:n*, to create the finite verb phrase 'I cried'. (21), on the other hand, shows the same verb base /azhu-/ in non-finite verbal participle form, /azhudhu-/. It is then followed by the auxiliary verb 'to be' / iru-/, followed by the future allomorph /-pp-/. The addition of this auxiliary makes the compound verb form adopt a future perfect aspect, with the meaning 'I would have cried'. Finally, (22) is another compound verb form, this time with two auxiliaries. The first auxiliary, *koNDu*, in combination with the second auxiliary, *iru*, work together to create the progressive future aspect. Crucially, the first auxiliary is in verbal participle form, to allow the last auxiliary, which is in finite form, to adjoin to the verb base. Note that the PNG morpheme always attaches to this final auxiliary.

(20)	*Example utterance*
Transc.ta	அழுதேன்
Trans.la	azhudhEn
MB	azhu-dh-En
MG	cry-PST-1S
Transl.	'I cried'

(21)	*Example utterance*
Transc.ta	அழுதிருப்பேன்
Trans.la	azhudh**iru**ppEn
MB	azhu-dhu-**iru**-pp-En
MG	cry-VBLP-**be**-FUT-1S
Transl.	'I would have cried'

(22)	*Example utterance*
Transc.ta	அழுதிண்டிருப்பேன்
Trans.la	azhudhi**NDiru**ppEn
MB	azhu-dhi-**ND-iru**-pp-En
UR	azhu-dhu-**koNDu-iru**-pp-En
MG	cry-VBLP–**hold.VBLP-be**-FUT-1S
Transl.	'I would be crying'

Modal auxiliaries

Several modal finite and non-finite verb types exist including the concessive, permissive, hortative, and conditional (Annamalai, 2019). There is a final set of modals that are expressed through the use of modal auxiliaries. A subset of these auxiliaries are "defective"

(Krishnamurti, 2003) because they are only conjugated with the third person. All modal auxiliaries adjoin to a verb base in infinitival form. (23) shows an example of this using the "defective" auxiliary *vENDum* ('need'/ 'must' / 'want'). See Appendix D for details on the other modal auxiliaries.

(23)	*Example utterance*
Transc.ta	உன் கூட பேசணும்
Trans.la	un kUDa pEsa**Num**
MB	un kUDa pEs-a-**N-um**
UR	un kUDa pEs-a **vEND-um**
MG	you with talk.speak-INF **want.need.must-FUT.3N**
Transl.	'I want to talk with you.'

2.2.5 Morphophonological processes in *koDuntamil*

Phonological distinctions between centamil and koDuntamil

A defining feature of *koDuntamil* that distinguishes it from *centamil* is extensive phonological and morphophonological reduction. A simple but prevalent example of phonological reduction is with word-final vowel and nasal consonant combinations. In this environment, the nasal consonant is frequently dropped and instead the preceding vowel becomes nasalized (Annamalai, 2019). Reductive processes are often more intense than this however, to the point that it becomes hard to see how a *centamil* form and its corresponding *koDuntamil* equivalent are even related, as will be shown in the next section.

Another phonological feature of spoken Tamil is vowel alternations in certain words. Underlying /u/ in the first syllable of some words often surfaces as /o/ if the second syllable has a low vowel (/a/, /a:/, or /ai/) and there's only one consonant between the two vowels.

Examples of this alternation are in words like குடை /kuDai/ ‘ umbrella’, உடம்பு /uDambu/ ‘body’, and உடை /uDai/ ‘break’ which surface as கொடை [koDai], ஒடம்பு [oDambu], and ஒடை [oDai] (Annamalai, 2017). The inverse of this process is seen in words like கொடு /koDu/ ‘give’ and ஒண்ணு /oNNu/[20] ‘one’ which can surface as [kuDu] and [uNNu] in some dialects of *koDuntamil*.

Morphophonological reduction in koDuntamil

As already discussed, Tamil is highly agglutinative, with compound verb forms in particular tending to be morpheme-heavy, resulting in very long words. In *centamil*, all of the morphemes are overtly realized and the auxiliaries are generally realized as separate words. As the register shifts to more colloquial, the surface realization of the morphemes begin getting “shorter” either by deletion or some form of assimilation, and the auxiliaries attach to the main verb in the form of suffixes. These reductions are productive, and the synthesis of strings of auxiliaries onto verb stems creates forms that resemble those seen in polysynthetic languages (Bright, 1990b). The morphophonological processes that take place in *koDuntamil* that contribute to this “shortening” are primarily processes of assimilation and reduction.

As discussed in 2.2.4, all auxiliaries are independent words in their own right and tend to be represented as such in *centamil*. However, most are realized as dependent suffixes in *koDuntamil*[21]. The frequency with which they are realized as suffixes rather than independent words is likely attributable to different degrees of grammaticalization. As was mentioned in 2.2.4, *koNDu,*[22] the auxiliary that is used in expressing the progressive aspect, has been nearly entirely grammaticalized in *koDuntamil* to suffixes such as *-ND-, -kiTT-,* or *-TT-,* based on the dialect. Another common reduction is *vENDum* to *-Num* (see (23)] earlier in

[20] This is itself a *koDuntamil* form; its *centamil* equivalent is ஒன்று /onRu/.
[21] Several paradigms for these in the Brahmin dialect are provided in Bright & Lindenfeld (1990).
[22] *koL* in verbal participle form.

2.2.4). In present and future perfect forms, it is permissible in the colloquial register to reduce *iru* to just the rhotic—essentially, the vowels from the morpheme are deleted to shorten the verb form. In the past perfect, the entire verb root, i.e., the entire form *iru*, can be deleted in casual conversation so that all that remains of the auxiliary is its past tense marker.

(24), which presents a compound verb form that is in the progressive aspect, provides a good example of these reductive processes. The first column shows an utterance in *centamil*, and the following two columns show the same utterance in *koDuntamil*, where the second column is of a slightly higher register than the last. Recall from 2.2.4 that the progressive aspect is expressed by adding the auxiliaries *koNDu* and *iru* to a compound verb form verb base. The tense marker that follows *iru* determines whether the compound verb form is in the present, past, or future progressive (in this case, future). (24) demonstrates how the degree of reduction increases as the register becomes more colloquial. It presents the phrase 'she must be cooking', uttered in (a) *centamil*, (b) not exactly *koDuntamil* but moving in that direction, and (c) a variety of *koDuntamil*. In the first column each of the three verbs in the phrase is uttered as its own word. The second column shows the same underlying compound verb form, but this time the first auxiliary, *koNDu*, has been largely incorporated into the main verb through the process of deletion described above: the verb stem is deleted entirely and only its accompanying verbal participle remains. The final column, which represents speech at the most colloquial end of the register, goes even further. The underlying past tense morpheme */thth/* changes to the *koDuntamil* [chch] (Subrahmanyam, 1971), and the vowels from *iru* are deleted to enable it to attach to the previous auxiliary. Thus, what is realized as three separate words and seven overtly realized morphemes in the first column has been reduced to only one word with six overtly realized morphemes in the last. Simply following the top line written in Tamil gives a clear visual

indication of the overall reduction in fully realized phonemes that occurs as the register becomes more colloquial.

(24)	Utterance	Morpheme reduction
Centamil Transliteration Morpheme breakdown Underlying representation Morpheme gloss Translation	சமைத்து கொண்டு இருப்பாள் samaiththu koNDu iruppAL samai-ththu ko-NDu iru-pp-AL samai-thth koL-ND iru-pp-AL cook-VBLP hold-VBLP be-FUT-3S.F 'she is probably cooking'	N/A
in-between *centamil* and *koDuntamil*	சமைத்திண்டு இருப்பாள் samaiththiNDu iruppAL	*koNDu* → *NDu*
koDuntamil	சமச்சிண்டிருப்போ samachchiNDrppo:	*ai* → *a* *thth* → *chch* *koNDu* → *NDu* *iru* → *r* *a:L* → *o:*

In addition to these auxiliary reductions, there are several other morphemes that either get reduced or changed in some way in the colloquial register. For example, *-kiR-*, the present tense allomorph used in verb classes I - V, is so often reduced to the trilled rhotic that it is taught as such in some conversational Tamil textbooks (Annamalai & Asher, 2002). There are several other deletion and assimilatory processes that complicate Tamil morphophonology.

Recall that *koDuntamil* itself consists of numerous *vattara vazhakkukaL*.[23] All *koDuntamil* utterances in this thesis that aren't produced by the speakers I worked with are, unless

[23] '*vazhakkukaL'* is the plural of *'vazhakku'*.

specified otherwise, from my *vattara vazhakku,* simply because it is what I am most familiar with; it is also largely the same *vattara vazhakku* as the one spoken by my participants. The same *centamil* utterance, however, would undergo different forms of reduction depending on the individual *vattara vazhakku* being spoken. Renganathan (2015) presents several clear examples of the variations present in different varieties of *koDuntamil*. These are duplicated below in (25) and (26) (transcriptions have been modified and reformatted to match the conventions being used in this paper.) Madurai, Thanjavur, and Tirunelveli are all regional dialects. The Brahmin dialect, as discussed in 2.1.1, is a caste dialect.

(25)	Utterance	Morpheme reduction
Centamil	வந்துவிடுவான்	N/A
Transliteration	vandhu**viDu**va:n	
Morpheme breakdown	vandhu-**viDu**-va:n	
Morpheme gloss	come.PST-**COMP**-FUT.3M	
Translation	'he will (definitely) come'	
Thanjavur dialect	vandhu**D**uvã:	*viDu* → *D*
Madurai dialect	vandhu**R**uvã	*viDu* → *R*
Tirunelveli dialect	vandhu**N**uvã:	*viDu* → *N*

(26)	Utterance	Morpheme reduction
Centamil *Transliteration* *Morpheme breakdown* *Morpheme gloss* Translation	*வந்துக்கொண்டிருக்க வேண்டும் என்கிறது* vandhukoNDirukka ve:Ndum enkiRadhu vandhu-koNDirukka ve:Ndum enkiRadhu come.PST-PROG.INF want the.fact.that 'the fact that one needs to come'	N/A
Thanjavur dialect	vandhukkiTTRukkNunRadhu	*koND → kiTT* *ve:NDum → N(u)* *enkiR → nR*
Thanjavur dialect (fast speech)	vandhuTTirukkNunRadhu	*koND → TT* *ve:NDum → N(u)* *enkiR → nR*
Tirunelveli dialect	vandhuNurukkaNunRadhu	*koND → N* *ve:NDum → N(u)* *enkiR → nR*
Brahmin dialect	vandhuNDirukkNunRadhu	*koND → ND* *ve:NDum → N(u)* *enkiR → nR*

A complete description of all of the reductive processes that occur in the morphophonology of *koDuntamil* across different *vattara vazhakkukaL* would merit its own thesis. In this section I merely hoped to underscore how the reductive processes in *koDuntamil* forms obscure underlying morphological patterns, which then makes learning verb forms challenging for learners. I also want to emphasize that *koDuntamil* is not a static variety and

contains multitudes of dialects (and thus multitudes of different reductive processes) within it.

2.3 Heritage language acquisition

As the focus of this work is on *heritage* Tamil, in this section I provide some background on factors that are unique to the process of heritage language acquisition in particular. In this work I use the definition of a heritage language speaker following the one presented by Polinsky & Kagan (2007) and Valdés (2000): A heritage language speaker is someone who was raised in a home where the heritage language was or is spoken, but where the dominant language of society was different from the heritage language. The speaker then develops some degree of bilingualism in both languages. Crucially, the heritage language is initially acquired as an L1, but is not fully acquired because of the switch to the dominant language of society. When exactly this switch happens, and to what degree, is not a constant, which means that heritage speakers span across a wide range of proficiencies (Polinsky & Kagan, 2007, pp. 3-4).

The actual process of heritage language acquisition shows huge variation from speaker to speaker. In addition to linguistic factors, the sociocultural and political environment the heritage language speaker is in has a huge effect on their acquisition. For instance, how much the family of the child chooses to speak to them in the heritage language, whether there are other community members they can hear the language from and speak with, the domains of use for the language, and the child's own interest in speaking the language all influence the child's acquisition and proficiency in the heritage language (Montrul, 2016).

Montrul in her 2016 book *The Acquisition of Heritage Languages*[24] discusses several other parameters that directly influence and/or help model the variation seen in heritage

[24] While there is also more recent work on heritage languages (see Montrul & Polinsky [2021]), I found that this book presented important variables around specifically acquisition in a clear way, that is still relevant in accordance with current research.

language acquisition. This section provides a brief overview of the most relevant ones: acquisition vs mastery, simultaneous vs sequential bilingualism, the age of acquisition vs age of onset of bilingualism, and attrition vs incomplete acquisition.

2.3.1 Acquisition vs mastery

Acquisition of language and specific linguistic structures generally refers to there being evidence of the individual having competence in said language or structure as part of their mental grammar, but Montrul (2016), citing Berman's definition from her (2004) book, makes the important distinction between acquisition and mastery. Mastery, distinct from mere acquisition, is when the individual demonstrates proficiency in "several dimensions of knowledge", and shows both competence *and* accurate production in more than just "chunks" (stock, memorized phrases) or select, limited contexts. Mastery of a form is said to represent the stage at which the learner produces the form in question accurately 90-100% of the time; contrasted with acquisition of the form, wherein the learner may use it accurately, but not in all the places they need to and with inconsistent accuracy.

An example of the distinction between acquisition and mastery is seen in Lakshmi Bai's (2004) study of acquisition of the dative case in L1 learners of Tamil. Lakshmi Bai discusses how children start producing the dative around the age of 2, but overextend it to contexts that are infelicitous till at least 2;2, with these infelicitous uses starting to reduce around 2;4. Thus, we could say that the dative is acquired in L1 learners of Tamil by age 2, but not mastered until at least half a year later.

2.3.2 Simultaneous vs sequential bilingualism

Heritage language acquisition is a type of bilingual acquisition, which is itself subject to a lot of variation. Montrul (2016) describes four dimensions in which a bi/multilingual individual's languages in their linguistic repertoire can be compared: order of acquisition (first language, second language, or simultaneous with another language(s)), dominance, function, and

socio-political status (majority or minority). These factors interact with each other: because of the minority status of the heritage language, the majority language of society generally becomes the dominant language of the heritage language speaker. As such, heritage language speakers are a classic example of "unbalanced bilingualism" (Montrul, 2016), where the speaker becomes more dominant in one of their languages (the majority one) than the other(s).

Heritage language speakers vary in terms of whether they are "sequential" or "simultaneous" bilinguals, and this distinction has a significant impact on their proficiency in the heritage language. A sequential bilingual is first only exposed to their heritage language, and then later as an L2 learner to the majority language of the society they are in. This could be as an early child L2 learner (ages 4-10), or late child L2 learner (10+) (Montrul, 2016).

Sequential bilinguals may be individuals who are born in a society where their heritage language is also a majority language, but then immigrate as children to a society where their heritage language is minoritized. Alternatively, they could be born and raised in a society where their heritage language is minoritized, but in the home were only exposed to this language, and so their first exposure to the majority language is in school (Montrul, 2016).

Simultaneous bilinguals, on the other hand, are exposed to both the heritage language and the dominant language from the same age, i.e. birth. Montrul cites several studies that show that sequential bilinguals are more proficient in their heritage language compared to simultaneous bilinguals, and that simultaneous bilinguals are at greater risk of attrition.

The distinction between sequential and simultaneous bilinguals is extremely significant; however, from my own informal experience I would add that it can be difficult to draw the line between the two. For example, some children are primarily exposed to the heritage language at home, but their caregivers may work in the majority language and

occasionally use this in the home as well, at a much lower frequency. Would this be simultaneous bilingualism? Consider also a child whose family immigrates to several different countries before the child is 10 years old, sometimes to countries where the heritage language is a majority and sometimes to ones where it isn't. Would this be sequential bilingualism? It seems to be the case that the distinction between simultaneous and sequential bilingualism is not always clear-cut, and may be a spectrum rather than a binary.

2.3.3 Age of acquisition vs age of onset of bilingualism (dominant language shift)

It is well known that both first and second language acquisition are sensitive to age effects (Mayberry, 2007), and that age of first exposure to the target language has a significant effect on the acquisition process. Children acquiring a heritage language initially acquire it as an L1, and so are (initially, at least) sensitive to the same age effects as any other L1 learner (i.e., reaching similar linguistic milestones at roughly the same times.) Possibly more importantly, however, the acquisition of heritage languages is also sensitive to the age of onset of bilingualism; that is to say, the age at which the child begins learning the dominant language of the society they are in leads to different outcomes in their continued acquisition of their heritage language (Montrul, 2016). This of course heavily intersects with the ideas discussed in the previous sections about the distinction between simultaneous and sequential bilinguals.

Language learning requires substantial input and opportunities to use and practice the language. In a heritage language situation, there is a point at which input and opportunity to produce the dominant language increase while the same decreases for the heritage language. This point generally arrives when the child begins schooling in the dominant language (though individual situations can and do vary significantly). The younger the child is when this language imbalance starts to arise, (or dominant language "shift" happens), the more vulnerable they are in the acquisition and retention of their heritage language. This is especially true if the shift occurs before puberty, which is thought to be the

point at which any L1 has properly solidified (i.e., been "mastered") in the speaker's mind (Montrul, 2016).

The school-age period is an important period for cognitive development, neuro-biological development, and critical language learning (Montrul, 2016). Not receiving instruction in the heritage language can severely impact future proficiency in the heritage language (while conversely, receiving any amount of schooling in the heritage language would be a significant asset [Cummins & Swain, 2014]). This means that, on the whole, children who immigrate from Tamil Nadu to Canada before attending school in India will likely be at greater risk of losing proficiency in Tamil than their peers who immigrated after doing a few years of school in India. Ultimately, age on its own cannot be taken as a predictor for language proficiency; rather, the quality and quantity of input as occurs as a result of the speaker's language experience is what matters (Ortega, 2019).

The interplay and intersection between age effects, sequential vs simultaneous bilingualism, and acquisition vs mastery (which is itself a function of the previous two factors), come together in the next section that discusses the distinction between incomplete acquisition and attrition.

2.3.4 Attrition vs incomplete acquisition

The gaps in linguistic knowledge that heritage learners often feel they have compared to homeland "native" speakers can be attributed to one of two things: attrition or incomplete acquisition.

Incomplete acquisition happens when the onset of bilingualism (and shift to the dominant language) occurs before the child has fully acquired the heritage language (Montrul, 2016). Not all the forms are acquired, because input has decreased at the expense of the dominant language, as have opportunities to practice speaking in the heritage language. This then results in a *divergent grammar* (Benmamoun et al., 2013, as cited by Montrul [2016]).

Attrition, on the other hand, is defined by Montrul as a situation where an aspect of linguistic structure was acquired, mastered, retained, and then subsequently lost. She cites several studies on attrition occurring in individuals who immigrate as adults to places where their heritage language is a minority language. The attrition generally surfaces as the speakers struggling with lexical retrieval, slower processing time, code-switching to fill in lexical gaps, being perceived by homeland speakers as having an "accent", and occasionally some pragmatic differences like using a subject where one is not required in a null subject language. In general, attrition in adults does not seem to significantly affect morphosyntax or phonology.

In heritage language contexts, however, attrition occurs with children who are still acquiring their heritage language, and so its effects are much starker than they are with adult speakers who are presumed to be fluent L1 speakers before the attrition occurred. Here, it is key to make the distinction previously discussed between acquisition and mastery: certain structures may have been acquired by the child before the shift to the dominant language, but they may not have been mastered and retained yet. For example, a certain form may be initially acquired in early child language development, but not mastered by the time the child starts school in the dominant language and begins to receive less input in the heritage language. As such, these structures are more vulnerable to attrition. In addition to this highlighting the important distinction to be made between acquisition and mastery, Montrul notes that this fact raises the question of whether acquisition of any L1 is truly "complete" by the ages of 3-4; rather, if we are including mastery and retention as necessary elements in "fully" acquiring a language, is the acquisition process more likely to continue till and around the age of puberty (or even later throughout the schooling period)?

Any given heritage language grammar can't be characterized wholly as affected only by attrition, or only by incomplete acquisition. Rather, certain aspects of the grammar may have been subject to attrition, while others were incompletely acquired. Structures that

were attrited may be easier to re-learn, since they won't be brand new to the learner, compared to structures that were incompletely acquired or not acquired at all. Informally, I recall a moving instance with one of my heritage Tamil adult students who commented that she "could hear [the voice of her] mother" in her head saying a specific phrase we were working on; that is to say, she was unearthing a long-held passive competence she had acquired but just hadn't had the opportunity to practice for years.

Attrition can of course be easily mixed up with incomplete acquisition, especially if there isn't any longitudinal data to compare against. Nevertheless, data about first language acquisition can be used to estimate which linguistic structures get acquired when, and this information combined with knowledge of when the onset of bilingualism occurred can help determine what linguistic gaps are a result of attrition rather than incomplete acquisition. For example, missing structures and forms that require more input, or gaps in vocabulary, are likely the result of incomplete acquisition rather than attrition (Montrul, 2016). This information can in turn inform what teaching strategies are used to (re)introduce these structures to students.

2.4 Tamil language acquisition

2.4.1 First language acquisition of Tamil

While the research looking specifically at the process of acquiring Tamil as a heritage language or of L1 English speakers acquiring it as an L2 is quite limited (to my knowledge), there is more literature on first language acquisition of Tamil in non-heritage contexts (though it should be said, much of the literature is based on the data of the same four speakers). This research, in conjunction with the previous section detailing factors relevant to heritage language acquisition more generally, can provide context for the specific situation of heritage Tamil acquisition.

Lakshmi Bai (2000) details a longitudinal study of her two children's speech, focusing on their phonological development growing up as simultaneous Tamil and Telugu bilinguals in Hyderabad. This could be argued to be a heritage Tamil situation, as Tamil is not a majority language in Hyderabad (which is the capital of the primarily Telugu-speaking state of Telangana). However, the "heritage" language situation in India is very different from that in Canada, so I will leave aside the implications of that for now. Lakshmi Bai describes how the Tamil retroflex consonants took a long time to acquire and, more importantly, stabilize—the voiceless retroflex stop /T/[25] ([ʈ] in IPA) stabilized by around 3 and a half years, the flapped allophone of its voiced counterpart /D/ [ɖ] by 3 for one child and past 5 for the other, the retroflex nasal /N/ [ɳ] between 3 to 4 and a half years of age, the retroflex lateral /L/ [ɭ] around 4, and the retroflex rhotic /zh/ [ɻ] between 5 to 6. The trilled rhotic /R/ [r] stabilized between 3 and a half to 4 and a half. I highlight these phonemes specifically, as the ones that are unique to Tamil and don't appear in "standard" North American English (excepting the retroflex rhotic). These phonemes are some of the last to be acquired for L1 Tamil learners and seem to stabilize (or be "mastered") very near to the age when most children begin school—for heritage learners, near the age when they begin shifting to the dominant language. I would hypothesize that these sounds are therefore the most vulnerable to attrition.[26]

Moving on to morphology, much of the literature provides evidence arguing that all verbal inflections are acquired early by L1 Tamil children, around the age of 2 (Lakshmanan, 2006; Raghavendra & Leonard, 1989; Sarma, 2013; Vaidyanathan, 1988, 1991). Lakshmanan (2000) notes that the data of one of the children shows tense and aspect inflections were

[25] Moving forward, I transcribe Tamil using only romanization and Tamil script as this is more easily accessible for Tamil speakers. A romanization-IPA-Tamil script mapping is included for reference in Appendix B.

[26] This is of course only representative of production knowledge. It is possible that speakers possess the underlying phonemic distinctions and can hear the differences, but may struggle to produce them for other reasons.

fully productive by age 2, but non-finite inflected verb forms appeared later, with the verbal participle and infinitive at 2;1.5, and the conditional at 2;2.28. Nevertheless, this still appears to be before 2 and a half years of age.

As discussed previously, verbal inflection in Tamil is extensive: it includes tense, aspect, modality, person-number-gender agreement, and some forms of interrogatives and negation. Raghavendra & Leonard (1989) argue that the reasons for the early acquisition of this system are similar to the reasons verbal inflections are also acquired early in Turkish, another agglutinative language: Tamil verbal inflections are all suffixes, each suffix represents one type of inflection, they are all syllabic (and therefore possibly more salient), they are obligatory, appear on content words, are semantically transparent, and are paradigmatically regular, productive, and consistent, (though as was discussed in 2.2, there is complexity introduced by the system of verb classes and by many underlyingly regular morphological patterns being obscured in *koDuntamil*.) Sarma (2013) argues that because Tamil is built around the verb and verbal inflections are obligatory, necessity for communication drives their early acquisition.

Whatever the reason for this early acquisition, it raises interesting questions for heritage language learners. If we assume that most heritage language learners hear Tamil regularly at home, then we would expect that as part of the usual process of L1 acquisition, verbal inflection would be fully acquired before they were exposed to English at school. The presence of significant variation between the verbal inflections produced by young adult and adult heritage Tamil speakers compared to those produced by L1 homeland speakers then indicates that these forms were maybe acquired but not mastered, and subsequently attrited; or perhaps they weren't acquired at all, and the L1 acquisition process is in fact different for heritage Tamil children than Tamil children growing up where it is a majority language.

While not exactly a heritage language situation, the findings from Gnanaseelan (2022) shed some light on the acquisition process. The author worked with a 6 year old Sri Lankan Tamil speaker who is growing up in Sri Lanka (where Tamil is a majority language) but attends school in English, and so is going through either simultaneous or sequential bilingualism. The child appears to be comfortable using verbal affixation in a manner comparable to how an adult speaker would. Tense suffixes are used extensively, though the child sometimes shows confusion with the past and future morphemes, and in general he does not use affixes to indicate plurality. Based on these results, perhaps it is the case that verbal morphology, such as tense affixation, is acquired as early as 2 and a half years old, but not fully mastered by the time the child starts school.

To my knowledge, there does not exist research directly comparing L1 Tamil acquisition in the homeland with heritage Tamil acquisition, so this would be a fruitful area for future study.

2.4.2 Heritage Tamil teaching

As discussed in 2.3, heritage speakers demonstrate significant variation in their language proficiencies. This variation makes it difficult to teach a heritage language class as there is no established benchmark for what a "typical" learner looks like.[27] Speakers can be at varying ranges of proficiencies, from those who understand few key phrases, to those who can communicate fairly fluently and only rely on the dominant language to fill in lexical gaps. This large variation can make designing a curriculum difficult, as you run the risk of either making it "too easy" or "too difficult". Thus, you cannot rely entirely on just a foreign language learning curriculum (which may be too basic for a more advanced heritage speaker), or on a curriculum designed for first language speakers (which may be too challenging for a less proficient heritage speaker) (Correa, 2011). Compared to L2 learners,

[27] This connects in a broader sense to the pedagogic challenge of teaching mixed-level classes; Nijhawan (2017) discusses this in the heritage Hindi-Urdu context.

heritage speakers generally don't have as strong metalinguistic skills nor do they depend on metalinguistic knowledge in the same way. Their oral proficiency tends to far outstrip their literacy skills because of hearing the language at home, in comparison to L2 learners who are often taught both at the same time (Montrul, 2016).

Through my teaching, I've learned first-hand that the level of proficiency a heritage Tamil speaker may have in one competency (e.g. written comprehension) often has little to no correlation with the level of proficiency they have in another (e.g. oral production). This is in contrast to foreign language learners, who have no prior knowledge of the language and so develop skills in each competency in a more predictable order, i.e., the order in which they are taught (though of course there can still be variability there).

Though age is not a correlate of language proficiency in the heritage language classroom, it can still play a critical role in determining which teaching strategies to employ. Explicit instruction can be beneficial for those with strong analytical skills (Ellis & Shintani, 2013, p.21). As very young learners (e.g., under the age of 10) are unlikely to have the analytical skills of an adult or even an adolescent learner, explicit instruction in the form of teaching grammar rules is therefore more likely to be effective with older students. Thus, older children and adults can benefit from both immersive teaching strategies and explicit instruction because of their increased analytical abilities, and likely also an increased metalinguistic awareness. In all cases, learner preferences are always vital to take into account when deciding on best teaching strategies (Nijhawan, 2011).

Finally, I want to acknowledge that language production does of course not provide a full picture of a learner/speaker's overall knowledge of a language. In the case of any speaker, there are various factors that can impact production (i.e., even fluent speakers can sometimes make "mistakes" in casual speech), and it is shown that learners often have underlying competency of morphological forms they may not always accurately produce (Lardiere, 2005). In the case of heritage languages in particular, learners often have a lot of

anxiety about speaking, which can negatively impact their production abilities (Jee, 2016, 2020; Sevinç & Backus, 2019). Also, many heritage speakers are "receptive bilinguals"; that is, they understand the language when it is spoken to them, but may struggle to speak it themselves. In these cases especially, production data does not do justice to a speaker's actual language understanding. Despite this, I chose to focus this study on production. I chose to do this because my objective was to identify challenges that my students and heritage Tamil peers face in *speaking* the language, and thereby provide practical suggestions to apply in the classroom. I discuss this choice and the research question in more detail in the next section.

2.5 Research question

This research is a pilot study investigating what challenges heritage Tamil speakers face in their production of verbs in informal, spoken conversational contexts. Because of the time constraints and the limited number of participants I was able to work with (see 3.3), this was designed to be an explorative and qualitative study**.** The findings from this research are intended to be a resource that heritage Tamil educators can draw on in their curriculum development and lesson planning, and to motivate more specific hypotheses that can be tested in future research on heritage Tamil speech and heritage language speech in general.

The study was deliberately designed to be broad in scope with respect to the areas of the grammar that it examines because of its exploratory nature. As such, the aim was to elicit both "simple" (finite and non-finite) verbs as well as compound verbs in order to capture as wide a range of the participants' knowledge of verbal morphology as possible. Eliciting compound verb forms in particular was thought to be beneficial since they, by necessity, contain simple verbs within them and are likely to exhibit more of the morphophonological processes briefly touched on in 2.2.5. There were four guiding questions that structured the design of the study and the analysis of the data. Because of the constraints described above, these were meant to also be exploratory questions rather than

subject to a detailed quantitative analysis. The questions were motivated by my own experience as a heritage speaker and language teacher, and the observations and suggestions of the heritage Tamil teachers that I work with. The guiding questions were as follows.

How do heritage Tamil speakers and fluent L1 Tamil speakers differ in their:

1) Production of PNG agreement morphemes?
2) Production of (past, present, future) tense morphemes and tense allomorphy?
3) Comprehension of the completion entailment[28] of the perfective?
4) Production of compound verb forms?

The verbal elements that are the focus of these questions are elements that the heritage Tamil teachers I work with identified as being challenging for students, and so I felt they would be relevant to structure my research around. PNG agreement morphemes and tense morphemes in particular are required in any finite verb, and so can reasonably be seen as a necessary foundation for constructing a simple Tamil verb. Ultimately the data lent itself to an analysis primarily only of the first two questions, but the second two are worth recording here as areas for future research.

I focus on production rather than comprehension as I know from my own personal experience, and that of the teachers and students I work with, that heritage speakers often feel they understand the language well enough but want more support in speaking. Thus, while it is true that production tasks don't give a full picture of a speaker's internal grammar, I limited the scope to production because the aim of this study is to produce results that are directly applicable for teaching.

[28] I.e., the use of the perfective to semantically entail the completion of an event.

3 Methodology and methods

I begin this chapter by describing the methodological framework and guiding principles I used to ground this work, before going into the details of the participants I worked with, the tasks I used, how I analysed the data, and providing an overview of the data I ultimately ended up working with.

3.1 Methodology

Tamil has a rich linguistic tradition dating back millennia to the Tamil literary and grammar treatise *Tholkaappiyam,*[29] and culturally there has been and remains great linguistic pride among Tamil peoples and speakers. Since my linguistics training has been entirely in the North American tradition, my research and analysis is primarily grounded in this framework; however, I incorporate elements of Tamil linguistic tradition by opting to use the native Tamil terms for linguistic phenomena which I feel are not as accurately or elegantly represented by their English counterparts (i.e., the term *vattara vazhakku* used in 2.1.1). I provide English translations wherever I do so. In future, I would like to ground my research more strongly in the Tamil linguistic tradition. I believe this would highlight many more linguistic patterns that are often hidden when viewed from the western lens.

For my data collection, I primarily drew on methodologies used in Indigenous language reclamation and revitalization work. This is because the values of relevance, responsibility, respect, and reciprocity (Gardner, 2012) that are integral to these methodologies resonate with me personally: These are values that I recognize as being indispensable and key to my culture as well, and so it felt natural to ground my work in such principles rather than in the more removed, individualistic methodologies I associate with traditional western academia which feel alien to the culture I was raised in.

[29] The exact date of composition of the *Tholkaappiyam* is still quite uncertain and the subject of significant debate.

This project was motivated by the needs of my community (as I have been able to identify to the best of my ability) and so I aspire to have the work and results be first and foremost relevant to this community. For example, the research question discussed in 2.5 was designed by me and the other heritage Tamil teachers collectively brainstorming areas of the grammar that seem to be challenging for our students, and so would be useful to know more about. In addition, several of the tasks to be described in 3.4 were either inspired by classroom activities we have used before, or can be used as classroom activities once the research is done; all of the tasks were designed with the idea that they can easily be adapted into teaching tools.

I wanted to avoid conducting exploitive linguistic work that removes the cultural and individual contexts from the language I am working with- respect for peoples and culture was a requirement in every activity I used. I make sure to present every form I discuss within the full context it was uttered in, and discuss the forms in relation to the individual speaker who uttered them rather than grouping the speakers together.

In doing this work, I want to make sure I acknowledge my privileges and avoid doing harm. As mentioned in 2.1.1, casteism continues to be pervasive in Indian communities, and in Tamil is socially marked through language variation. I do not want to, even unintentionally, promote my dialect, which reflects my caste-privilege, as the standard dialect or afford it inordinate prestige and power when doing this work. At the same time, I do not wish to erase the fact that this linguistic variation exists, as I feel that acknowledging this difference is one part in recognizing the reality that casteism still exists. Unfortunately, I do not have the sociolinguistic awareness to be able to discern which words and morphemes in my speech are markers of my caste. I worked with a fluent L1 speaker to identify these markers. I then made sure not to privilege specific dialect markers when conducting analysis of the participants' speech, and where relevant, highlighted forms that are obvious dialect markers and noted that I am not intending to present these as either "standard" or "inferior" forms.

While not intentional, I ended up working with participants who all spoke dialects nearly identical to my own. Though this made a lot of the analysis work easier, I want to affirm that the examples presented in this work are *not* meant to be representative of all Tamil speakers – they just represent the speech of the people I worked with, and this speech is not more or less "standard" than any other variety of the language. While it is highly prevalent in the Tamil context (see discussion in Annamalai [2011a]), I do not subscribe to any ideologies of language purism.

Ultimately, I saw this project as being a shared journey with my peers. I myself am a heritage language learner of Tamil, and I hoped to improve my own oral proficiency through doing this work. By working and learning with my heritage language learning peers through the process of conducting this research, I hope that everyone came out of the experience feeling that they gained something, whether it be insights into their own oral proficiency, increased confidence, or motivation to connect with other speakers with whom to continue a language learning journey. The success of my project is determined as much by its adherence to these values as it is by the more technical aspect of analysis of forms.

3.2 Structure of participant interviews

I conducted two sessions with each participant. I asked them to fill out a background questionnaire (detailed in 3.3.1) prior to the first session. This first session consisted of me explaining the goals of the research, how the sessions would run, and then conducting a sociolinguistic interview. The language backgrounds of the speakers in 3.3 were described entirely using the information shared with me as part of the background questionnaires and sociolinguistic interviews. The second session included an oral narrative task, picture description tasks, and a picture matching task (all detailed in 3.4). The language data analysed in chapter 4 comes entirely from this second session. The sociolinguistic interviews took between twenty to thirty minutes per participant. The oral narrative task and each production/comprehension task combined also took about twenty to thirty minutes in total.

Cumulatively, the two sessions combined were about 40-60 minutes per participant. Sessions were conducted virtually over Zoom and recorded using the built-in Zoom audio recording feature. The interviews were then manually transcribed using ELAN.

3.3 Speakers

3.3.1 Language background questionnaire and sociolinguistic interview

Developing an understanding of the speaker's background is essential when working with heritage language speaking populations, and the most common ways to do this are through the use of a language background questionnaire and/or sociolinguistic interview. Montrul (2016) lists several important sociolinguistic variables that she says must be captured when working with heritage language speaking populations. Prior to the sessions with each of the participants, I provided them with a background questionnaire to fill out that touched on some of these variables. Subsequent to this, I conducted a sociolinguistic interview that followed-up on any unclear answers and explored the variables that are harder to capture in a questionnaire. With each participant, I made sure to touch on the key sociolinguistic variables discussed in 2.3: type of bilingualism, age of acquisition, and age of onset of bilingualism. It so happened that these variables were not especially relevant to the details of the analysis and discussion in chapters 4 and 5, but I include description of them here anyway in case it is relevant for future research/researchers.

Table 3 lists variables that were captured as part of the background questionnaire. Table 4 lists the topics explored in the sociolinguistic interviews. The interviews were conducted one-on-one with each participant, and the conversations were in a mix of Tamil and English matching the comfort of the participant.

The next section describes each participant's language background based on the details they shared in their questionnaire and as part of their sociolinguistic interview.

Background questionnaire
Place of birth
Current age
Languages spoken + self-assessed proficiency
Age of acquisition of languages spoken
Level of education
Current occupation
Age at immigration (if not born in the GTA)
Languages of schooling
Languages spoken by parents

Table 3: Sociolinguistic variables covered in background questionnaire

Sociolinguistic interview
Participant's attitude towards Tamil, and whether/how it has changed through the course of their life so far
Participant's own perception of their fluency in the languages they speak (building on what they shared in the questionnaire)
Number of people who speak to participant in Tamil, and their relationship
Number of people participant speaks to in Tamil, and their relationship
Domains of use of Tamil in their life
Any changes in the domains of use of Tamil throughout their life (e.g. preschool, childhood, high school, post-secondary)
Language(s) heard the most in the home
Parents' attitude towards Tamil
Parents' beliefs about passing Tamil down to their children
Education and occupation of parents

Table 4: Sociolinguistic variables covered in sociolinguistic interview

3.3.2 Language backgrounds of participants

I worked with four speakers from the Greater Toronto Area (GTA) Tamil community, the community I am part of. Two of the speakers were heritage speakers (DS and AB) and the other two were L1 speakers whose speech served as a baseline (AS and RB). This section provides a summary of their language backgrounds as was shared with me through the language background questionnaires and sociolinguistic interviews described in the previous section.

Rather than strict ages, the selection criteria for choosing the heritage speaking participants was based on the age of shift to bilingualism; I worked with speakers who were

either born in the GTA or who immigrated to the GTA from India before adulthood, i.e. before or while they were still in school.

The heritage speakers I worked with were self-motivated in either wanting to improve their oral proficiency in Tamil, or otherwise motivated in using the language in their daily life. It is difficult to accurately evaluate the proficiency of a heritage language speaker, because as mentioned in section 2.4.2, heritage language learners demonstrate a range of proficiencies across the various dimensions of language competencies. I decided to work with heritage language learners that have similar levels of proficiency across a specific set of competencies that for the most part match my own production and comprehension proficiencies, as listed in Table 5. Note that these aren't meant to reflect any specific external "score" of proficiencies, but are instead meant to be descriptive of the proficiencies of myself and the speakers I worked with. I chose to work with speakers who fit within this range of proficiencies for simplicity (since they match my own proficiencies), but also because I found in conversations with my heritage language speaking peers who fit within this range that they were the ones who felt especially frustrated at being "stuck" at a proficiency they felt was challenging to improve from. As such, I hoped that this work could be useful for them. Since I worked with people from my community that I already knew, I was able to informally assess if they fell within this range of proficiencies, and invited them to participate accordingly. Future research would need to check for this criteria in a more rigorous manner.

As mentioned in 3.1, it just so happened that all of the speakers I worked with spoke essentially the same *vattara vazhakku* as myself, which made it easier for me to recognize forms that are markers of this dialect. I want to reiterate that I don't intend to privilege my/our dialect through this work, or falsely present this variety as a more "standard" variety.

Dimensions of Tamil language competence	Proficiency
Oral comprehension	Can understand colloquial speech in the contexts of conversations between family and friends, as well as popular media
Oral production	Can hold an informal conversation in their own dialect, but might have to code-switch into English for more complex verb structures or to fill vocabulary gaps

Table 5: Tamil oral language proficiency requirements of heritage speaking participants

The two heritage speakers I worked with, DS (aged 13) and AB (aged 14), were raised in the GTA. DS was born in Canada, while AB was born in India but immigrated with her family to Canada around the age of two. Both speakers were between the ages of 13 and 18 and are attending school in English. Both report learning and being exposed to English and Tamil from the same age, and so both can be considered simultaneous bilinguals. Both of their schooling has been in English to date, with the addition of French as an L2 (as is required in Ontario.) AB comprehends but doesn't speak Hindi. DS briefly attended both Tamil and Sanskrit community-run classes when she was younger, but felt she didn't learn much in the former apart from nursery rhymes. Neither DS or AB currently attend any Tamil language classes.

DS and AB are both being raised by L1 Tamil-speaking parents who are themselves multilingual in English and other Indian languages. DS shared that her parents speak Tamil, Hindi, and English, but primarily use Tamil in the home. AB said her family spoke a mix of English, Telugu, Hindi, and Tamil, and couldn't gauge between whether Tamil or English was used most at home. Both speakers reported that the primary domain in which they heard and spoke Tamil was in the home, especially in familiar, daily contexts like food.

I asked both heritage speakers to self-assess their own proficiency. Neither can read or write in the language, and both reported varying degrees of oral proficiency. AB felt that she could understand the language in familiar contexts but couldn't speak.[30] DS felt that she could understand most conversations, and though she found some pronunciations tricky, nevertheless felt confident she could speak enough "to get by" in India.

DS expressed that she associates Tamil with elder family members. She clarified that "elder" didn't include anyone in her own generation; indeed, she didn't speak as much Tamil with her elder sibling because she felt doing so would feel like she was emotionally distancing them rather than viewing them as a friend. She mentioned having Tamil friends, and not speaking much Tamil with them for the same reason. At the same time, she emphasized that the language she felt most comfortable in is "Tanglish" (a common portmanteau to describe the mix of English and Tamil), and that she would occasionally slip into this mixed language with friends once she felt comfortable enough around them. She expressed a strong affinity for "Tanglish" as being a special and unique marker of her identity.

Tamil is an intrinsic part of DS's roots and cultural identity. She feels that while she uses English in public, Tamil for her is more private and therefore something she mainly speaks at home. She emphasized that this is not related to feelings of shame or language inferiority, but rather to feelings of comfort. Home for her is the safest learning environment, a space free from being judged. While she is comfortable with her current proficiency, she expressed an interest in more explicitly learning verb conjugation. She envisioned this happening within the home environment, rather than a classroom.

AB speaks Tamil with her family in the home (parents and elder sibling), and occasionally with one close friend. Her primary use of Tamil is when speaking to her family

[30] Anecdotally, I've had conversations with many heritage speakers who will say that they "don't speak" the language, despite having measurable oral production skills. This feeling is likely a result of lack of confidence, rather than true lack of ability.

living in India. Conversations tend to be about simple, every-day things, and not any complex topics. She expressed that Tamil is important to her as a marker of cultural identity and connection to her roots. In addition to Tamil, she said she felt inclined to learn to speak Hindi to connect with her Indian culture and with her peers who spoke Hindi.

The baseline speakers AS and RB are the parents of the heritage speakers DS and AB, respectively. I chose AS and RB as baseline speakers deliberately to provide me with a clear indication of the type of language input the heritage speakers must have received and be receiving in the home; this proved to be an influencing factor in the analysis, specifically when looking at the heritage speakers' phonology in 4.5. The baseline speakers were both born and raised in India and moved to Canada in their late 20s or early 30s. They both currently work in English-speaking workplaces, but continue to speak Tamil in the home, with family back home in India, and with friends. Both identify Tamil as one of their mother tongues, and continue to be multilingual in English and several other Indian languages.

3.4 Tasks

My second session with each participant consisted of a set of specific language tasks. First I asked each participant to retell an oral narrative using a storyboard that was designed to elicit different types of compound verb forms. The goal of this activity was to test for proficiency of these targeted forms, but also to get a general sense of the participant's language abilities that might not otherwise be possible in a more restricted task. Then I asked the participant to complete two picture description tasks that were designed to elicit production of PNG agreement morphemes and tense allomorphs respectively.

Having separate tasks that looked specifically at just tense and PNG helped isolate which divergent forms were connected to differences between the heritage speakers' and L1 speakers' morphological knowledge versus those that were a result of the morphophonological processes that arise more in compound verb forms.

I initially also included a picture matching task adapted from Wagner (2002) that was designed to assess speakers' comfort with the completion entailment of the perfective (question three in 2.5)[31]. I ended up not using the results from this task because I didn't adequately take into account the different semantic shades and morphosyntactic properties of this morpheme (described in more depth in Annamalai [2019]; Paramasivam, [1979]) by erroneously simplifying it to being merely a perfective aspectual marker; this made the task confusing, even for the L1 speakers. Because of this I chose not to include the results from this task in my analysis.

The following sections detail the storyboard (3.4.1) and picture description (3.4.2) tasks that took place during the second sessions with each participant.

3.4.1 Oral narrative storyboard task

Oral narratives, as a type of semi-structured elicitation, are an effective way to get a general sense of a speaker's language abilities, as well as elicit freer speech (Eisenbeiss, 2010). For this reason, narratives and stories are commonly used in language acquisition studies (Polinsky, 2008; Slobin & Berman, 1994). For this project, I used storyboards to implement this task.

Storyboards are stories told through a series of pictures with no text, often with a surprise ending. Traditionally, in fieldwork settings, the linguist tells the story first a couple of times using the aid of the pictures, and then asks the consultant to tell the story back to them in the target language, with or without the aid of the pictures as they prefer. The idea behind this exercise is that the consultant is likely to be more focused on the act of telling a story than on their language, so it should elicit more natural speech (Burton & Matthewson, 2015).

[31] To briefly illustrate, one of the tasks was to listen to me saying a sentence like "I drank water" in Tamil, and based on whether the sentence contained a perfective or not, correctly match it to either a picture of an empty glass of water or a half-empty glass of water.

I used storyboards from the website Totem Field Storyboards (http://totemfieldstoryboards.org/), which was created by linguists, community members, educators, and students. It is licensed under a Creative Commons license that allows the storyboards to be adapted and used freely. A benefit of using these storyboards is that they are each clearly constructed to target specific grammatical forms. What this means for my purposes is that I was able to filter through the storyboards to pick one that would elicit a range of different compound verb forms. I did this by using a storyboard designed to elicit modals. The reason I picked modals is because, other than the two finite forms of hortative and permissive, most modals in Tamil necessarily need to be part of compound verb forms; they are either non-finite verbs that need to be accompanied by a finite predicate, or they are modal auxiliaries which are by definition part of compound verb forms. In this way, I hoped to elicit a large number and variety of different compound verb forms that were not likely to occur in the picture description and picture matching activities discussed later in this section.

I used the storyboard "Feeding Fluffy", which I anticipated would elicit modals including expressions of possibility and obligation/desire, conditionals, and the perfect and progressive aspects. Ultimately I didn't end up eliciting many compound verb forms. I suspect this was a result of individual speaker variation rather than a direct function of the task itself as I have used this storyboard often with my students, many of whom tell the story using compound verb forms.

When I facilitated the storyboard task, instead of telling the story to the participants first, I adopted a similar method to the one used by Polinsky in her 2008 study of heritage language narratives working with heritage Russian speakers. I provided the participants with the storyboard and asked them to look through it for as long as they needed and to then retell the story in Tamil whenever they were ready. Following this, I asked the heritage speakers to retell the story in English. DS retold the story in English immediately after telling

it in Tamil, whereas AB retold the story in English in a separate session a few days later. The baseline speakers were only asked to tell the story in Tamil.

By having the participants look through the story first themselves rather than have it read to them, I avoided a) priming the participants for English if it was told to them in English, and b) having them memorize parts of the story if it was told to them in Tamil. The downside to this approach was that each participant interpreted the story slightly differently, and sometimes not how I had intended them to, which meant they didn't produce all the forms I was hoping to elicit. In addition, the heritage speakers both expressed confusion in understanding the story from just the pictures, and so it was difficult to disentangle grammatical errors from miscomprehension of the story itself.

Having the participants retell the story in both Tamil and English did nevertheless allow me to compare both and better evaluate whether potential issues that arose in the Tamil retelling were due to language difficulties or something else (e.g. misunderstanding of the story). However, it would be incorrect to assume that the two retellings were direct translations of each other, or even that the retellings didn't influence each other as a sort of task repetition effect. Indeed, during her second retelling, DS remarked that she understood the story only after telling it in English, and that she "didn't get the story before." (Both baseline speakers, on the other hand, expressed no confusion when telling the story in Tamil.) In addition to a quasi-task repetition effect, differences between the English and Tamil retellings could also have been conditioned by differing language fluencies or by cultural code-switching conditioned by the participants "thinking in English" vs "thinking in Tamil". This isn't a variable I could precisely account for, but the hope is that the effect is incidental in context to the broader goals of the analysis.

For future research, it may be advisable to tell the story to the participants once in English or Tamil,[32] so as to mitigate the risk of not understanding the story. In the current study, I counterbalanced this risk by supplementing the storyboard task with a selection of picture description tasks, described in the next section, that elicited target forms in a more structured and consistent way.

3.4.2 Picture description tasks: memory game

While there are many benefits to the free style of speech that is produced with narrative tasks, the restriction in topic and content can be a drawback if participants don't produce certain forms that are of interest. In this situation, having follow-up elicitation tasks that are more controlled can be beneficial (Montrul, 2015). An example of this is the picture description task that was used in Montrul & Sánchez-Walker (2013). Montrul describes how in this task, participants were asked to construct sentences to describe images they were presented with. I adapted this picture description task to look at heritage speakers' production of PNG agreement morphemes and tense allomorphy, by reconfiguring it into a game that is a cross between "memory" and "battleship", shown in Figure 1.

[32] I imagine the risk of telling the story in English would be priming the participant for English, and the risk of telling the story in Tamil would be participants simply repeating what they have just heard, and thus not providing an accurate representation of their own language.

Figure 1: Memory game – PNG agreement

The goal of this game is to find matching pairs. As in the classic “memory” game, the player uncovers two cards and scores a point if they are a match. The twist I added in this version is that instead of allowing the player to uncover the cards on their own, I control the board and they have to tell me which card they would like me to uncover. They do this by creating a sentence with the images that head the row and column of the card that they want to uncover. The rows are meant to represent different combinations of PNG agreement (3F, 3M, 3HON and 3PL respectively.) The columns represent different commonly used verbs (drinking water, singing, playing (a game), going to school, and eating). For example, if the participant wanted to uncover the card that is at the intersection of the first row and first column (Figure 2), then they would have to construct the sentence “she is drinking water” in Tamil. Uttering this sentence in any other tense or aspect would also be fine, as long as the appropriate PNG agreement was present.

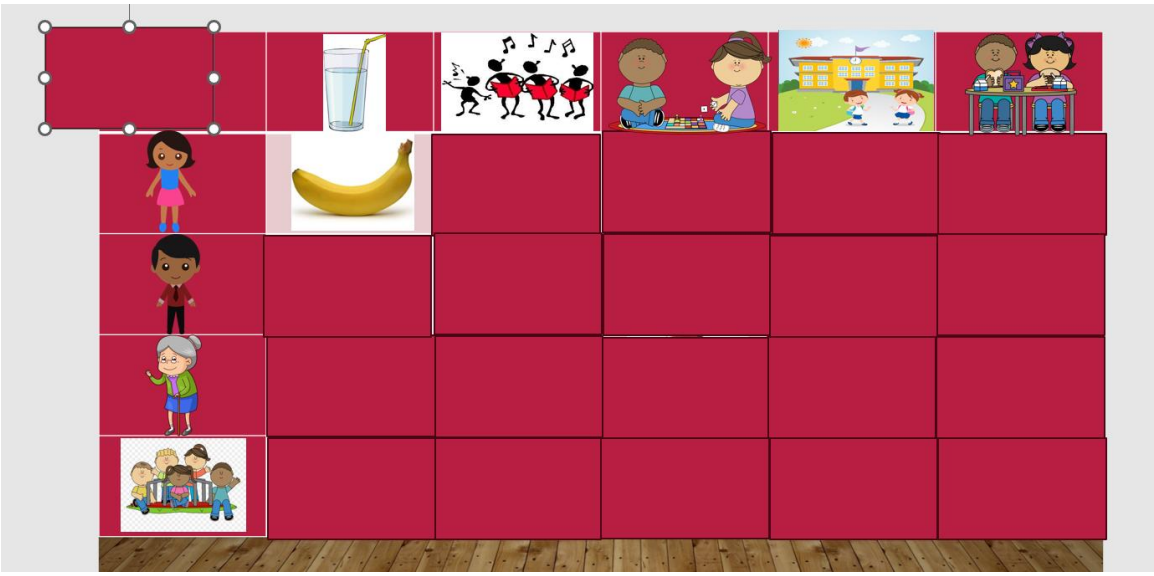

Figure 2: Playing the memory game

In addition to the PNG-agreement board, I also designed two boards to target tense allomorphy. The tense boards contained verbs from each of the seven different verb classes as the columns, and the different tenses (past, present, future) represented using images for "morning", "now", and "night" as the rows. Participants were instructed to conjugate these verbs in the first person. The tense boards were initially designed to be one large board, but following feedback from trial runs of the task with a couple of L1 speakers, they were split up into two boards to reduce cognitive load. Both boards needed to have four verbs (3 rows x 4 columns = 6 pairs), and so one of the verb classes had to be represented twice. I chose verb class III, as this allowed me to include the commonly used verb *pO* 'go'.

The memory game boards were designed to represent verbs from all seven verb classes (described in Table 1 in 2.2.2.) The verbs elicited in the PNG-agreement board (shown above in Figure 1) are *kuDi* 'drink' (verb class VI), *pa:Du* 'sing' (III), *viLaiya:Du* 'play' (III), *po:* 'go' (III), and *sa:ppiDu* 'eat' (IV). The first tense board elicits the verbs *azhu* 'cry' (I), *vizhu* 'fall' (II), *pa:Du* 'sing' (III), and *sa:ppiDu* 'eat' (IV). The second tense board elicits the verbs *ke:L* 'ask' (V), *kuDi* 'drink' (VI), *naDa* 'walk' (VII), and *po:* 'go' (III). I've included a picture of each of the boards used in Appendix A.

I have used the memory game activity in the Tamil class I teach and the students have found it very engaging. I have found it an effective way to practice PNG paradigms without it being tedious and repetitive for the students. Similarly, the game format helps reduce any anxiety the participants have producing the target forms, and makes it feel less like a test. Participants mentioned enjoying the activity, and DS even said that it was a fun way for her to practice her Tamil. Another benefit of this activity is that saying the target sentences more than once is built into its mechanics (because of the requirements of the game that entails flipping over the same card more than once). This provides valuable insight on distinguishing between utterances that were performance errors (which may improve or go away with repetition) vs genuine instances of divergent grammar development.

There were two limitations of the tense storyboards specifically that emerged during the sessions. The first was that the verbs *azhu* "cry" and *vizhu* "fall" were challenging to the heritage speakers because they are less commonly used verbs, and because of the retroflex rhotic *zh* present stem-finally in both. This phoneme is articulatorily marked in Tamil; both informally as a speaker myself, speaking with my peers, and through working with the students at the Tamil school, I have noticed that this is one of the trickiest phonemes to learn to articulate.[33] As discussed in 2.4.1, Lakshmi Bai (2000) demonstrated that this is one of the last phonemes mastered by a child acquiring Tamil as an L1[34]. Indeed, this phoneme stabilized for the children in her study only between the ages of 5 and 6, which is around when the heritage speakers I worked with would have started switching over to English, thus making it unlikely they would have fully mastered it. All of this to say, using these verbs made the task trickier for phonological rather than morphological reasons, and so was a limitation

[33] From a sociolinguistic perspective, it is interesting to note that this phoneme, written using the Tamil character ழ, is an extreme source of language pride among Tamils because it is perceived both as being unique as well as difficult to articulate. At the same time, in many dialects, most prominently the Sri Lankan one, this phoneme has merged with /l/.

[34] It would be instructive in a future study to determine if this phoneme is articulatorily complex for either L1 or heritage Tamil speakers, following the example of Patience (2022) with L2 Spanish.

in terms of providing an accurate read on whether challenges the speakers faced in its production were a result of its phonology or the morphology of the forms it was a part of.

Another general limitation of both the tense storyboards is that speakers didn't always distinguish tense the way I had intended; technically, using either the past or the present tense to speak about all three times of day could be grammatically (and pragmatically) appropriate, and so speakers frequently used the same tense across the different times. This happened more with the baseline speakers than the heritage speakers, but did happen across the board.[35] What this meant for the data was that I wasn't getting an accurate sample of whether speakers had mastered each of the three different tense allomorphs equally because they didn't produce them all in equal proportions to each other (though they did produce enough of each for analysis). On the other hand, it was a good indication of pragmatic competence in the speakers.

An obvious limitation of the memory board activity is that the boards I designed only elicited verbs that agreed with 1SING, 3F, 3M, 3HON, and 3PL. My analysis is therefore limited as such. The results still proved to be of value because gender distinctions only appear in the 3rd person anyway, and the persons included are some of the most commonly used in communicative contexts. Additionally, verbs from class III are definitely overrepresented in the PNG-agreement board, as I was prioritizing using familiar, commonly used verbs. In a future iteration, I would make sure to have more verb classes and persons represented across all the boards.

3.5 Conventions for data analysis

3.5.1 Target vs non-target: a note on terminology

The goal of this work is to identify what areas of Tamil grammar, specifically of verbal morphology, heritage speakers struggle with. In order to identify these areas, I compare

[35] DS even commented on it at one point, expressing concern that she had used the wrong tense than I had intended at times.

heritage speakers' speech to a baseline of fluent L1 (non-heritage) speakers, following in the vein of "comparative logic" as advocated for by Dominguez & Arche (2021) in the SLA context and described by Polinsky & Kagan (2007) in the context of describing a continuum of heritage language acquisition and proficiencies. This terminology is not meant to imply that the heritage speakers' speech is inferior to that of first language speakers. The intention is not to measure the heritage speakers' speech against that of monolinguals (note that both baseline speakers are multilingual), and neither do I want to contribute to the feeling of linguistic insecurity that many heritage speakers (including myself) experience as a result of language purism ideologies (Ortega, 2019). Rather, the terminology I'm using here is meant to be in recognition of the fact that the heritage speakers' speech *differs* from and is primarily influenced by the L1 speakers they grew up with. As discussed in 3.3, I deliberately worked with parent-child pairs to limit the possibility that differences in speech are a result of dialect variation. Throughout this thesis, I use the term "target form" to refer to forms used by this baseline of L1 speakers (RB and AS as described in 3.3), and "non-target form" to refer to forms that differ from those used by L1 speakers. Some forms may be target-like in one homeland dialect of Tamil but not in another; where the heritage speakers use these forms, I label them as target-like and elaborate on the dialectal variation in the accompanying prose. Again, the terms "target" and "non-target" are not intended to carry a value judgement; rather, they are terms meant to distinguish between forms that are common to heritage and homeland varieties (target-like), and forms that are unique to heritage varieties (non-target). I use the term "target" because of the assumption that the target forms for heritage speakers are the forms they hear the L1 speakers in their community use, and are the forms learners are generally trying to replicate. At the same time, it will be shown that the "non-target" forms are often representative of heritage Tamil as its own variety of the language, distinct from the variet(ies) spoken by L1 speakers.

3.5.2 Transcribing

The analysis of the data was based on transcriptions of the sessions. The sessions were recorded on Zoom. I did not transcribe the sociolinguistic interviews, instead just took notes and drew on the information provided through these and the background questionnaires to build my description of the speakers in 3.3. I transcribed the storyboard, picture description, and picture matching tasks for each participant in ELAN.

There were some general conventions I followed when transcribing speakers' speech. As the focus of my analysis is largely morphological rather than phonetic, I deliberately adopted a broad transcription style[36] and avoided detailed phonetic description. Geminates were not consistently transcribed; instead, a voicing distinction was included (consonant devoicing is a cue for gemination.) Alveolar and trilled rhotics (*r* ɾ and *R* r̥ respectively) are merged outside of geminate contexts in nearly all varieties of *koDuntamil* and so these phonemes were not distinguished or subject to analysis; as a default I transcribed both phonemes as *r* (the alveolar rhotic) in surface forms, though I retain the distinction when transcribing underlying forms. Some vowels were pronounced using a non-standard allophone; these weren't analysed further, but I included a note in the transcription to describe their vowel quality.

As was discussed briefly in 2.2.5, word-final nasal consonants often drop and the preceding vowel becomes nasalized in *koDuntamil*. I have transcribed nasal vowels as if they were nasal consonants following the vowel, in order to make the underlying phonology transparent.

Some challenges of transcription were distinguishing between retroflex and alveolar laterals and nasals. While retroflexion is still a phonemic feature in Tamil, retroflex consonants are starting to merge with their alveolar counterparts in some varieties of

[36] Rather than using IPA, transcriptions in this thesis follow the convention of Tamil linguists writing in English and instead are written using a Roman-alphabet based writing system that is more intuitive to Tamil speakers. IPA equivalencies are included in Appendix B.

koDuntamil; as such, I decided not to analyse these distinctions in the speech of my participants and transcribed these sounds according to their underlying phoneme.

Another challenge was accurately perceiving word-final nasal vowels, specifically in the heritage speakers' speech. It was often very unclear if an expected word-final nasal was present or not. While this issue could definitely be attributed in part to the fact that the interviews were conducted over Zoom and so the audio was not as clear as it would have been in person, it is also worth noting that this ambiguity was not present with the baseline L1 speakers, whose interviews were also conducted over Zoom. Thus, either a) the word-final nasal vowels genuinely were sometimes missing in the heritage speakers' speech, b) the nasality was acoustically more pronounced in the baseline speakers' speech, or c) the baseline speakers used other phonological cues associated with nasal vowels that the heritage speakers didn't. Future research, ideally conducted in person and with high quality recording equipment, would be valuable in determining which explanation is most likely.

As a general rule, if there appeared minor phonetic variations in the speech of my participants, but they were not relevant to the target morphology, I transcribed the speech according to the underlying phonology of the target forms. Again, the focus of my analysis is morphological and not phonetic, and so while the data collected could have been used to analyse pronunciation, these distinctions were decided to be less relevant for this work.

3.5.3 Tagging

I broke up the transcriptions for each speaker by clause in ELAN, and for each clause, tagged whether they contained non-target forms or not. From here I exported the tagged transcription files into Excel and continued tagging there.

For every clause that contained a non-target form, I tagged in a separate tier what aspect was non-target-like (morphological, phonological,[37] morphophonological,

[37] As mentioned before, I did not focus on pronunciation, so this tag was just to track instances of wholesale segment replacements of one phoneme for another.

morphosyntactic, syntactic, other). I also tagged whether it specifically contained a non-target-like verb form.

For every clause that contained a verb (which for both speakers, was over 85% of all Tamil utterances), in separate columns I recorded a) whether the verb was target-like, b) the full verb form produced by the speaker, c) the stem of the target verb, d) the target tense, and e) the target PNG form. Where the verb form was non-target-like, I included separate columns to tag whether the a) verb stem, b) verb tense, and c) verb PNG were target-like. I additionally recorded a) the speaker's produced verb stem, b) the speaker's produced tense, and c) the speaker's produced PNG. I included a separate columns to track any additional observations that didn't fit within these categories.

3.5.4 Inter-coder reliability

In addition to the four speakers described in 3.3, I worked with a fluent L1 Tamil speaker, Vasumathi Srikanth - my mother and a teacher at the heritage Tamil school I volunteer at (previously mentioned in 1.2 and 3.1) - to assist in the identification of non-target-like forms. She listened to the recordings of the storyboard, picture description, and picture matching tasks. Then in an Excel spreadsheet, she identified the time stamps where she identified non-target-like forms, recorded what the non-target form was, and what she thought the intended target form likely was. The analysis in chapter 4 is built around the non-target-like forms identified by myself and my mother.

In total, I identified 87 non-target-like forms, of which 62 involved non-target-like verbal morphology. In total, my mother identified 68 non-target-like forms. We had differing interpretations for 14 tokens. For each of these, we reviewed the forms together and came to an agreement on each of them.

Of the 62 tokens that I had tagged as having non-target-like verbal morphology, all but 7 were also tagged as such by my mother. For three of these differing tokens, my mother told me that the forms were target-like in dialects that the speakers (heritage and L1) spoke,

and so I revisited these forms to tag them as target-like instead. For the remaining four tokens, I had not perceived an expected word-final nasal vowel, but my mother had. In reviewing these tokens, we both just acknowledged that neither of us were really able to definitively say if the nasal vowel was present and were just making our best guess. As discussed in the previous section, because of the uncertainty around word-final nasal vowels, they were not included in the analysis.

Of the 68 tokens my mother had tagged as non-target-like, I had also identified all but 7. For three of these 7, my mother had identified instances where the 3rd person plural PNG was marked with the dialectically-marked *-a:* form, instead of the *centamil* form *-a:nga(L)*. Since differences stemming from dialectal variation is not the focus of my study, I opted to ignore these and leave them tagged as target-like. The remaining four tokens my mother identified that I didn't were where the forms appeared to be semantically anomalous because of the context set up by the game; but otherwise were target-like forms. For instance, a sentence that translates to 'I run in the night' is grammatical, even though in the context of the game I was hoping to elicit the future form 'I will run in the night' instead. For these forms as well, I left them tagged as target-like.

3.6 Data

3.6.1 Verb types produced by DS and AB

Table 6 lists every verb that DS used, grouped by which verb class it belongs to, and Table 7 presents the same for the verbs used by AB. Recall from 2.2.2 that the verb class a verb belongs to determines which set of tense allomorphs it carries.

The verbs that were deliberately elicited as part of the "memory game" picture description tasks are highlighted in bold. The others were produced either as part of the storyboard task or incidentally during or between, but not related to, the tasks.

Verb Class*	Verb	# of tokens	Total # of tokens in verb class
I	**azhu** *(cry)*	6	6
II	va: *(come)*	2	9
	vizhu *(fall)*	7	
III	**pa:Du** *(sing)*	14	55
	(English verb) + paNNu *(do)*	4	
	pe:su *(speak)*	1	
	po: *(go)*	17	
	sol *(say)*	7	
	viLaiya:Du *(play)*	12	
IV	**sa:ppiDu** *(eat)*	18	18
V	**ke:Lu** *(ask)*	11	11
VI	**kuDi** *(drink)*	15	16
	pa:ru *(look / watch)*	1	
VII	iru *(be)*	5	14
	mara (*forget*)	1	
	naDa *(walk)*	8	
***Dative subject verbs**	muDiyum *(able to)*	1	4
	therile *(don't know)*	2	
	theriyum *(know)*	1	
Total			**133**

Table 6: Verbs produced by DS

Verb Class*	Verb	# of tokens	Total # of tokens in verb class
I	**azhu** *(cry)*	9	9
II	va: *(come)*	1	11
	vizhu *(fall)*	10	
III	**pa:Du** *(sing)*	16	47
	(English verb) + paNNu *(do)*	1	
	po: *(go)*	20	
	sol *(say)*	1	
	va:ngu *(buy)*	1	
	viLaiya:Du *(play)*	8	
IV	**sa:ppiDu** *(eat)*	17	17
V	**ke:Lu** *(ask)*	9	9
VI	**kuDi** *(drink)*	16	19
	kuDu *(give)*	1	
	ninai *(think)*	2	
VII	iru *(be)*	6	11
	naDa *(walk)*	5	
***Dative subject verbs**	piDikkum *(like)*	1	2
	theriyum *(know)*	1	
Total			**125**

Table 7: Verbs produced by AB

The last three verbs in Table 6 and the last two verbs in Table 7 (“Dative subject verbs”) are technically part of classes VI and VII but have not been grouped with them because they are

extremely common forms that appear as memorized "chunks" in heritage speakers' speech (as defined by Polinsky & Kagan [2007]). In addition, they don't tend to carry any inflections other than the ones shown here. This is because these verbs are "defective verbs": they cannot occur with every tense, PNG, and non-finite suffix (Lehmann, 1989). These four verbs in particular cannot inflect for any PNG suffixes other than 3N and can only appear with a dative subject (or instrumental, in the case of *muDiyum*). *MuDiyum* is a grammaticalized auxiliary formed from the lexical verb *muDi* (to finish, complete) and the fusional FUT+3N suffix *-um*. *Therile* and *theriyum* are the negative and positive inflections for the lexical verb *theri* (to know),[38] carrying the negative suffix *-illai* and the fusional FUT+3N suffix *-um* respectively. *piDikkum* is a grammaticalized auxiliary formed from the lexical verb *piDi* (to like) and the fusional FUT+3N suffix *-um*.

Knowledge of verb class is not required to form these verbs, because the suffixes they take attach directly to the verb stem. The forms included here are ubiquitous and likely to be memorized by heritage speakers, and so including them in the analysis would misrepresent speakers' actual knowledge of verbal inflections. Note that this is why the total number of tokens in Table 6 is 133 while the total number in Table 8 in the following section is 129, and the total number of tokens in Table 7 is 125 while the total in Table 9 in the following section is 123.

3.6.2 Verb types and tense morphology produced by DS and AB

Table 8 below details the verbs DS used and with which tense inflection, and Table 9 presents the same for the verbs used by AB. As described in 3.4.2, some of the elicitation contexts pragmatically allowed for more than one tense, and so it was not possible to always determine which tense the speaker meant to use. I opted to analyse the forms according to their morphology rather than their intended semantics, even if the forms were not entirely

[38] It also means "to see"; but that meaning is not relevant to this discussion.

target-like. This was simple to do as all of the tense morphemes have very distinct forms. For example, if it was unclear from context if the speaker intended to convey a future or present meaning in an utterance, but the tense allomorph she used clearly matched a future tense morpheme (even if not entirely target-like), I would treat it as an instance of a verb carrying a future tense inflection.

In some instances, the speaker would start a form, and then change the form part-way through. I have coded these tokens based on the final utterance she produced, but included the partial forms in the transcription of the examples as they make the analysis more transparent. In other instances, the speaker produced a complete form and then immediately followed it up with a self-correction. In these cases I have coded both utterances separately, but when presenting the examples individually I make sure to note that the speaker made a self-correction.

Some tokens were compound verb forms, and in these cases I have included the token under the category of the lexical verb, with the following auxiliary verb (that carries tense and PNG morphology) and its verb class in brackets after it. I have also marked instances where the verb carried a fused 3rd person neuter and tense morpheme.[39]

Some of the tokens produced by AB were non-finite forms that headed their own dependent clause, with a following finite verbal main clause (different from compound verb forms). These are counted in a separate column as “Non-Finite”.

Finally, some utterances were only partial forms, either because the speaker started saying something and then changed her mind mid-sentence, or because she was sort of “testing out” saying the word to herself quietly first before committing to a form. These utterances are counted in a separate column as “Partial Forms”.

[39] For all other persons, PNG and tense always remain distinct morphemes.

Verb Class	Verb	Past	Present	Future	Partial Forms
I	azhu *(cry)*	1	1	4	
	# of tokens/tense:	**1**	**1**	**4**	
II	va: *(come)*	1 (*a:ku*, III) (3N)		1 (3N)	
	vizhu *(fall)*	4 (*viDu*, IV)		2	1
	# of tokens/tense:	**5**	**0**	**3**	**1**
III	pa:Du *(sing)*		12	2	
	(English verb) + paNNu *(do)*	1 (*viDu*, IV)	3		
	pe:su *(speak)*		1		
	po: *(go)*		12	5	
	sol *(say)*	1 (*viDu*, IV)	5	1 (*iru*, VII)	
	viLaiya:Du *(play)*		12		
	# of tokens/tense:	**2**	**45**	**8**	
IV	sa:ppiDu *(eat)*	2	8	8	
	# of tokens/tense:	**2**	**8**	**8**	

V	ke:Lu *(ask)*	2 (*viDu*, IV)	4	5	
	# of tokens/tense:	**2**	**4**	**5**	
VI	kuDi *(drink)*		13	2	
	pa:ru *(look / watch)*		1		
	# of tokens/tense:	**0**	**14**	**2**	
VII	iru *(be)*	1	4 (3N)		
	naDa *(walk)*	0	1	5	2
	mara *(forget)*	1 (*po*, III)			
	# of tokens/tense:	**2**	**5**	**5**	**2**
Total # of tokens / tense	**129**	**14**	**77**	**35**	**3**

Table 8: Verbs produced by DS and their tense morphology

Verb Class	Verb	Past	Present	Future	Non-Finite	Partial Forms
I	azhu *(cry)*	6	1	2		
	# of tokens/tense:	**6**	**1**	**2**	**0**	
II	va: *(come)*	1				
	vizhu *(fall)*	6	3	1		
	# of tokens/tense:	**7**	**3**	**1**	**0**	
III	pa:Du *(sing)*	3	11	2		
	(English verb) + paNNu *(do)*				1	
	po: *(go)*	6	12	2		
	sol *(say)*	1				
	viLaiya:Du *(play)*	1	7			
	va:ngu *(buy)*	1				
	# of tokens/tense :	**12**	**30**	**4**	**1**	
IV	sa :ppiDu *(eat)*	4	12	1		
	# of tokens/tense :	**4**	**12**	**1**	**0**	
V	ke :Lu *(ask)*	6	2	1		
	# of tokens/tense :	**6**	**2**	**1**	**0**	
VI	kuDi *(drink)*	4	9	2		1
	kuDu *(give)*	1				
	ninai *(think)*	2				
	# of tokens/tense:	**7**	**9**	**2**	**0**	**1**
VII	iru *(be)*		4	2		
	naDa *(walk)*	2	1	2		

	# of tokens/tense:	2	5	4		
Total # of tokens / tense	123	44	62	15	1	1

Table 9: Verbs produced by AB and their tense morphology

There is large variation in the distribution of tenses produced by each speaker. This is largely a function of the tasks performed; an event that occurs in the past or future can frequently still be talked about using a present tense morpheme, and so we see that it is overrepresented compared to the other tenses. Despite the relatively fewer number of past tense tokens (for DS) and future tense tokens (for AB), the patterns that emerged from them were still very clear and worthy of discussion in the analysis.

3.6.3 Verb types and PNG morphology produced by DS and AB

Due to time limitations, this study does not attempt to provide a full analysis of heritage speaker PNG morphology. In particular, the 2nd persons are not represented in the data at all, and so I leave a comprehensive discussion of this aspect of the heritage grammar to a future study.

Table 10 details which PNG morphemes DS produced and how many times she produced them, and Table 11 does the same for AB. The second row in Table 10 is matched with both the first singular and plural because there were two instances where DS produced what was phonologically the 1st plural form, but it was unclear from context whether she meant to refer to a singular or plural subject. The rows matched with the third person feminine and honorific forms in both tables refers to instances where the subject of the sentence was a grandmother, and so the intended subject could feasibly have been either. Either way, the target form used by both speakers would have been the same. Finally, the

second last row in both tables refers to instances where it was unclear whether DS meant to refer to a male or female third person subject.

Person	# of tokens
1S	48
1S / 1PL.INC[40]	2
3SN	16 ±1
3F	18
3F/3HON	11
3M	17
3F or 3M	2 ±1
3PL	12

Table 10: Number of PNG morphemes produced by DS per person type

Person	# of tokens
1S	63
3SN	10
3F	18
3F/3HON	11
3M	7
3F or 3M	2
3PL	8

Table 11: Number of PNG morphemes produced by AB per person type

[40] There was only one possible instance of the first-person plural (plausible from both form and context). I never directly elicited it. DS spontaneously produced it once, but I can't be sure whether she intended a plural or singular meaning.

4 Analysis

This work was designed to explore heritage speakers' production of verbal morphology, and the analysis will be restricted as such. The bulk of the analysis is focused on the speakers' morphological realizations of tense (4.1 and 4.3) and person-number-gender agreement (sections 4.2 and 4.4), with most of the focus being on the former. Following this is a brief discussion of observed phonological variation (4.5). I look at each speakers' speech separately to represent them and their language as holistically as possible, and then compare the emergent patterns from both of them in chapter 5. I will not be including analysis or discussion of any syntactic variation present in the participants' speech as the data was not well-suited to investigating this. I hope that this could instead be an avenue of research for others or myself to pursue in future.

In the following chapters (and throughout this thesis) I use the term "non-target-like" to refer to forms and structures produced by the heritage language speakers that differ from how homeland L1 speakers (represented by the baseline speakers RB and AS) would produce them (see 3.5.1 for discussion on why I use this terminology). When these differences are morphological (rather than purely phonological), they are marked in the glossed examples with the abbreviation NTARG.

Throughout this chapter, the patterns apparent in DS and AB's speech provide clear evidence that their grammars, heritage language grammars, are natural language grammars. In particular, though the structure may differ from that of the baseline L1 speakers, it will be shown that both speakers undoubtedly have knowledge of Tamil tense and PNG morphology.

4.1 Speaker 1 (DS) tense morphology

This section is a description and analysis of DS's tense morphology. 4.2 presents an analysis of her PNG morphology.

4.1.1 Overview of tense-marking strategies

DS used different strategies in how she expressed tense morphology on verbs. Table 12 summarizes the distribution of pronounced verbs in terms of which strategy DS used in expressing tense on them. Partial forms were not included here.

Overall, most of the time DS used a target-like tense allomorph, matching those used by the baseline L1 speakers. In the instances where she didn't, she used one of two strategies: she created a compound verb construction,[41] or she employed "tense-stacking". This latter strategy involved lexicalizing a tense morpheme (either the past or the present) and incorporating it into the verb stem, and adding another common / unmarked tense morpheme on top of this as an explicit tense suffix. I am calling this process "tense-stacking" to reflect the fact that the result is a double marking of tense: the lexicalized piece plus the additional suffix. This strategy will be described in more detail in 4.1.2. It is analogous to the process of overmarking that appears in other heritage languages (Polinsky, 2018a; Putnam et al., 2021) and in language acquisition more generally (e.g., the use of the form "childrens" by English learners, which double-marks the plural, in place of "children,") and is discussed further in 5.2.

[41] Compound verb constructions also contain tense suffixes, but the distinction is that the suffix appears on the (in this case, aspectual) auxiliary verb, rather than on the main lexical verb.

Verb class	Target-like tense allomorph	Tense-stacking	Compound Verb Constructions
I	6		
II	3	±4	1 ±4
III	52	±2	1 ±2
IV	18		
V	±4	5 ±6	±2
VI	±14	2 ±14	
VII	5	6	1
TOTAL	**85 ±18**	**13 ±26**	**10 ±8**

Table 12: Tense-marking strategies DS used with each verb class

It is important to note that, though this information isn't reflected in this more general table, which strategy she used was often dependent on which tense she was trying to express. This will be described in the following sections, which go into detail on DS's production of the future, present, and past tenses in turn.

4.1.2 Future

Table 13 shows what strategy DS used to form the future for each verb. The first column in the table lists the class of each verb in the second column, and in brackets a reminder of what the expected future tense allomorph for this class is. PNG and tense morphemes fuse when the verb agrees with the 3rd person neuter, so (3N) is included to mark when this was the case for a token. There was only one instance of DS using a compound verb construction, and the auxiliary verb she used and its verb class is included in brackets.

Cells are highlighted in green to mark where her strategy resulted in target-like forms. From the table we can see that DS consistently used the target-like future tense allomorph for verbs from classes I to IV, created a compound verb construction for one class

III token, and then used the strategy of tense stacking for verbs in classes V to VII. Note that classes I to IV all carry the *-v-* allomorph, in contrast to classes V to VII which don't - this will be relevant later.

Verb Class	Verb	target-like future allomorph	tense-stacking	compound-verb construction
I *(-v-)*	azhu *(cry)*	4		
II *(-v-)*	va *(come)*	1 (3N)		
	vizhu *(fall)*	2		
III *(-v-)*	pa:Du *(sing)*	2		
	po: *(go)*	5		
	sol *(say)*			1 (*iru*, VII)
IV *(-v-)*	sa:ppiDu *(eat)*	8		
V *(-p-)*	ke:Lu *(ask)*		5	
VI *(-pp-)*	kuDi *(drink)*		2	
VII *(-pp-)*	naDa *(walk)*		5	
Total	**35**	**22**	**12**	**1**

Table 13: Strategies DS used to mark the future tense

Nearly two-thirds of the time, DS used a target-like future tense allomorph when expressing the future tense. She used tense-stacking with about a third of the tokens, and in one instance, used a compound verb construction. I will discuss and provide examples of each of these three strategies in turn.

Strategies 1 and 3: target-like tense allomorph and compound verb construction

The majority of the time, DS used the strategy of using a target-like future tense allomorph. **(27)** is an utterance DS made using this strategy, where you can see her use the *-v-* allomorph to mark the future. As the verb here *pa:Du* 'to sing' is a class III verb, *-v-* is the target-like

future tense allomorph that we expect. In this utterance and several more upcoming, DS used the first-person plural PNG morpheme rather than the first-person singular (presented in the second column for comparison); this has no bearing on the actual tense morphology being discussed and so will instead be addressed towards the end of this section.

(27)	*Speaker utterance*[42]	*Target form*
Transc.ta	*night* ல நா பாடுவோம்	பாடுவேன்
Trans.la	night le na: pa:Duv**o:m**	pa:Duv**e:n**
MB	night le na: pa:Du-v-**o:m**	pa:Du-v-**e:n**
MG	night LOC I sing-FUT-**1PL**	sing-FUT-**1S**
Transl.	'I will sing at night' (intended)	

A target-like future form was also used in **(28).** In **(28)**, the verb *va* 'come' is conjugated with the 3rd person neutral. In Tamil, the future tense and 3rd person neutral morphemes are merged for all verbs, and so this is the target-like form we see DS using here.

(28)	*Speaker utterance*
Transc.ta	*then* அப்போ ஒரு *snake* வரும்
Trans.la	then appo: oru snake var**on**
MB	then appo: oru snake var-**on**
UR	then appo: oru snake var-**um**
MG	then at.that.time a snake come-**FUT.3N**
Transl.	'then at that moment, a snake will come'

[42] The complete speaker utterance is transcribed in column two and the word form under discussion is underlined. The specific morphemes being discussed within the word form are highlighted in bold typeface. Where the morphology of the speaker's form differs significantly from the target form of the underlined word and this difference is relevant to the discussion, the expected target form (in the speaker's specific *vattara vazhakku*) is presented for comparison in the third column.

There was one instance in which DS created a compound verb construction. Specifically, she created a future perfect construction using *iru*, the verb 'be', which when used as an auxiliary verb in a compound verb form represents the perfect aspect. She added this auxiliary to the main verb *sol* 'say', to create the form *sollirppe:n* 'I would have / must have said'. As described in 2.2.4, in compound verb forms such as this, tense is marked on (and thus must agree with) the final auxiliary verb. *Iru* is a class VII verb. This takes the *-pp-* future tense allomorph, which is what we see DS accurately producing in (**29**).

(29)	*Speaker utterance*
Transc.ta	[43][...] *நா English ல சொல்லிருப்பேன்*
Trans.la	na: English la sollir**pp**e:n
MB	na: English la soll-ir-**pp**-e:n
UR	na: English la sol-i-iru-**pp**-e:n
MG	I English LOC say- VBLP-be-**FUT**-1S
Transl.	'[…] I'd say it in English / I must have said it in English'

To sum, DS created target-like future tense forms for all verbs she produced that were from classes I to IV. For the most part, this meant using the future tense allomorph *-v-*. However there were two exceptions: once with the class III verb *va* 'come' that carried a fused 3rd person and future morpheme *-um*, and once with a future perfect construction that used the *-pp-* future allomorph on the perfect aspectual auxiliary *iru* 'be'.

The verbs *va* 'come' and *iru* 'be' are both common verbs that DS likely hears used daily. Not only are they commonly used lexical verbs, they also appear frequently in

[43] This example (and some future examples) includes some morphophonological reductive processes, as were touched on in 2.2.5. In all such examples, I've opted to include a line (MB) that shows the morphological breakdown of the surface *koDuntamil* forms in addition to a line (UR) that shows a breakdown of the underlying unreduced morphemes, for ease of comprehension.

compound verb constructions.[44] It's possible that these forms are memorized "chunks" (similar to the dative subject verbs discussed at the top of this chapter). More research, ideally through the form of a long-term study, would help clarify this.

Strategy 2: Tense-stacking

In nearly a third of her future tense utterances, DS used a strategy I am calling "tense-stacking", which I briefly described earlier in this section. To recap, this is a phenomenon where she produces a verb stem and follows it with what appears to be two tense morphemes: the first one has been lexicalized into the verb stem, and the second is used to semantically mark tense (in this case, the future tense). DS used the tense-stacking strategy to express the future tense with the verbs *ke:L, kuDi,* and *naDa*; class V, VI, and VII verbs respectively. While the surrounding phrasal context varied slightly with each utterance, the form she produced for each verb was always the same as the forms shown in **(30), (31),** and **(32)**.

First let us look at the future form DS produced for the class V verb *keL* 'ask', as shown in **(30)**.

(30)	*Speaker utterance*	*Target form*
Transc.ta	*night* ல நா க்வெஷன் கேக்குவோம்	கே(ட்)ப்பேன்
Trans.la	night la na queshun ke:k**kuv**o:m	ke:(T)**pp**e:n
MB	night la na: queshun ke:k-**k(u)-v**-o:m	ke:(T)-**pp**-e:n
UR	night la na: queshun ke:L-[45]**k(u)-v-**o:m	ke:L-**pp**-e:n
MG	night LOC I question [ask-**PRES**]-**NTARG.FUT**-1PL	ask- **FUT**-1S
Transl.	'I'll ask a question at night'	

[44] It's interesting to note that the compound verb in (**29**) was an unsolicited utterance from DS, not connected to any of the structured activities.

[45]The lateral *L* in *ke:L* does not surface due to some processes of sandhi: first it undergoes a form of manner assimilation with the following *k* to become a *T* and surfaces as the future stem form *ke:Tp-*. Then in colloquial speech the segment drops entirely to create the future stem form *ke:pp-*. When followed by the present morpheme, however, this segment instead assimilates to a *k*, which is what we see happening here.

In this example, we can see the verb *ke:L* 'ask' followed by a present tense morpheme *and* a future tense morpheme. The form of the present tense morpheme used here is *-k-*. This is the onset of *-kiR-,* which is the present tense allomorph we would expect with this verb (a class V verb). The form of the future tense morpheme, on the other hand, is not what we would associate with a class V verb: The future tense allomorph for class V verbs is *-p-*, but the one used by DS here is *-v-*.

Next is presented an utterance **(31)** using the class VI verb *kuDi '*drink'. Once again, the verb is followed by a present tense morpheme, and then by a future tense morpheme. Once again, the form used for the present tense is *-kk-*¸ which is similar to the *-kkiR-* we would expect with a verb in this verb class. And once again, the future tense allomorph is *-v-*, which is different from the form *-pp-* that we would normally expect with a class VI verb.

(31)	*Speaker utterance*	*Target form*
Transc.ta	*night* ல நா தண்ணி குடிக்குவோம்	குடிப்பேன்
Trans.la	night la na: thanni kuDi**kk(u)v**o:m	kuDi**pp**-e:n
MB	night la na: thanni kuDi-**kk(u)-v**-o:m	kuDi-**pp**-e:n
MG	night LOC I water [drink-**PRES]-NTARG.FUT**-1PL	drink- **FUT**-1S
Transl.	'I'll drink water at night'	

Finally, let's look at an utterance using the class VII verb *naDa* 'walk' **(32)**. This time, DS followed the verb with a past tense morpheme instead of a present tense one. The form used here is *-ndh-*, which is the allomorph we would expect with a class VII verb. Following the past tense morpheme is the future tense, and just as in the previous two examples, the form used here is *-v-*, rather than the *-pp-* form we would expect with a class VII verb.

(32)	*Speaker utterance*	*Target form*
Transc.ta	*night* ல நா நடுந்துவோம்	நடப்பேன்
Trans.la	night la na: naDu**ndhuv**o:m	naDa**pp**e:n
MB	night la na: naDu-**ndh(u)-v**-o:m	naDa-**pp**-e:n
MG	night LOC I [walk-**PST]-NTARG.FUT**-1PL	walk-**FUT**-1S
Transl.	'I'll walk at night'	

The three examples presented above have two things in common: 1) The first tense morpheme following the verb is always the exact form of, or clearly derived from, the allomorph we would expect for that verb's verb class; and 2) The second tense morpheme, the future tense, is always of the form *-v-*, which is not the form we would expect for any of those verbs.

With respect to point 1), I believe that the first tense morpheme in these examples has been reanalysed by DS and lexicalized into the verb stem. I will elaborate in chapter 5, but in brief, I believe this lexicalization is motivated by a general process of regularization that tries to "resolve" irregular patterns of allomorphy in functional morphemes by lexicalizing their most "unpredictable" (read: not transparently phonologically-determined) forms, and then simplifying the allomorphy to just a single allomorph. This falls under the broader phenomenon of overmarking present in several heritage languages cross-linguistically (Putnam et al., 2021), which I discuss further in chapter 5.

Regarding point 2), it is interesting to note that *-v-* is the future tense allomorph for every verb *except* the ones from classes V, VI, and VII. Additionally, commonly used everyday words like *po, va*, and *sa:ppiDu*, ("go," "come", and "eat" respectively) all belong to verb classes that use the *-v-* form; it is likely the future allomorph DS hears the most at home[46],

[46] Further research looking at precisely which verbs are most commonly used and heard by heritage speakers in the home could provide evidence to support (or disprove) this hypothesis.

and has become the most salient to her. Indeed, as discussed in earlier this chapter, we see DS accurately producing *-v-* to mark the future with the remaining verbs from class I to IV. I propose that because of how frequently it marks the future, *-v-* has become an unmarked future tense form, at least for DS. As such, she seems to have regularized future tense allomorphy to just this single allomorph, opting to largely dispense with the *-p-* and *-pp-* forms normally associated with class V and classes VI and VII respectively.[47]

I conclude the analysis of DS's future tense forms with one final observation: whenever DS used the future tense morpheme *-v-* , it was always followed by the 1st person plural PNG morpheme *-o:m* to create the combined suffix *-vo:m*, even in contexts where it was clear she was speaking only about either herself (see **[27], [30], [31], [32]**) or a third person neuter subject (**33**).[48] I don't have a solid explanation for why this is; my guess is that she may be used to her family using the 1PL when talking about things they are going to do as a family, and so this may be most familiar to her. It's also common for parents to adopt a first-person inclusive subject when speaking directly to their child about something they are encouraging the child to do (think about how an English-speaking parent may say to their child, "**let's** finish eating"), so it's possible DS has adopted this kind of phrasing. It is also possible that DS simply interprets the entire form *-vo:m* as a future tense marker, or perhaps as a fused tense and PNG marker, and is omitting an explicit PNG morpheme entirely. I discuss this latter hypothesis more in 4.2. Ultimately, more research is needed to further explain this phenomenon and whether it is more widespread.

[47] (**29**) presents a counter-example to this hypothesis; I will discuss the implications of this in chapter 5.
[48] In L1 speech, the future tense morpheme and 3rd person PNG fuse to create the form *-um* seen in **29**, so in fact the form produced here by DS diverges from the target-form we would expect. This is discussed in more detail in 4.2.

(33)	*Speaker utterance*	*Target form*
Transc.ta	*dog-ஓட food-உ, bone* <u>*சாப்புவோம்*</u>	*சாப்பிடும்*
Trans.la	dog-o:Da food-u bone <u>sa:ppu**vo:m**</u>	sa:ppiD**um**
MB	dog-o:Da food-u bone <u>sa:ppu-**v-o:m**</u>	sa:ppiD-**um**
MG	dog-SOC food-EUPH bone <u>eat-**FUT-1PL**</u>	eat-**FUT.3SN**
Transl.	‘a dog’s food; (it) eats/will eat bone(s)’	
	(intended)	

4.1.3 Present

Table 14: Strategies DS used to mark the present tenseTable 14 shows what strategy DS used to form the present tense for each verb. The first column in the table lists the class of each verb in the second column, and in brackets a reminder of what the expected present tense allomorph for this class is. As with the future tense, here also DS used both the strategies of a target-like present tense allomorph and of tense-stacking. Some tokens are categorized under both strategies; I will discuss these examples in more detail later on in this section. Cells are highlighted in green to indicate when the form used was target-like.

Verb Class	Verb	target-like present allomorph	tense-stacking
I *(-kir-)*	azhu *(cry)*	1	
III *(-kir-)*	pa:Du *(sing)*	12	
	(English verb) + paNNu *(do)*	3	
	pe:su *(speak)*	1	
	po: *(go)*	12	
	sol *(say)*	5	
	viLaiya:Du *(play)*	12	
IV *(-kir-)*	sa:ppiDu *(eat)*	8	
V *(-kir-)*	ke:Lu *(ask)*	4	
VI *(-kkir-)*	kuDi *(drink)*	13	
	pa:ru *(look / watch)*	1	
VII *(-kkir-)*	iru *(be)*	4 (3N)	
	naDa *(walk)*		1
Total		**58 (±18)**	**1 (±18)**

Table 14: Strategies DS used to mark the present tense

Strategy 1: target-like tense allomorph

In this section I will discuss examples that I have only categorized under the first strategy, "target-like present allomorph".

Before presenting the examples, it is important to make clear the distinguishing feature between the two present tense allomorphs, *-kiR-* and *-kkiR-*. Compared to the future tense allomorphs, these two forms seem nearly identical. Recall from 2.2.5, however, that in colloquial speech, it is common to reduce *-kiR-* to just *-R-* (Annamalai & Asher, 2002). This

reduction is not permissible with the *-kkiR-* allomorph. Colloquially, then, the two present tense allomorphs look more like *-R-* and *-kkiR-*. The distinguishing factor, thus, is the presence, or absence, of the *-kk-* segment.

Nearly all of the verb tokens with which DS only used the first strategy come from verb classes I to IV. These all carry the *-kiR-* allomorph, and as would be expected in colloquial speech, in all instances DS realized this morpheme as the reduced *-R-*. An example is seen in **(34)**, with the class III verb *viLaiya:Du*.

(34)	*Speaker utterance*
Transc.ta	பாட்டி விளையாடுறா
Trans.la	pa:TTi vaLa:D**R**a:
MB	pa:TTi vaLa:D-**R**-a:
UR	pa:TTi viLaiya:Du-**kiR**-a:L
MG	grandmother play-**PRS**-3F
Transl.	'grandmother plays'

Similar to the future tense table, I have marked in brackets the tokens where the verb agreed with the third person neuter. With the present tense, this only occurred with the verb *iru*. Unlike the future tense, the 3rd person neuter generally does not fuse with the present tense morpheme. Nevertheless, I have chosen to distinguish these tokens in particular because the combination of present tense *iru* with the 3rd person neuter contracts irregularly in colloquial speech; notably, the PNG morpheme drops off, and *-kkiR-* reduces to *-kk-*, as shown in (35)**.**

(35)	*Speaker utterance*
Transc.ta	எல்லாம் *body* ல இருக்கு
Trans.la	yella:m body la iru**kku**
MB	yella:m body la iru-**kku**
UR	yella:m body la iru-**kkiR**-adhu
MG	all body LOC be-**PRS**-3N
Transl.	'everything is in (its) body'

Note that this is not a feature of all class VII verbs. Dropping the PNG agreement and reducing *-kkiR-* to *-kk-* is not permissible for any other verb; this is just a feature of the verb *iru*. The form seen here, *irukku*, is also the existential 'there is', and is ubiquitous and unavoidable in everyday speech. It is highly likely that this form is a memorized "chunk" for DS.

Strategy 2: Tense-stacking – maybe?

If we return to Table 14, we see that an overwhelming majority of the time, DS used target-like present tense allomorphs in her constructions. Indeed, only one out of the 77 present tense utterances was non-target-like. However, for roughly a quarter of these target-like utterances, though superficially it appeared that the strategy DS used was simply producing a target-like present tense allomorph, I propose that she was actually using the tense-stacking strategy that we have already seen at play with the future tense.

First we look at a present tense example of a class V verb, *ke:L* 'ask'. This verb carries the *-kiR-* allomorph, and based on the discussion in the previous section, we would expect that in colloquial speech this would get reduced to *-R-*. However, *ke:L* behaves a bit differently: morphonological processes of assimilation mean that though the underlying present-tense stem of the verb is *ke:TkiR-*, it is more commonly colloquially realized as

ke:kkiR-[49]*.* Thus, though *ke:L* takes the *-kiR-* allomorph underlyingly, in this particular instance the *-kiR-* cannot be reduced to *-R-* because phonetically, it's actually realized as *-kkiR-,* with the first *k* being underlying part of the verb stem. Accordingly, we see DS faithfully produce the full *-kiR-* form in (36)**.**

(36)	*Speaker utterance*
Transc.ta	[50]அப்ப பயென் கேக்கறா
Trans.la	appa payen ke:k**kaR**a:
MB	appa payen ke:k-**kaR**-a:
UR	appo payen ke:L-**kiR**-a:n
MG	then boy ask-**PRS**-3M
Transl.	'then the boy asks' (intended)

It is interesting to compare this utterance with its future counterpart in (**37)**. Here, DS preceded the future morpheme *-v-* with *-kk-*, similar to what we see in the present form. Next we look at a class VI verb, *kuDi* 'drink'. We would expect the present allomorph *-kkiR-* here, which is indeed what DS produces in (37)**.**

(37)	*Speaker utterance*
Transc.ta	இப்பொவே தண்ணி குடிக்கிறேன்
Trans.la	ippove: thaNNi kuDi**kkR**e:n
MB	ippove: thaNNi kuDi-**kkR**-e:n
UR	ippove: thaNNi kuDi-**kkiR**-e:n
MG	now.EMPH water drink-**PRS**-1S
Transl.	'I'm drinking water now itself'

[49] The process is as follows: underlying *L* transforms to the stop *T* when preceding the stop *k* in *-kiR-*. Then in colloquial speech the segment assimilates entirely to become a *k*. This then creates a surface form *ke:kkiR-* that deceptively looks, at first glance, like a root *ke:* plus a present allomorph *-kkiR-*.
[50] This utterance is technically non-target like because of a mismatch in the PNG morpheme, but I leave discussion of this to 4.2**.**

Once again, if we compare this to its future counterpart in **(31)**, we see that in that utterance too, DS included the *-kk-* segment before the future morpheme.

Recall that the difference, in colloquial speech, between the two present allomorphs is the presence or lack thereof of the *kk* segment. Recall also that most verb classes take the *–(ki)R-* allomorph; i.e., the allomorph that (generally) doesn't surface with *kk*. I hypothesize that since the *-R-* allomorph appears so frequently, and indeed the segment *R* is part of the *-kkiR-* allomorph as well, that it may be the least marked present allomorph for DS. Consequently, I think she may have generalized it to use as a present morpheme for all verb classes and is lexicalizing the *kk* of the *-kkiR-* allomorph into the stems of the verbs that would normally carry this form. As such, I hypothesize that though the utterances in (36) and (37) represent target-like utterances, the strategy DS used with the class V and VI verbs was, once again, tense-stacking. The phonological parallels in the future tense tense-stacking examples support this idea. Here are the present tense utterances again, but with a different underlying parse to illustrate my hypothesis:

(38)	*Speaker utterance*
Transc.ta	அப்ப பயென் கேக்கறா
Trans.la	appa payen ke:k**kaR**a:
MB	appa payen ke:**kk(a)-R**-a:
MG	then boy ask-**PRS**-3M
Transl.	'then the boy asks' (intended)

(39)	*Speaker utterance*
Transc.ta	*இப்பொவே தண்ணி குடிக்கிறேன்*
Trans.la	ippove: thaNNi kuDi**kkR**e:n
MB	ippove: thaNNi kuDi**kk-R**-e:n
MG	now.EMPH water drink-**PRS**-1S
Transl.	'I'm drinking water now itself'

The last example I will be presenting is a clear-cut case of tense-stacking. We revisit class VII verb *naDa* **(40)**.

(40)	*Speaker utterance*	*Target form*
Transc.ta	*அப்போ morning ல நடுந்துவோம் – or நடுந்தறேன்*	*நடப்பேன் – or நடக்கிறேன்*
Trans.la	apa morning la naDu**ndhuv**o:m – or naDu**ndhaR**e:n	naDa**pp**e:n– or naDa**kkiR**e:n
MB	apa morning la na: naDu-**ndh(u)-v**-o:m or naDu-**ndh(a)-R**-e:n	naDa-**pp**-e:n– or naDa-**kkiR**-e:n
MG	then morning LOC I [walk-**PST]-NTARG.FUT**-1PL or [walk-**PST]-NTARG.PRES**-1S	walk- **FUT**-1S or walk- **PRES**-1S
Transl.	'then I'll walk – I walk/I am walking in the morning' (self-correction)	

This utterance contains an instance of future-tense tense-stacking directly followed by present-tense tense-stacking. In both, the initial morpheme is the past allomorph *-ndh-*, which is an appropriate past allomorph for a class VII verb. The same past allomorph was lexicalized with this verb when forming a future tense utterance in **(32)**. As discussed in 4.1.2,

the future allomorph used in **(32)** was what I proposed was the least marked future form; and likewise, the present allomorph used in **(40)** is what I am arguing is the least marked present form, *-R-*.

In sum, though nearly all present-tense utterances were target-like, I argue that the verbs from classes V-VII (excepting *iru*, which was already discussed) were built using the tense-stacking strategy. This matches the distribution seen with the future tense verbs, where all verbs from classes V to VII were built using tense-stacking. I argue that the allomorphs associated with classes I to IV are the least marked (at least for DS) and have been generalized (and replaced) the allomorphs normally associated with classes V to VII, for both the future and present tenses.

4.1.4 Past

Finally, we turn to the past tense morpheme. I want to start by acknowledging the limitations here; I elicited significantly fewer past tense tokens compared to the other two tenses. There are only fourteen utterances to analyse compared to the 77 present tense tokens and 35 future tense tokens. In addition, an entire verb class, class VI, is not represented at all. Nevertheless, I believe there were clear patterns worthy of discussion that emerged from this limited data.

The strategies that DS used to create past tense forms are shown in Table 15. As with the future and present tables, the expected allomorphs for each verb class are shown for reference in brackets in the first column. As with the present and future tenses, DS continued to use the strategy of producing a target-like past allomorph for some utterances. Unlike the future and the present tenses however, she never used the tense-stacking strategy. Instead, most of the time she used the strategy of creating compound verb constructions (CVCs), primarily using the completive (sometimes also called perfective) aspectual auxiliary *viDu*. All cells are highlighted in green because although some tokens might have been pragmatically a bit odd if they were coming from an L1 speaker, they were all still morphosyntactically well-

formed. One instance is included under both columns 5 and 6 because the form was a CVC that incorporated *viDu* in addition to another auxiliary. Some instances are under both columns 4 and 5 because, similar to what was discussed in the previous section, some utterances could arguably be categorized under both strategies.

Verb Class	Verb	target-like past allomorph	tense-stacking	CVC *-TT- (viDu)*	CVC other
I (*-dh-*)	azhu *(cry)*	1			
II (*-ndh-*)	va: *(come)*				1 (*a:ka,* III 3N)
	vizhu *(fall)*			4	
III (*-in-* or *-i-*)	(English verb) + paNNu *(do)*			1	
	sol *(say)*			1	
IV (*-TT-*)	sa:ppiDu *(eat)*	2			
V (*-T-*)	ke:Lu *(ask)*			2	
VII (*-ndh-*)	iru *(be)*	1			
	maRa *(forget)*			1 (*po*, III 3N)	
Total		**4**	**±8**	**8 ± 1 +-8**	**1 ± 1**

Table 15: Strategies DS used to mark the past tense

Strategy 1: target-like tense allomorph

With a little over a quarter of the tokens, DS simply used a target-like past tense allomorph to mark the past, as shown in (41)**.** This example shows DS producing a class IV verb, using the reduplicative past tense allomorph that doubles stem final */T/* to create the geminate *TT*, as we would expect from an L1 speaker.

(41)	*Speaker utterance*
Transc.ta	அப்ப *morning* ல சாப்ட்டேன்
Trans.la	appa morning la na: sa:pT**T**e:n
MB	appa morning la na: sa:pT-**T**-e:n
UR	appa morning la na: sa:ppiT-**T**-e:n
MG	then morning LOC I eat-**PST**-1S
Transl.	'then, I ate in the morning'

The other two allomorphs DS used are the class I *-dh-* as shown in (42) and the class VII *-ndh-* as shown in (43)**.** In (42), DS hesitated a bit over how to produce the verb stem. It was ultimately target-like, but I think the hesitation threw her off and is the reason she produced a third person male PNG agreement form *-a:n* at the end rather than the 1st person singular *-e:n*, seeing as in no other instance did she struggle with 1st person singular marking.

(42)	*Speaker utterance*	*Target form*
Transc.ta	*morning* ல அழ்—ழு--தான்	அழுதேன்
Trans.la	morning la na: azh—zhu—**dh**a:n	azhu**dh**e:n
MB	morning la azh—zhu-**dh**-a:n	azhu-**dh**-e:n
MG	morning LOC cry-**PST**-3M	cry-**PST**-1S
Transl.	'I cried in the morning' (intended)	

(43)	*Speaker utterance*
Transc.ta	நன்னா இருந்தது
Trans.la	nanna: iru**ndh**udhu
MB	nanna: iru-**ndh**-udhu
UR	nanna: iru-**ndh**-adhu
MG	good.ADV be-**PST**-3NS
Transl.	'it was good'

DS used target-like past allomorphs in four separate tokens that consisted of these three verbs, from classes I, IV, and VII.

Strategy 2: compound verb constructions

In ten out of the fourteen past tense tokens, DS used the strategy of creating compound verb constructions (CVCs). As discussed in 2.2.4, CVCs are verb phrases that incorporate aspectual and/or modal auxiliaries to the verb phrase. With the tokens we are about to discuss, the auxiliaries DS used were all aspectual. She produced one form that incorporated the auxiliary *a:ka* 'become', one form that incorporated both *po* 'go' and *viDu* 'leave', and eight forms that incorporated just *viDu*. I will discuss these in order.

The verb *a:ka* is a lexical verb meaning 'become'; but in modern use it has more or less become a phonologically reduced, bound morpheme. When used in a CVC, it adds a completive meaning to certain verbs. The utterance DS produced is shown in (44).

(44)	*Speaker utterance*
Transc.ta	*snake* வந்தாச்சு
Trans.la	snake vandh**a:chchu**
MB	snake va-ndh-**a:-chchu**[51]
MG	snake come-VBPL-**become-PST.3SN**
Transl.	'the snake has come/arrived'

The use of *a:ka* in this utterance adds an element of finality, and even possibly an expression of shock or surprise. Since its use in CVCs is as a completive, it makes sense that it is associated with the past tense.

In the next example (45)**,** DS creates a CVC using the auxiliary *po* 'go' followed by the auxiliary *viDu* 'leave'. (Note that *pey-* is a common pronunciation variant of *po:i.*) *Po* is used commonly as a lexical verb (if you look at the future and present verb tables again, you'll see that DS used it in this way several times). As an auxiliary, *po* is used with intransitive verbs to indicate a sudden change of state (Krishnamurti, 2003). *ViDu* is also used as a lexical verb; as an auxiliary it functions as a perfective or completive. Thus, used together, the *po* and *viDu* auxiliaries indicate a sudden, completed change of state. Note from 2.2.2 that the combination of *viDu* plus the past tense plus the 3rd singular neuter is realized as the fused form *–(u)rththu* in some *vattara vazhakkukaL*, in stark contrast to its underlying analysable -*viTTadhu*. This is the form we see DS produce in (45)**.**

[51]As was touched on in 2.2.2**,** *a:chchu* is a colloquial form, compared to the high register *a:giyadhu* or *a:yiTRRu*; the former maintains the *ka* of the underlying stem, while the form used by DS uses the reduced stem *a:-* .

(45)	*Speaker utterance*
Transc.ta	*full-*ஆ மறந்து பெயிர்த்து
Trans.la	fulla: maRandhu **peyurththu**
MB	full-a: maRa-ndhu **pe-y-[urththu]**
UR	full-a:ka maRa-ndhu **po:-i-[viT-T-adhu]**
MG	full-ADV forget-VBPL **go-VBPL-[leave-VBPL-3SN]**
Transl.	'I've completely forgotten'

In this utterance, the combination of the two auxiliaries works to emphasize the fact that DS has well and truly forgotten whatever it is she's forgotten.

In both (44) and (45), the auxiliaries are conjugated with the past and the 3rd person singular neuter. While there is a 3rd person singular neuter morpheme that attaches independently to any past allomorph, in colloquial speech it is common to use forms that fuse the past tense and 3rd singular neuter morphemes, which is what we see DS using here. These merged forms can vary slightly with different verbs, which is why we see a phonological difference in the forms here.

The forms used here are highly irregular in that the markers of the past tense are not consistent and so difficult to predict. However, the auxiliaries *a:ka*, *po,* and *viDu* are extremely common, and though CVCs in general are morphologically "complex", they are in fact very common in daily use. It's impossible to go a day with hearing and using them. As such, once again I am curious about whether the auxiliaries DS is using in these two CVCs are memorized chunks. This would require more research to confirm; it could be interesting to ask the speakers what their intuitions are about which forms carry which meanings.

By far the auxiliary verb that DS used most in her past tense utterances was the verb *viDu*. As a lexical verb it means 'leave', but as discussed above, in CVCs it marks the perfective or completive aspect. When appearing as this aspect marker, it frequently

undergoes phonological reduction and loses its initial syllable (*vi*). Conjugated with the past tense, this results in a surface form of just the past allomorph *TT*[52], as seen in (46)**.**

(46)	*Speaker utterance*
Transc.ta	*playdate* பண்ணிட்டா
Trans.la	playdate paNNi**TT**a:
MB	playdate paNN-i-**ø-TT**-a:
UR	playdate paNN-i-**viT-T**-a:
MG	then do-VBPL-**leave-PST**-3F
Transl.	'they finished the playdate'

The phonological realization of the combination of *viDu* and a past morpheme, though in a reduced form, remains quite salient. Not only does -*TT*- have a longer duration because it is a geminate, but unlike the other past tense allomorphs (except that of class III), it tends to appear in its own syllable separate from the verb stem as opposed to within a consonant cluster: Notice that in (47), -*TT*- is surrounded by vowels.

(47)	*Speaker utterance*
Transc.ta	அப்ப சொல்லிட்டா
Trans.la	appa solli**TT**a:
MB	appa soll-i-**ø-TT**-a:
UR	appa sol-i-**viT-T**-a:
MG	then say-VBPL-**leave-PST**-3F
Transl.	'then she said' (with an air of finality)

[52] *Vidu* is itself a class IV verb, so the past "allomorph" is actually just reduplicating whatever the final consonant of the verb stem is. The verb stem is *viD-*. Recall that *D* and *T* are both allophones of the phoneme *T*. Reduplication of an obstruent is the same as creating a geminate, and as part of a geminate, voiced allophones become voiceless, so the past stem of *viDu* surfaces as *viTT-*.

Because this form -*TT*- is so salient, and because DS would frequently hear it used in the context of past events (since it is a completive), it is possible that she reanalysed the form to itself *be* a past tense marker. Furthermore, she may be in the early stages of generalizing this morpheme to replace all the other past tense allomorphs, which are (as we know) not phonologically regular and therefore more difficult to predict. -*TT*- can be used with any verb, regardless of class, and so this may feel like a more consistent way of forming past tense utterances.[53]

In addition to its phonological salience, my mother pointed out that the combination of *viDu* and the past morpheme is frequently used by parents when talking to their children, because it is always used when asking questions of the form "have you finished/done xyz". Children would likely be used to hearing their parents asking them if they've got out of bed, brushed their teeth, washed their face, cleaned their room, eaten their breakfast - all of these questions would contain the *viDu* auxiliary conjugated in the past tense, as exemplified in (48)**.**

(48)	*Example utterance*
Transc.ta	எழுந்துட்டியா?
Trans.la	ezhundhu**TT**iya:
MB	ezhu-ndhu-**TT**-iy-a:
UR	ezhu-ndhu-**viTT**-a:y-a:
MG	wake.up/get.up-VBLP-**leave.PST**-2S-INTERR
Transl.	'Are you up/awake yet?'

It is important to note that all three auxiliaries used by DS -- *a:ka, po̱* and *viDu* -- must attach to a verbal participle; that is, the preceding lexical verb must be in verbal participle form. In all of the utterances where DS formed a CVC, she used a target-like verbal participle. Since

[53] Informally, I have noticed several of my heritage Tamil students also use this strategy, and so I believe it might be more widespread across the younger heritage speaking population.

the verbal participle is homophonous to the past tense, this is essentially the same as needing to know what past tense allomorph the lexical verb takes. Indeed, it is possible that when using the auxiliary *viDu*, this is also actually just tense-stacking and she is overgeneralizing the common *-TT-* form to be a past tense morpheme for all verbs. Similar to the present tense examples discussed in 4.1.3, the strategy of tense-stacking here would be less apparent because it happened to produce target-like forms.

When we look at the verbs with which DS formed a CVC, we see they belong to classes II ([(44)] and [(49)] shown below), III ([(46)] and [(47)]), V ([(50)]), and VII ([(45)]). Classes II and VII both take the same past tense allomorph (i.e., form of the verbal participle), *-ndh-*. The two class III verbs DS used both take the form *-i-*. The class V verb *ke:L*, once again proves interesting; because of the underlying sandhi that we have discussed before, the verbal participle almost seems to surface as *-TT-*, just like past tense *viDu* does.

(49)	*Speaker utterance*
Transc.ta	அப்ப *morning* ல நா விழுந்துட்டேன்
Trans.la	appa morning la na: vizhndhu**TT**e:n
MB	appa morning la na: vizh-ndhu-**ø-TT**-e:n
UR	appa morning la na: vizh-ndhu-**viT-T**-e:n
MG	then morning LOC I fall-VBPL-**leave-PST**-1S
Transl.	'then, in the morning I fell' (implication of 'I finished falling' or 'I went and fell')

(50)	*Speaker utterance*
Transc.ta	*அப்ப morning ல நா question கேட்டுட்டேன்*
Trans.la	appa morning la na: question ke:TTu**TT**e:n
MB	appa morning la na: question ke:T-Tu-**ø-TT**-e:n
UR	appa morning la na: question ke:L-Tu-**viT-T-**e:n
MG	then morning LOC I question ask-VBPL-**leave-PST-**1S
Transl.	'then I asked/ finished the task of asking the question in the morning'

In all CVC utterances, DS produced the verbal participle of the lexical verb accurately the way an L1 speaker would. In fact, with the exception of verb class VI (of which I was not able to get a past tense token), DS produced the past tense allomorphs associated with every verb class accurately, either through the creation of a simple present or through the formation of a verbal participle used in a CVC. This to me reinforces the idea that even if she sometimes uses different strategies to express tense, DS definitely has knowledge of tense allomorphy, especially knowledge and recognition of their phonological shapes.

4.1.5 Summary of tense morphology

Overwhelmingly, most of the time DS produced utterances with target-like tense morphology. Around two-thirds of the future tense tokens were target-like, every present tense token except one was target-like, and all the past tense tokens were morphosyntactically well-formed (if pragmatically unconventional at times). DS used three strategies to express tense: using a tense morpheme, "tense-stacking", or creating a CVC. DS had the most non-target-like forms when expressing the future tense; but even here, she was consistent in using the strategy of tense-stacking to create these forms. Tense-stacking is a strategy that depends on the speaker having knowledge of lexically-determined tense

allomorphy, lexicalizing marked allomorphs and incorporating them into verb stems, and generalizing an unmarked allomorph to be used in their stead. Whether the forms DS produced were what one would expect from an L1 speaker or not, they all clearly demonstrated DS's remarkable understanding and knowledge of tense allomorphy, which is arguably one of the more challenging aspects of Tamil verbal morphology to acquire.

DS used the strategy of tense-stacking exclusively with verbs from classes V to VII, and only to create future or present tense utterances. With the class V and VI verbs *ke:L*, *kuDi,* (and arguably *pa:ru*), DS lexicalized the present tense allomorph, and with the class VII verb *naDa* she lexicalized the past allomorph. The future (*-pp-*) and present (*-kkiR-*) allomorphs for classes V, VI, and VII are phonologically identical,[54] and contrast with the future (*-v-*) and present (*-[ki]R-*) allomorphs that classes I to IV share. I argued in the previous sections that the future and present allomorphs for classes I to IV are unmarked for DS compared to the ones for classes V to VII. At the same time, DS is very familiar with hearing all of these allomorphs because they are produced by the Tamil speakers around her. In order to reconcile the information she is hearing with her own process of regularization and generalization, I argue she is lexicalizing the marked allomorphs and incorporating them into the stems of the verbs they are associated with, and then generalizing the less marked allophones to "stack" on top as overt markers of tense.

Except for with the verb *naDa*, where she used the past tense, DS always lexicalized the present tense allomorph when tense-stacking. It makes sense that she never lexicalized the future tense as she probably hears it less often; in contrast, the present will get used all the time in conversation to describe current actions, and the verbal participle (i.e. past

[54] Technically, the allomorphs for VI and VII are geminates while the ones for V aren't underlyingly; but as has already been discussed a lot in this section, the sandhi effects of the verb *ke:L*, (which was the only class V verb in the data), ultimately end up making the tense allomorphs surface as geminates.

allomorph homophone) is used all the time to form CVCs. Also, future tense allomorphy is comparatively less phonologically irregular than that of the past and the present.

Considering how unpredictable past tense allomorphy is though, it does seem surprising that DS rarely lexicalized these forms. In fact, she was always target-like in her production of the different verbal participle allomorphs, which are homophonous to past tense morphemes. Interestingly, she comprehends[55] and produces past tense allomorphy the way an L1 speaker would, but she uses different strategies (i.e. tense-stacking) with producing the present and future tenses.[56] Either this is simply variation that she displays consistently, or instead an indication that her mental grammar is still changing. If the latter hypothesis is true, and her mental grammar has not fossilized, this would be in accordance with the literature that proposes that reanalysis of the heritage grammar continues past puberty and into adulthood (Montrul, 2016; Polinsky, 2008, 2011). It would be illuminative to see how her grammar changes over the years; will she move towards adopting tense-stacking with all tenses? Will she use more tense allomorphy similar to L1 speakers? Or will it be yet another mix of both, and other strategies like creating CVCs?

4.2 Speaker 1 (DS) PNG morphology

In general, DS used target-like PNG forms and so in this section I provide just a brief overview of each of the forms she used. Table 2 from 2.2.2 is partially duplicated here for reference, including only the persons that DS produced in her speech.

[55] The evidence for this is admittedly anecdotal: she clearly comprehended me when I spoke in Tamil during the interviews, and I am also drawing here on my interactions with her outside of the research context in social situations, where I know she comprehends all that is said to her in Tamil.

[56] This distinction is especially stark with the verb *ke:L*. With the future (and I argue also present) tenses, DS uses tense-stacking by lexicalizing the partial present tense allomorph *kk*. However, with the past, she produces a target-like verbal participle to form a CVC, thus showing a productive understanding of past tense allomorphy.

Person-number-gender	*Centamil* form	Common *koDuntamil* forms
1S	*-e:n*	*-ẽ:*
1PL.EXC and 1PL.INC	*-o:m*	*-õ:*
3F	*-a:L*	*-a:*
3M	*-a:n*	*-ã:*
3HON	*-a:r*	*-a:*
3SN	*-adhu*	
	(for class III verbs) *-iRRu* (incorporates past tense)	*-chchu* (incorporates past tense)
	-um (incorporates future tense)	Many compound verb forms exist that fuse the 3SN morpheme with preceding auxiliaries.
3PL	*-a :rgaL*	*-a:ngaL* *-a:nga* ***-a:*** *-a:L*

Table 16: Person-number-gender morphemes in Tamil (reduced for DS)

The forms DS used, on the whole, matched with the forms used by the L1 baseline speakers. She used the form *-a:* to refer to 3rd feminine, honorific, and plural subjects; though in the

formal register, these persons each correspond to unique PNG morphemes, in the dialect DS (and myself) speak, they are frequently all referred to with just this one form.[57] In one non-target-like instance, she also used this form when referring to a 3rd person singular neuter subject (51). This utterance was one out of 16 otherwise target-like 3rd person singular neuter PNG morphemes, so there isn't really enough of a pattern to help explain this one utterance.

(51)	*Speaker utterance*	*Target form*
Transc.ta	…and then at the end அது snake எல்லா	சாப்பிட்டது
Trans.la	சாப்பிட்டா	
	and then at the end adhu snake yella: sha:pT**a:**	sa:ppiTT(**adhu/icchu**)
MB	and then at the end adhu snake yella: sha:p-T-**a:**	sa:ppiT-T-(**adhu/icchu**)
MG	and then at the end it.that snake all eat-PST-**3F/PL/HON**	eat-PST-**3SN**
Transl.	'and then at the end it, the snake ate all of it'	

When referring to 1st person subjects, DS used the target-like PNG form *-e:n*. The exceptions are one instance (42) which was already discussed in 4.1.4, and instances where the verb was conjugated in the future tense and DS used the fused form *-vo:m* instead; this was discussed in 4.1.2.

As was touched on in 3.5.2, it was often unclear whether expected word-final nasal vowels were indeed nasalized, and in three out of twenty-three instances this was the case with the *-e:n* form. This was even more frequently the case with the 3rd person masculine PNG form, where the word-final nasal vowel in the form *-a:n* was only indisputably present in one out of seventeen utterances, and was otherwise unclear.

[57] This usage is actually a marker of our dialect.

DS used the widest variety of PNG forms in expressing the 3rd singular neuter form. As was discussed in 2.2.2, the 3rd singular neuter behaves the least regularly of all the PNG forms and has many allomorphs where it fuses with different tenses. DS used a variety of these target-like fused forms, as was seen in (44) and (45) in 4.1.4. In addition to these target-like forms, she also produced two other non-target-like variants, as seen in (52) and (53) below. The full verb form seen in (52) was produced in five separate utterances, and the full verb form in (53) was produced twice.

(52)	*Speaker utterance*	*Target form*
Transc.ta	dog-ஓட food-உ, bone சாப்புவோம்	சாப்பிடும்
Trans.la	dog-o:Da food-u bone sa:ppu**vo:m**	sa:ppiD**um**
MB	dog-o:Da food-u bone sa:ppu-**vo:m**	sa:ppiD-**um**
MG	dog-SOC food-EUPH bone eat-**FUT.3SN**	eat-**FUT.3SN**
Transl.	'a dog's food; (it) eats/will eat bone(s)'	

(53)	*Speaker utterance*	*Target form*
Transc.ta	snake-உ என்ன சாப்புவோன்	சாப்பிடும்
Trans.la	snake-u yenna sha:ppu**vo:n**	sa:ppiD**um**
MB	snake-u yenna sha:ppu-**vo:n**	sa:ppiD-**um**
MG	snake-EUPH what eat-**FUT.3SN**	eat-**FUT.3SN**
Transl.	'what does/will a snake eat'	

On the surface, the *-vo:m* and *-vo:n* in (52) and (53) appear to be 1st person plural PNG forms. The context, however, tells us that DS intended a 3rd singular neuter (recall that the neuter gender is generally used for non-human subjects). I present two analyses of this form. Both rely on the assumption that *-vo:m* and *-vo:n* are underlyingly *-vo:m*, where *-vo:n* is a

free variation word-final nasal vowel allophone of *-vo:m* (see 2.2.5). The first analysis is a more morphological explanation, the second is phonological.

The first analysis I offer is that these tokens are further evidence to support my hypothesis from 4.1.2 that the form *-vo:m* represents a fused future and PNG form for DS, where the PNG element is subject-neutral. We saw in 4.1.2 that she used this form when referring to what were clearly 1st person singular subjects, and here we see her use it with 3rd person neuter subjects. Data about what forms she uses when forming the future tense with other persons is required to equivocally accept or reject this hypothesis.

The second analysis I propose is that in these instances, *-vo:m* is an allophone of *-um* that is triggered by appearing in an intervocalic environment. Note that the stem DS produces here is *saapa-*, not *saapiD-* as we would expect with an L1 speaker. Since the *D* is missing, the combination of the stem plus the *-um* suffix would create the form *saapaum*. The double vowel combination is dispreferred by Tamil syllable structure (Annamalai, 2019). In such environments of vowel hiatus, an epenthetic *v* or *y* (depending on the quality of the following vowel)[58] is inserted between the two vowels. This now gives us the form *saapavum*. Now, recall from 2.2.5 that there are several colloquial pronunciations of words that indicate a *u~o* merger may be underway. If we accept that this merger is at play here, we can then expect the surface form *saapavom*. This doesn't fully explain the vowel lengthening in the actual produced form; possibly it just sounds more familiar to forms DS has heard (since *-vo:m* is actually the 1st plural form). This analysis would also need evidence from other future tense verbs with 3rd person neuter subjects to further support or dismiss it.

[58] Annamalai (2019) calls this phoneme an "onglide"; every vowel corresponds to one of these two onglides.

4.3 Speaker 2 (AB) tense morphology

This section is a description and analysis of AB's tense morphology. The following chapter is an analysis of her PNG morphology.

4.3.1 Overview of tense-marking strategies

AB used similar tense-marking strategies to DS, shown in Table 17. In addition to the strategies of using a target-like tense allomorph, tense-stacking, and compound verb constructions that both speakers used, AB also uniquely used the strategy of replacing the expected tense allomorph with a generalized "unmarked" allomorph for some utterances. Sometimes, AB used multiple strategies for the same utterance. As such, the sum total in Table 17 is higher than the total number of utterances reported in Table 9.

Overwhelmingly, AB used target-like tense allomorphs. In the instances where she didn't, she preferred compound verb constructions, then tense-stacking, then a generalized unmarked tense allomorph, in that order. Of course, as with DS, the strategy AB used was heavily dependent on the tense she was expressing.[59] Each tense will be discussed in turn.

Verb class	Target-like tense allomorph	Generalized unmarked tense allomorph	Tense-stacking	Compound Verb Constructions
I	1	6	0	2
II	3	2	7	3
III	36	1	1	9
IV	17	0	0	0
V	5	0	5	1
VI	5	0	12	6
VII	6	0	5	2
TOTAL	**73**	**9**	**30**	**23**

[59] I deliberately do not include percentages in this table, to discourage blind comparison of the numbers outside of their phrasal contexts.

Table 17: Tense-marking strategies AB used with each verb class

4.3.2 Future

As seen in Table 18 below, AB strongly preferred using compound verb constructions to express the future tense, doing this with over half of the future tense utterances. She also employed tense-stacking with five of these compound verb constructions, and I will discuss these examples in detail later in this section. Finally, in three cases, AB used the strategy of a target-like tense allomorph. I will discuss AB's use of each of these strategies in the next sections. As with the tables in 4.1, green highlighting signifies target-like constructions for all of the tables in 4.3.

Verb Class	Verb	target-like future allomorph	tense-stacking	compound-verb construction *(po:)*
I *(-v-)*	azhu *(cry)*			2
II *(-v-)*	vizhu *(fall)*		1	
III *(-v-)*	pa:Du *(sing)*			2
	po: *(go)*			2
IV *(-v-)*	sa:ppiDu *(eat)*	1 (3N)		
V *(-p-)*	ke:Lu *(ask)*			1
VI *(-pp-)*	kuDi *(drink)*		2	
VII *(-pp-)*	iru *(be)*	2 (3N)		
	naDa *(walk)*		2	
Total	**15**	**3**	**5**	**12**

Table 18: Strategies AB used to mark the future tense

Strategy 1: target-like tense allomorph

AB used a single target-like tense allomorph to express the future with two verbs, shown in (54) and (55). In both instances the form used was the 3rd person neuter morpheme, which,

as discussed in 2.2.2 and 4.1.2, has only one form and doesn't change based on verb class. Indeed, the verb used in (54) *saappiDu*, is a class IV verb while the verb in (55) *iru* is a class VII verb. Note that these two verbs are extremely commonly used in daily speech.

(54)	*Speaker utterance*
Transc.ta	*snake-உ என்ன சாப்டும்*
Trans.la	snake-u enna sa:pD**um**
MB	snake-u enna sa:pD-**um**
UR	snake-u enna sa:ppiD-**um**
MG	snake-EUPH what eat-**FUT.3SN**
Transl.	"what does a/the snake eat?"

(55)	*Speaker utterance*
Transc.ta	*...and fish cat* க்கு இருக்கும்
Trans.la	and fish cat kku irukk**um**
MB	and fish cat kku iru-kk-**um**
MG	and fish cat-DAT be-STR-**FUT.3SN**
Transl.	"…and the fish is / will be / would be / must be for the cat"

Strategies 2 and 3: tense-stacking and compound verb construction

In an overwhelming majority of utterances, AB expressed the future tense by forming compound verb constructions using the auxiliary *po:* 'go'. *Po:* is used in different ways in different compound verb constructions; the way AB is using it here is to express the meaning

"I'm going to X", where X represents any activity.[60] When used in this way, *po:* needs to attach to a verb in infinitival form.[61]

Recall from 2.2.4, the infinitive is formed by adding the infinitival suffix *-a* to the verb stem, as AB does in (56) with the verb *pa:Du* 'sing'. With verbs with middle or strong stems (classes V to VII), this suffix is preceded by a *-k-* or *-kk-* respectively, so on the surface, with these verbs the infinitival suffix appears to take the form *-kka.* AB produces such a form using the class V verb *ke:L* 'ask', as seen in (57)**.**

(56)	*Speaker utterance*
Transc.ta	பாட்டு பாடப்போறன்
Trans.la	pa:TTu pa:D**a**ppo:re:
MB	pa:TTu pa:D-**a**-po:-r-e:n
UR	pa:TTu pa:D-**a-**po:-kiR-e:n
MG	song sing-**INF**-go-PRS-1S
Transl.	"I'm going to sing a song"

(57)	*Speaker utterance*
Transc.ta	நா கேக்கப்போறன்
Trans.la	na: ke:k**ka**ppo:re
MB	na: ke:k-**k-a**-ppo:-r-e:n
UR	na: ke:L-**k-a**-po:-kiR-e:n
MG	I ask-**MID-INF**-go-PRS-1S
Transl.	"I'm going to ask"

[60] Note for non-Tamil speakers: this native Tamil construction basically has the same meaning and structure as the English future construction "I'm going to X."

[61] Recall that this is different from the auxiliaries we have seen so far, in DS's speech, *viDu* and *iru*. Both of these auxiliaries were aspectual, and therefore needed to attach to a verb in verbal participle form.

The verb *po:* ‘go’ is somewhat irregular, and can appear in these sorts of compound verb construction without an infinitive when it is itself the main verb, as in (58)**.** AB, however, retains the infinitive. This may be to differentiate lexical *po:* (bolded in this example) from auxiliary *po*, which she produces as normal*:*, or perhaps it is for emphasis.

(58)	*Speaker utterance*
Transc.ta	நா *school*-க்கு போகப்போறேன்
Trans.la	na: school-kku **po:ga**ppo:re
MB	na: school-kku **po:g-a**-po:-r-e
UR	na: school-kku **po:g-a**-po:-r-e
MG	I school-DAT **go-INF**-go-PRS-1S
Transl.	“I’m going to go to school”

AB does not always use the infinitival form of the verb in these compound verb constructions, and instead frequently uses the verbal participle form instead, as seen in (59)**.** The target-like verb form we would generally expect instead is *vizhappo:re:n*.

(59)	*Speaker utterance*	*Target form*
Transc.ta	நா வீந்துப்போறேன்	நா விழப்போறேன்
Trans.la	na: vi:**ndhu**ppo:re	na: vizh**a**ppo:re:n
MB	na: [vizhu-**ndhu]**-po:-r-e:n	na: vizh(u)-**a**-po:-(ki)R-e:n
MG	I [fall-**VBPL]**-go-PRS-1S	I fall-**INF**-go-PRS-1S
Transl.	“I’m going to fall” (intended)	

Interestingly, twice with the verb *kuDi* ‘drink’ she started to form the infinitival form we would expect and then switched to the verbal participle form instead, as seen in (60)**.** In both instances, the stem she first uses is *kuDu*, and the stem she corrects to is *kuDi*. *kuDu* is in fact

a separate verb meaning 'give',[62] which AB used once earlier in the session (see [(83)] in 4.3.4); it would need more evidence to support or disprove, but my instinct is that AB was trying to differentiate between *kuDu* and *kuDi* (because in colloquial speech the two final short vowels are very close in quality), and in order to do this, changed the allomorph of the following morpheme from *kk* to *chch*; possibly she thought that the two verbs were distinguished by carrying different tense allomorphs as a result of belonging to different verb classes,[63] or she is incorporating *kk* into the *kuDu* stem and *chch* into the *kuDi* stem.

(60)	*Speaker utterance*	*Target form*
Transc.ta	தண்ணி குடுக்கப்போறேன்- குடிச்சிப்போறேன்	குடிக்கப்போறேன்
Trans.la	thanni kuD**ukka**ppo:re— kuD**ichchi**ppo:re:	kuDi**kka**ppo:re:n
MB	thanni kuD**u-kk-a**-po:-r-e—[kuD**i-chchi]**-po:-r-e:n	kuDi-**kk-a**-ppo:-r-e:n
UR		kuDi-**kk-a**-po:-kiR-e:n
MG	water drink-**STR-INF**-go-PRS-1S— [drink-**VBPL]**-go-PRS-1S	drink-**STR-INF**-go-PRS-1S
Transl.	"I'm going to drink water" (intended)	

AB uses the verbal participle instead of the infinitive again with the verb *naDa* 'walk' in (61) and (62)**;**[64] with the infinitive we would expect the form *naDakkappo:re:n*.

[62] Note the *centamil* form of this verb is *koDu* rather than *kuDu*. Since in AB's dialect (and mine), it appears to be *kuDu*, I hereon out only transcribe this verb as *kuDu* underlyingly.

[63]In fact *kuDi* and *kuDu* both belong to verb class VI. However, the colloquial variants of their tense allomorphs do vary, in that the *-thth-* past allomorph is often realized as *-chch-* with the verb *kuDi*, but never with the verb *kuDu*.

[64]In these examples, the switch from infinitive to verbal participle actually changes the meaning of the sentence. Though I believe she intended to say "I will walk", the sentences mean something more along the lines of "the way I will go, is by walking". Generally, this phrasing is interpreted as the speaker drawing attention not to the time of the event (future), but instead the manner by which they are going to some pre-discussed destination (by walking).

(61)	*Speaker utterance*	*Target form*
Transc.ta	நந்து போகப்போறேன்	நடக்கப்போறேன்
Trans.la	na**ndhu**ppo:gapo:re:	naDa**kka**ppo:re:n
MB	[naDu-**ndhu]**-po:g-a-po:-r-e:n	naDa-**kk-a**-po:-r-e:n
UR		naDa-**kk-a**-po:-kiR-e:n
MG	[walk-**VBPL]** go-INF-go-PRS-1S	walk-**STR-INF**-go-PRS-1S
Transl.	"I'm going to walk" (intended)	

(62)	*Speaker utterance*	*Target form*
Transc.ta	நா நந்துப்போறேன்	நடக்கப்போறேன்
Trans.la	na: na**ndhu**ppo:re	naDa**kka**ppo:re:n
MB	na: [naDu-**ndhu]**-po:-r-e:n	naDa-**kk-a**-po:-r-e:n
UR		naDa-**kk-a**-po:-kiR-e:n
MG	I [walk-**VBPL]**-go-PRS-1S	walk-**STR-INF**-go-PRS-1S
Transl.	"I'm going to walk" (intended)	

In the examples (59), (60), (61), and (62), the compound verb constructions are clear, but there doesn't seem to be any tense-stacking. After all, what seems to be happening is that the infinitive is being dropped in favour of the verbal participle. I have chosen to still group this under the strategy of tense-stacking however, because the verbal participle is homophonous to the past tense. As such, I believe that what is happening is at its core, is the same as what happens with the tense-stacking seen in DS's speech: the "irregular" or "unpredictable" tense / verbal participle allomorph is being reanalysed and lexicalized into the verb stem. It no longer carries the function of a verbal participle. Unlike tense, however, a verbal participle doesn't really carry a clear semantic meaning. I hypothesize that because of this, AB doesn't feel the need to explicitly mark the verbal participle the way she would

tense, explaining why she then doesn't stack verbal participle allomorphs. Nevertheless I am arguing that the reanalysis and lexicalization of a functional morpheme, which is the essence of tense-stacking, is still what is going on here. This will become more apparent when looking at the present and past constructions.

4.3.3 Present

AB largely used target-like present tense allomorphs to express the present, and in a few instances used tense-stacking, as enumerated in Table 19.

Verb Class	Verb	*target-like present allomorph*	tense-stacking
I *(-kiR-)*	azhu *(cry)*	1	
II *(-kiR-)*	vizhu *(fall)*	3	
III *(-kiR-)*	pa:Du *(sing)*	11	
	po: *(go)*	12	
	viLaiya:Du *(play)*	7	
IV *(-kiR-)*	sa:ppiDu *(eat)*	12	
V *(-kiR-)*	ke:Lu *(ask)*	2	
VI *(-kkiR-)*	kuDi *(drink)*		6
		3	
VII *(-kkir-)*	iru *(be)*	4	
	naDa *(walk)*		1
Total	**62**	**55**	**15**

Table 19: Strategies AB used to mark the present tense

Strategy 1: target-like tense allomorph

AB used target-like present tense allomorphs in most of her present tense utterances. With the exception of verbs *iru* and *ke:L* which belong to verb classes VII and VI respectively (the latter of which I discuss more in 4.3.4), AB used the strategy of target-like present tense allomorphy exclusively with verbs from classes I to IV. Recall that verbs from these classes take the present tense allomorph *-kiR-*, which is colloquially reduced to *-R-* or *-r-*. This is the form that AB used in most cases, as exemplified in (63)**.**

(63)	*Speaker utterance*
Transc.ta	*புண்ணு பாட்டு பாடுறா*
Trans.la	puNNu pa:TTu pa:Du**r**a:
MB	puNNu pa:TTu pa:Du-**r**-a:
UR	poNNu pa:TTu pa:Du-**kiR**-a:L
MG	girl song sing-**PRS**-3F
Transl.	“(the) girl is singing a song”

Just as DS did in (35) in 4.1.3, AB uses the colloquial reduced form *irukku* for the present tense form of *iru*, as seen in (64)**.**

(64)	*Speaker utterance*
Transc.ta	*எங்க Fluffy இருக்கு*
Trans.la	engu fluffy irukku
MB	enga fluffy iru-kku
UR	enga fluffy iru-[kkiR-adhu]
MG	where fluffy be-[PRS-3N]
Transl.	“where is Fluffy”

In one case (65), AB used *-zh-* to mark the present tense instead of *-R-*. I believe this is a case of unintentional metathesis, where the *zh* in the verb stem was switched with the *R* of the tense morpheme. This seems suggestive of underlying challenges with perception or production (or both) of the /zh/ phoneme. As was discussed in 2.4.1, /zh/ is one of the phonemes that is last to be acquired. Examples in 4.3.4 show that AB often produces a lenited allophone of /zh/. I briefly touch on this pattern of lenition in 4.5.

(65)	*Speaker utterance*
Transc.ta	இப்போ அரேழன்
Trans.la	ippo a**r**a**zh**e:n
MB	ippo a**r**a-**zh**-e:n
UR	ippo a**zh**u-**kiR**-e:n
MG	now cry-**PRS**-1S
Transl.	"I'm crying now"

Strategy 2: tense-stacking

AB used tense-stacking in about a quarter of her present tense utterances. She used this strategy with the verbs *vizhu*, *kuDi*, *naDa*, and I will argue *ke:L* as well. It is interesting to note that AB used the tense-stacking strategy when expressing the future tense with these verbs as well. Also, with the exception of *vizhu*, these verbs belong to classes V, VI and VII; which are the verb classes DS used the tense-stacking strategy with as well. This parallel will be discussed in more depth in chapter 5.

AB, like DS, appears to view *-R-*, the colloquially reduced form of the present tense allomorph *-kiR-*, as the most unmarked form of the present tense, and has generalized this for verbs across verb classes. This is the expected allomorph for verbs from classes I to IV, and we can see from Table 19 that this is indeed the form AB used with verbs from these classes. With the verb from class II *vizhu*, though, she preceded the target-like present tense

allomorph with tense-stacking, as seen in (66)**.** The present tense *-R-* is preceded by the target-like past tense allomorph *-ndh-*.

(66)	*Speaker utterance*	*Target form*
Transc.ta	*நா வீந்துறே*	*விழுறேன்*
Trans.la	na: vi:**ndh**u**r**e:	vizhu**r**e:n
MB	na: [vi:-**ndh]**-u-**r**-e:	vizhu-**r**-e:n
UR		vizhu-**kiR**-e:n
MG	I [fall-**PST]**-EUPH-**PRS**-1S	fall-**PRS**-1S
Transl.	"I'm falling" (intended)	

The expected form of this verb would be *vizhuRe:n*. I discuss the process more in 4.5, but to summarize, I believe that with the verb *vizhu*, AB has lenited the /zh/ phoneme to the point of deletion. She likely still retains the three-syllable rhythm of the target-like form though, and has added the past tense allomorph she is used to hearing to take the place of the deleted syllable of the stem.

AB does a similar thing with the present tense form of the class VII verb *naDa*, as seen in (67), where she deletes the second syllable of the stem. Once again, she stacks the unmarked present tense *-R-* on top of the target-like past tense form *-ndh-*.

(67)	*Speaker utterance*	*Target form*
Transc.ta	*இப்போ நந்தறே*	*நடக்கிறேன்*
Trans.la	ippo na**ndh**a**r**e:	naDa**kkar**e:n
MB	ippo [na-**ndh]**-a-**r**-e:	naDa-**kkar**-e:n
UR		naDa-**kkiR**-e:n
MG	now [walk-**PST]**-EUPH-**NTARG.PRS**-1S	walk-**PRS**-1S
Transl.	"I'm walking now"	

With the class V verb *ke:L*, AB produces target-like present tense utterances, as seen in (68). However, just as with DS, I argue that AB is still using the tense-stacking strategy to create these forms. With DS, I compared the present tense form of *ke:L* that she produced with the future tense form that she produced, which incorporated the *kk* of the present tense before stacking an unmarked future allomorph on top. In AB's case, she incorporates the *kk* of the present tense before stacking a past tense form on top, as seen in (69). She then immediately self-corrects to create a present-tense form. I believe the fact that the stacked past and present morphemes change in her self-correction but the *kk* stays is clear evidence of the underlying morphological breakdown of her utterance, where I argue the *kk* of the -*kkiR*- present allomorph is in fact being reanalysed to be part of the verb stem instead.

(68)	*Speaker utterance*	*Target form*
Transc.ta	நா கேக்குறே	கேக்கிறேன்
Trans.la	na: ke:**kkur**e:	ke:**kkir**e:n
MB	na: ke:-**kkur**-e:	ke:k-**kir**-e:n
UR	na: ke:L-**kiR**-e:n	ke:L-**kiR**-e:n
MG	I ask-**PRS-**1S	I ask-**PRS**-1S
Transl.	"I'm asking"	

(69)	*Speaker utterance*	*Target form*
Transc.ta	நா கேக்கிட்டே uh நா கேக்குறே	கேக்கறேன்
Trans.la	na: ke:**kk**iTTe: uh na: ke:**kk**u**r**e:	ke:**kki**r**e**:n
MB	na: ke:-**kk**-i-TT-e: uh na: ke:-**kk-**u-**r**-e:	ke:k-**ki**r-e:n
UR		ke:L-**kiR**-e:n
MG	I [ask-**PRS**-EUPH]-PST-1S uh I [ask-**PRS-**EUPH]**-PRS-**1S	I ask-**PRS**-1S
Transl.	"I asked- uh- I'm asking"	

AB produced the verb *kuDi* in various ways, and it was interesting to see how and when she would choose to self-correct. These self-corrections provided a valuable insight into how she was forming these words in her head, and consequently made the shape of their underlying morphology in her mental grammar much more transparent. Initially, AB produced the present tense form of *kuDi* by stacking the unmarked *-R-* on top of the target-like past allomorph *-chch-*, as seen in (70).

(70)	*Speaker utterance*	*Target form*
Transc.ta	பாட்டி தண்ணி குடி- குடிச்சிறா	குடிக்கிறா[65]
Trans.la	pa:Ti thanni kuDi— kuDi**ch**i**r**a:	kuDi**kkir**a:
MB	pa:Ti thanni kuDi— [kuDi-**ch]**-i-**r**-a:	kuDi-**kkir**-a:
UR		kuDi-**kkiR**-
		a:L/a:ngaL/a:r
MG	grandma water drink— [drink-**PST]-**EUPH**-**	I drink-**PRS**-
	NTARG.PRS-1S	3F/3PL/3HON
Transl.	"Grandma is drinking water" (intended)	

In a subsequent utterance (71), however, AB briefly starts forming the target-like form by incorporating the present tense *kk*, before changing her mind and reverting back to the past tense allomorph instead. In this utterance, she doesn't add a present tense morpheme at all, opting to leave the verb at just the stem form.

(71)	*Speaker utterance*	*Target form*
Transc.ta	பசங்கு தண்ணி குடிக்கு- குடுச்சி	கிடிக்கிறா
Trans.la	pasangu thanni kuDi**k**u— kuDu**ch**i	kuDi**kkir**a:
MB	pasangu thanni kuDi-**k-**u—[kuDu-**ch]**-i	kuDi-**kkir**-a:
UR		kuDi-**kkiR**-
		a:ngaL
MG	kids water drink-**PRS**-EUPH [drink-**PST]**-EUPH	I drink-**PRS**-3PL
Transl.	"(the) kids are drinking water" (intended)	

[65] Note that the target forms in this section are based on the *vattara vazhakku* spoken by AB and her mother RB. The *-a:* variant of the 3rd feminine and plural is a marker of this variety, and I do not intend to imply with its inclusion here that it is the *definitive* variant to express these morphemes, just that it is the variant used by my speakers.

A few utterances later, AB plays around with different forms, starting with incorporating the past, then the present, and then definitively settling on the past allomorph (72). She stacks the unmarked present tense *-R-* on top of this. It is interesting to see that, similar to (60), when she is testing out the target-like *kk* form, she is incorporating this morpheme with the verb stem *kuDu*, rather than *kuDi*. It suggests that the self-correction may be partially motivated by uncertainty about which vowel is appropriate for the verb stem.

(72)	*Speaker utterance*	*Target form*
Transc.ta	*புண்ணு தண்ணி குடிச்சி- குடுக்கு- குடிச்சிறா*	*கிடிக்கிறா*
Trans.la	puNNu thanni kuDi**chch**i— kudu**kk**u— kuDi**chch**ira:	kuDi**kkir**a:
MB	puNNu thanni kuDi-**chch**-i—kudu-**kk**-u— [kuDi-**chch]**-i-**r**-a:	kuDi-**kkir**-a:
UR		kuDi-**kkiR**- a:L
MG	girl water drink-**PST**-EUPH drink-**PRS**-EUPH [drink-**PST]**-EUPH-**NTARG.PRS**-3F	I drink-**PRS**-3F
Transl.	“the girl is drinking water” (intended)	

Following this utterance, AB seems to settle on stacking the present tense allomorph on top of the past tense. She does this in two similar utterances, one of which is shown in (73).

(73)	*Speaker utterance*	*Target form*
Transc.ta	பாட்டி தண்ணி குடிச்சிறா	கிடிக்கிறா
Trans.la	pa:TTi thaNNi kuDi**chchi**ra:	kuDi**kkir**a:
MB	pa:TTi thaNNi [kuDi-**chch]**-i-**r**-a:	kuDi-**kkir**-a:
UR		kuDi-**kkiR**-
		a:L/a:ngaL/a:r
MG	grandma water [drink-**PST]**-EUPH-**NTARG.PRS**-3F	I drink-**PRS**-
		3F/3PL/3HON
Transl.	“Grandma is drinking water” (intended)	

Following these two utterances, however, she then produces two target-like forms by incorporating the present tense *kk* before stacking the unmarked *-R-*; one of these utterances is shown in (74).

(74)	*Speaker utterance*	*Target form*
Transc.ta	பயென் குடிக்கிறா	கிடிக்கிறன்
Trans.la	payen kuDi**kkir**a:	kuDi**kkir**a:n
MB	payen kuDi-**kkir**-a:	kuDi-**kkir**-a:n
UR	payen kuDi-**kkir**-a:n	kuDi-**kkiR**-a:n
MG	boy drink-**PRS**-3M	I drink-**PRS**-3M
Transl.	“(the) boy is drinking” (intended)	

In her last present tense utterance of the verb *kuDi* (75), AB once again plays around with different forms. First she stacks two past tense forms, then she stacks two present tense forms, and then she settles on stacking the present *-R-* on top of the past *-chch-*, as she did in the majority of the present tense *kuDi* utterances. I believe the uncertainty here was

initiated by competing self-priming from some of her preceding utterances, which included past and future utterances of *kuDi* ([(60)] shown in 4.3.2 and [(88)] shown in 4.3.4).

(75)	*Speaker utterance*	*Target form*
Transc.ta	*நா குடிச்சிட்- uh நா குடுக்குறே uh குடுச்சிறே*	குடிக்கிறேன்
Trans.la	na: kuDi**chiT**— uh na: kuDu**kkur**e: uh kuDu**chi**re:	kuDi**kki**re:n
MB	na: [kuDi-**ch]**-i-**T**— uh na: kuDu-**kkur**-e: uh [kuDu-**ch]**-i-**r**-e:	kuDi-**kki**r-e:n
UR		kuDi-**kkiR**-e:n
MG	I [drink-**PST]**-EUPH-**leave.PST** uh I drink-**PRS**-1S uh [drink-**PST]**-EUPH-**NTARG.PRS**-1S	I drink-**PRS**-1S
Transl.	"I'm drinking" (intended)	

Again, the target-like *kk* form is attached to the verb stem *kuDu*, rather than *kuDi*, leading me to strongly believe that AB associates *kuDukk-* with the verb 'give', and *kuDichch* with the verb 'drink', with the incorporated stem-final geminates *kk* vs *chch* being the distinguishing feature between the two forms rather than the final vowel.[66]

4.3.4 Past

Of the three tenses, AB demonstrated the most variety with how she constructed past tense utterances. She used a combination of four overarching strategies summarized in Table 20: using a target-like past allomorph, using an unmarked, generalized past allomorph, tense-

[66] Even for L1 speakers, tense allomorphy sort of is a distinguishing feature between the verbs *kuDu* and *kuDi*, though they both belong to class VI. Recall that *kuDi* can take the colloquial form *-chch-* for its past allomorph in place of the formal *-thth-*, while *kuDu* can't. This, of course, isn't the distinction that AB makes, but it is interesting that she does still lexically distinguish the verbs using past tense allomorphy.

stacking, or creating a compound verb construction. I discuss each strategy in turn in the following sections.

Note that the single past utterance of the verb *paNNu* is actually a non-finite verbal participle form. I have opted to include it here as the verbal participle is homophonous with the past tense.

Verb Class	Verb	target-like past allomorph	CVC	unmarked past allomorph *-i(n)*	tense-stacking
I (*-dh-*)	azhu *(cry)*			6	
II (*-ndh-*)	va: *(come)*		1 (3N, *a:ka*)		
	vizhu *(fall)*		1 *(viDu)*	2	3
III (*-in-* or *-i-*)	(English verb) + paNNu *(do)*		1 *(viDu)*		
	pa:Du *(sing)*	2		1	
	po: *(go)*	2	3 *(viDu)*		1
	sol *(say)*		1 *(viDu)*		
	va:ngu	1			
	viLaiya:Du	1			
IV (*-TT-*)	sa:ppiDu *(eat)*	4 (2 are 3N)			
V (*-T-*)	ke:Lu *(ask)*	3			3
VI (*-thth-*)	kuDi *(drink)*		4 *(viDu)*		
	kuDu *(give)*	1			
	ninai *(think)*	1			1
VII (*-ndh-*)	naDa *(walk)*				2
Total	45	15	11	9	10

Table 20: Strategies AB used to mark the past tense

Strategy 1: target-like past allomorph

A third of the time, AB used the strategy of target-like past allomorphy to express the past tense. AB used this strategy with verbs from classes III, IV, V, and VI. I present examples of each in turn. Note that some of the forms may not be target-like as a whole because of variations in their PNG morphology, but these features are left to be discussed in chapter 4.4.

AB uses the class III past allomorph *–(i)n-* with the class III verbs *paDu* (76), *po* (77), *va:ngu* (78), and *viLaiya:Du* (79), as would be expected.

(76)	*Speaker utterance*
Transc.ta	*நா பாட்டு பாடினேன்*
Trans.la	na: pa:TTu pa:D**in**e:n
MB	na: pa:TTu pa:D-**in**-e:n
MG	I song sing-**PST**-1S
Transl.	"I sang (music/songs)"

(77)	*Speaker utterance*
Transc.ta	*அப்புறம் the friend pet store போ[...] போனம்*
Trans.la	appram the friend pet store po[…] po**n**am
MB	apram the friend pet store po[…] po-**n**-am
MG	afterwards the friend pet store go […] go-**PST**-PNG.UNC
Transl.	"afterwards, the friend went to the pet store" (intended)

(78)	*Speaker utterance*
Transc.ta	அதுக்கு அப்புறம் *bone fish and carrot* வாங்கினம்
Trans.la	adhukkappram avaL bone fish and carrot va:ng**in**am
MB	adhu-kk-appram avaL bone fish and carrot va:ng-**in**-am
MG	that-DAT-after she bone fish and carrot buy-**PST**-PNG.UNC
Transl.	"after that, she bought bone, fish, and carrot" (intended)

(79)	*Speaker utterance*
Transc.ta	நான் விலையாடினேன்
Trans.la	na:n vilaya:D**in**e:n
MB	na:n vilaya:D-**in**-e:n
UR	na:n viLaiya:D-**in**-e:n
MG	I play-**PST**-1S
Transl.	"I played"

AB uses the form *-TT-* to mark the past with the class IV verb *sa:ppiDu* (80) and the class V verb *ke:L* (81). Recall that the expected past allomorph for class IV is reduplication of the stem-final consonant, and the expected past allomorph for class V is *-T-*, and that because of their respective phonologies, ultimately this results in the identical form *-TT-* for both verbs. AB uses the *-TT-* form twice with the verb *sa:ppiDu*, and three times with the verb *ke:L*.

(80)	*Speaker utterance*
Transc.ta	நா *lunch* சாப்பிட்டே
Trans.la	na: lunch sa:p**TT**e:
MB	na: lunch sa:pT-**T**-e:
UR	na: lunch sa:ppiT-**T**-e:n
MG	I lunch eat-**PST**-1S
Transl.	"I ate lunch"

(81)	*Speaker utterance*
Transc.ta	நா கேட்டே
Trans.la	na: ke:**TT**e:
MB	na: ke:T-**T**-e:
UR	na:(n) keL-**T**-e:n
MG	I ask-**PST**-1S
Transl.	"I asked"

AB uses the colloquial third singular neuter past form *-chchu-* twice with the verb *sa:ppiDu*, as exemplified in (82).

(82)	*Speaker utterance*
Transc.ta	*cat* சாப்டிச்சு
Trans.la	cat sa:pTi**chchu**
MB	cat sa:pT-**ichchu**
MG	cat eat-**PST.3SN**
Transl.	"the cat ate (it)"

Finally, AB produces the class VI past form *-thth-*, and its colloquial variant *-chch-*, with the class VI verbs *kuDu* 'give' (83) and *ninai* 'think' (84). While most class VI verbs can alternate between the *-thth-* and the *-chch-* past tense forms, *kuDu* cannot and can only carry the *-thth-* form. We see here that AB is likely aware of this exception, as she uses the colloquial *-chch-* with *ninai* but not with *kuDu*.

(83)	*Speaker utterance*
Transc.ta	*அப்புறம் snake-உ bone, fish, and carrot குட் – குடுத்தா*
Trans.la	appram snake-u bone, fish, and carrot kuD—kuD**th**a:
MB	appram snake-u bone, fish, and carrot kuD—kuD-**th**-a:
UR	appram snake-u bone, fish, and carrot kuD—kuDu-**thth**-a:L
MG	after snake-acc bone fish and carrot give-- give-**PST**-3F
Transl.	"then, she gave the snake (the) bone, fish, and carrot"

(84)	*Speaker utterance*	*Target form*
Transc.ta	*அப்புறம் நன்ச்சி எந்து animal Fluffy*	*நினைச்சா*
Trans.la	appram nan**chi** endhu animal Fluffy	ninai**chch**a:
MB	appram nan-**ch(i)** endhu animal Fluffy	ninai-**chch**-a:
UR		ninai-**thth**-a:L
MG	after think-**PST** which animal Fluffy	think-**PST**-3F
Transl.	"then (she) thought, what (kind of) animal is Fluffy?"(intended)	

Strategy 2: compound verb constructions

The strategy used the most after using a target-like past allomorph was compound verb constructions, used in about a quarter of past-tense utterances. Like DS, AB's most-used

auxiliary for creating past-tense CVCs was the verb *viDu*. Most of the time, this surfaced as the colloquially reduced form *-TT-*, as demonstrated in (85), (86), (87), and (88).

(85)	*Speaker utterance*	*Target form*
Transc.ta	*நா வீந்துட்டே*	*வழிந்துட்டேன்*
Trans.la	na: vi:ndhu**TT**e:	vizhundhu**TT**e:n
MB	na: vi:-ndhu-**TT**-e:	vizhu-ndhu-**TT**-e:n
UR	na: vizhu-ndh(u)-**(vi)TT**-e:n	vizhu-ndh(u)-**viTT**-e:n
MG	I fall-VBPL-(**leave).PST**-1S	fall-VBPL-(**leave).PST**-1S
Transl.	"I fell"	

(86)	*Speaker utterance*	*Target form*
Transc.ta	*நா school-க்கு பெய்ட்டே*	*பொயிட்டேன்*
Trans.la	na: school-kku pei**TT**e:	po:i**TT**e:n
MB	na: school-kku pe-i-**TT**-e:	po:-i-**TT**-e:n
UR	na: school-kku po:-i-**(vi)TT**-e:n	po:-i-**(vi)TT**-e:n
MG	I school-DAT go-VBPL-**(leave).PST**-1S	go-VBPL-**(leave).PST**-1S
Transl.	"I went to school"	

(87)	*Speaker utterance*
Transc.ta	*அவல் செரி சொல்ட்டா*
Trans.la	aval seri sol**T**a:
MB	aval seri sol-ø-**T**-a:
UR	avaL seri sol-i-**(vi)TT**-a:L
MG	she ok say-VBPL-**(leave).PST**-3F
Transl.	"she said ok"

(88)	*Speaker utterance*	*Target form*
Transc.ta	தன்னி குடுச்சிட்டே	குடுச்சிட்டே
Trans.la	thanni kuDuchi**TT**e:	kuDuchi**TT**e:n
MB	thanni kuDu-chi-**TT**-e:	kuDu-chi-**TT**-e:n
UR	thaNNi kuDi-chi-**(vi)TT**-e:n	kuDi-chi-**(vi)TT**-e:n
MG	water drink-VBPL-**(leave).PST**-1S	water drink-VBPL-**(leave).PST**-1S
Transl.	"I drank water / I finished drinking water"	

Note that *-TT-* is identical to the past tense form AB used in (80) and (81) with class IV and V verbs *sa:ppiDu* and *ke:L*. As such, it is possible that what is actually happening here is tense-stacking, with *-TT-* being another generalized unmarked past allomorph like *–(i)n-* (discussed later in this section) that is being stacked on top of an incorporated target-like past tense morpheme.

In one instance, AB used the 3rd singular neuter form *-chchu*, as seen in (89). As *va* is a fairly common verb, it's possible this is a memorized chunk, but it's hard to draw any strong conclusions from just this singular token.

(89)	*Speaker utterance*
Transc.ta	*and* உரு *snake* வந்துச்சு
Trans.la	and uru snake vandhu**chchu**
MB	and uru snake va-ndhu-**chchu**
MG	and a snake come-VBPL- **leave.PST.3SN**
Transl.	"and a snake came/arrived"

Strategy 3: unmarked generalized past allomorph

In a fifth of the cases, AB used a less marked "generalized" past tense allomorph in place of the target-like one we would expect based on verb class. Most of the time, this was the form

-in-, which is the form normally associated with verb class III. (Note that most of the target-like utterances from at the start of this section were with class III verbs.) An example of this is shown in (90) with the class I verb *azhu* 'cry'.

(90)	*Speaker utterance*	*Target form*
Transc.ta	நா அரினே	அழுதேன்
Trans.la	na: ar**in**e:	azhu**dh**e:n
MB	na: ar-**in**-e:	azhu-**dh**-e:n
MG	I cry-**NTARG.PST**-1S	cry-**PST**-1S
Transl.	"I cried" (intended)	

Another instance of using the *-in-* allomorph was with the class II verb *vizhu* 'fall', as seen in (91). Here, it is interesting that AB produces the stem *vi:ng-* rather than *vizh-*. As was briefly touched on in 4.3.3 and will be elaborated in 4.5, AB appears to delete the /zh/ of the stem. I'm not sure why she replaces it with /ng/; one guess is that it's to maintain the rhythm and syllable structure of the target form; or possibly she is recalling the intransitive verb *vi:ngu* 'to swell' and is conflating the two forms? Ultimately it was unclear to me what the underlying morphology of her utterance was here.

(91)	*Speaker utterance*	*Target form*
Transc.ta	நா வீங்கினே	விழுந்தேன்
Trans.la	na: vi:ng**in**e:	vizhu**ndh**e:n
MB	na: vi:ng-**in**-e:	vizhu-**ndh**-e:n
MG	I fall-**NTARG.PST**-1S	fall-**PST**-1S
Transl.	"I fell" (intended)	

In one instance, AB briefly uses the past allomorph *-TT-* to create a non-target-like past form of *paDu* (92), before immediately self-correcting to the past allomorph *-in-* to create the target-like form *pa:Dine:n* 'I sang'.

(92)	*Speaker utterance*	*Target form*
Transc.ta	நா பாட்டு பாட்டே — பாடினேன்	பாடினேன்
Trans.la	na: pa:TTu pa:**TT**e: — pa:D**in**e:n	
		pa:D**in**e:n
MB	na: pa:TTu pa:-**TT**-e: — pa:D-**in**-e:n	pa:D-**in**-e:n
MG	I song sing-**NTARG.PST**-1S — sing-**PST**-1S	sing-**PST**-1S
Transl.	"I sang a song"	

AB produced this form shortly after producing the form *sa:pTe:* 'I ate', where *-TT-* is used to mark the past tense with the class IV verb *sa:ppiDu*. I strongly suspect this previous utterance "primed" AB into instinctively using the *-TT-* allomorph, before switching to the expected *-in-*. It is nevertheless interesting that she did so, as the second most common past tense allomorph AB used (as described earlier), was indeed *-TT-*. Additionally, as was discussed, it is also the phonological form of the past tense of the auxiliary *viDu*, which was used frequently in past tense CVCs. Thus, it is possible that *-TT-* is also a relatively unmarked past allomorph for AB and may be being used in alternation with *-in-* as a generalized past tense morpheme.

Strategy 4: tense-stacking

AB employed the tense-stacking strategy with the verbs *vizhu* (II), *po:* (III), *ke:L* (V), *ninai* (VI) and *naDa* (VII). With the exception of *po:*, AB also used tense-stacking with verbs from each of these classes in either future and/or present constructions. Most of the time, AB stacked the unmarked allomorph *-in-* on top of an existing tense marker, as seen in (93) with the class II verb *vizhu* (which we saw previously in (66)] in 4.3.3) and in (94) with the class VII

verb *naDa*. In both these cases, *-in-* is stacked on top of the expected target-like past tense marker *-ndh-*.

(93)	*Speaker utterance*	*Target form*
Transc.ta	*நா வீந்துனே? or நா வீந்துறே*	*விழுந்தேன்*
Trans.la	na: vi:**ndhun**e? or na: vi:ndhure	vizhu**ndh**e:n
MB	na: [vi:-**ndh]-u-n**-e? or na: vi:-ndh-u-r-e	vizhu-**ndh**-e:n
MG	I [fall-**PST]-EUPH-NTARG.PST**-1S or I fall-PST-EUPH-PRS-1S	fall-**PST**-1S
Transl.	"'I fell'? Or 'I am falling'" (intended)	

(94)	*Speaker utterance*	*Target form*
Transc.ta	*நா நந்தினே*	*நடந்தேன்*
Trans.la	na: na**ndhin**e:	naDa**ndh**e:n
MB	na: [na-**ndh]-in**-e:	naDa-**ndh**-e:n
MG	I [walk-**PST]-NTARG.PST**-1S	walk-**PST**-1S
Transl.	"I walked" (intended)	

In (95), while *-in-* is the expected past tense marker for the class III verb *po:*, it is following the stem *po:ga* rather than *po:*. This is in fact a form from Old Tamil (Subrahmanyam, 1971), but is uncommonly used now in this context. It is interesting that in (77), AB uses the modern form *po:* but uses the Old Tamil form *po:gu* in this utterance. Note that *po:ga* is the infinitival form of *po:*, with */-ka-/ [-ga-]* being the infinitive morpheme. Thus it is possible that AB here isn't using an Old Tamil form, but is instead employing a form of tense-stacking by incorporating the infinitive and adding the past morpheme on top. She is stacking two morphemes that are both indeed expected with the verb *po:* - they're just not expected to

appear together anymore in modern parlance.[67] Again, this is more evidence that these heritage language structures are in fact natural language structures.

(95)	*Speaker utterance*	*Target form in Modern Tamil*
Transc.ta	*நா school-க்கு போகினே*	*போனேன்*
Trans.la	na: school-kku po:**gin**e:	po:**n**e:n
MB	na: school-kku po:**g-in**-e:	po:-**n**-e:n
UR		po:-**n**-e:n
MG	I school-DAT go-**PST**-1S	go-**PST**-1S
Transl.	"I went" (intended)	

AB used tense-stacking with the class V verb *ke:L* several times, but she showed a lot of variation with the morphemes she chose to stack in each instance. In (96), she stacks the unmarked *-in-* on top of the present tense *-kk-*.

(96)	*Speaker utterance*	*Target form*
Transc.ta	*நா கேக்கினே*	*கேட்டேன்*
Transc.la	na: ke:**kkin**e:	ke:**TT**e:n
MB	na: [ke:-**kk]-in**-e:	ke:T-**T**-e:n
UR		ke:L-**T**-e:n
MG	I [ask-**PRS]-NTARG.PST**-1S	ask-**PST**-1S
Transl.	"I asked" (intended)	

[67] Note it is definitely possible that this form does still get commonly used in a *vattara vazhakku* I am not familiar with. However, AB and I speak mostly the same *vattara vazhakku* so I feel fairly confident in positing it is not a form her family would frequently use.

Then in (97), she stacks the same unmarked *–(i)n-*, but this time on top of the expected past tense *-TT-*.

(97)	*Speaker utterance*	*Target form*
Transc.ta	*நா கேட்டுனே*	*கேட்டேன்*
Transc.la	na: ke:**TTun**e:	ke:**TT**e:n
MB	na: [ke:-**TT]**-un-e:	ke:T-**T**-e:n
UR		ke:L-**T**-e:n
MG	I [ask-**PST]-NTARG.PST**-1S	ask-**PST**-1S
Transl.	"I asked" (intended)	

Finally, in (98) (duplicated from [(69)] in 4.3.3) she stacks the target-like past allomorph *-TT-* (which is also the other unmarked form briefly touched on earlier), on top of the present tense *-kk-*.

(98)	*Speaker utterance*	*Target form*
Transc.ta	*நா கேக்கிட்டே uh நா கேக்குறே*	*கேட்டேன்*
Transc.la	na: ke:**kkiTT**e: uh na: ke:kkure:	ke:**TT**e:n
MB	na: [ke:-**kk]-i-TT**-e: uh na: ke:-kk-u-r-e:	ke:T-**T**-e:n
UR		ke:L-**T**-e:n
MG	I [ask-**PRS]-EUPH-PST**-1S uh I [ask-PRS-EUPH]-PRS-1S	ask-**PST**-1S
Transl.	"I asked- uh- I'm asking"	

Thus, she variably incorporates either the present or the past tense morphemes into the *ke:L* verb stem, and variably stacks either of the two unmarked past markers, *-in-* or *-TT-*, on top.

Finally, (99) is an example where AB doesn't stack either of the unmarked past morphemes we've seen so far. In this example she is producing the class VI verb *ninai*, which takes as its past tense morpheme either the formal *-thth-* or the colloquial *-chch-*. In this utterance, AB stacks the formal variant on top of the colloquial one.

(99)	*Speaker utterance*	*Target form*
Transc.ta	அதுக்கு அப்புறம் அந்த ஆலு நன்ச்சித்தே	நினைச்சான் / நினைத்தான்
Transc.la	adhukkappram andha aalu nan**chiththh**e:	ninai**chch**a:n / ninai**thth**a:n
MB	adhu-kk-appram andha aalu [nan-**ch]-i-thth**-e:	ninai-**chch/thth**-a:n
UR		ninai-**thth**-a:n
MG	that-DAT-after that guy [think-**PST]-EUPH-PST**-1S	think-**PST**-3M
Transl.	"after that, that guy thought […]" (intended)	

4.3.5 Summary of tense morphology

AB produced target-like utterances the majority of the time. With the target-like utterances, she either used the strategy of a target-like tense allomorph (about 50% of the total utterances), or the strategy of creating CVCs (a little under 20% of total utterances.) In about a quarter of the tokens, AB used the strategy of tense-stacking, and in 9 cases she replaced the expected tense allomorph with its generalized unmarked counterpart.

The strategy that AB preferred depended on the tense. For the future, she strongly preferred creating CVCs using the *po:* 'go' auxiliary. The challenges that arose here were with creating the non-finite version of the preceding lexical verb: rather than the expected infinitive (which patterns with the present tense), AB frequently used the verbal participle form (which is homophonous to the past tense) As such, these utterances could also

reasonably be examples of tense-stacking. With the present tense utterances, AB preferred using target-like tense allomorphs, while still employing some tense stacking. With the past tense, she used all four strategies with a relatively even distribution.

AB showed some similarities with DS in her use of tense-stacking, as well as some differences. As with DS, AB never incorporated a future tense morpheme into a verb stem, and I suspect for the same reasons as I posited with DS. DS mainly lexicalized the present tense, while AB lexicalized both the present and past, generally preferring the past. Sometimes she would switch between lexicalizing either the past or present with the same verb, as was seen with the *ke:L* and *kuDi* examples. This last point in particular I believe suggests that AB's grammar, like DS's, contains stable variation, or may still be changing and undergoing reanalysis.

As with DS, AB appears to consider the most commonly used or most salient allomorphs for each tense to be "unmarked" and generalizes them across verb classes,[68] and then incorporates the more marked, "irregular" allomorphs into the verb stem. This means with the present tense, she uses the tense-stacking strategy with verbs from classes V, VI, and VII to either incorporate the more marked *-kkiR-* allomorph and/or stack the less marked present allomorph *-R-* on top.

With the past tense, AB considers both the class III *-in-* and the versatile *-TT-* to be unmarked. She often either replaces the expected past tense allomorph with one of these forms, or stacks it on top of an incorporated "irregular" past allomorph like *-ndh-* or *-chch-*. I believe AB views *-TT-* as an unmarked past allomorph because it appears in many contexts: not only is it the surface realization of the past allomorphs of the class IV and V verbs used in this study, it is also the colloquially reduced form of the completive auxiliary *viDu*, which both DS and AB used frequently in building CVCs to express the past tense. The allomorph -

[68] Again, here, I'm simplifying by considering the most common allomorph to be the one used with the most verb classes. This would be best validated by also checking the distribution of verbs across all classes and cross-validating with the most commonly used verbs.

in-, on the other hand, I believe is unmarked because it is more perceptually salient than the other past allomorphs. It is phonemically very different from the other forms, containing no coronal obstruents. It is also the only form that contains a vowel, making it the only form that can stand alone as its own syllable since it contains a nucleus. I believe the fact that this form is its own syllable makes it more perceptually salient, because it is now adding to the rhythm of a word in a way the other forms don't, and so AB is recognizing and identifying it more overtly as an explicit marker of the past tense. Ultimately, the fact that AB varies which unmarked form she uses when tense-stacking, often with the same verb, is either an indicator of stable variation, or that her grammar is still changing; a long-term study or a follow-up study in a few years would be required to validate either hypothesis.

With the exception of a few of the *ke:L* and *kuDi* utterances, AB incorporated past allomorphs rather than present allomorphs into verb stems when tense-stacking. I believe this is because tense allomorphs show the most variation with the past forms, compared to the present and future, which only really have two or three forms that are clearly phonologically related. Unless past tense morphology has already been acquired, it is difficult to intuitively make the connection that the forms *-dh, -ndh-, -in-, -T, -thth-* (and its variant *-chch-*), and the reduplicated allomorph all have the same semantic and functional meaning. It feels much more natural to assume that these are not distinct morphemes, but rather segments that are part of the verb stem; and thus that another tense allomorph has to be added in order to explicitly convey tense.

Across all three tenses, AB used tense-stacking with the class II verb *vizhu*. I believe that this was motivated purely by phonology. In every *vizhu* utterance, the final retroflex rhotic (which I transcribe as /zh/), was omitted, which reduced the number of syllables in the stem. I believe the past allomorph was incorporated into the stem to make up for this gap. I discuss this example, and the general phonological process of lenition present in both speakers, in more depth in 4.5.

It bears repeating that the strategy of tense-stacking that both AB and DS used demonstrates deep understanding of Tamil verb morphology. It involves identifying both irregular and regular (read: salient, frequent, and predictable) morphology, and lexicalizing the first and generalizing the latter. AB put her own spin on this by using the strategy of a generalized unmarked allomorph: this is basically tense-stacking but skips the step of incorporating an "irregular" morpheme into the verb stem. AB showed a lot of variation in her forms, using different strategies within the same verb class or even the same verb. As well, though she frequently incorporated past tense allomorphs into verb stems as part of tense stacking, it is important to note that she did also produce many target-like past allomorphs, accurately producing forms that in other utterances, she lexicalized. All of this variation present in AB's speech could be because her grammar is not fixed and continues to change. It would be interesting to see if the variation ever settles, and if so, whether it will be conditioned by verb class, tense, phonology, the lexicon, or a combination of different factors.

4.4 Speaker 2 (AB) PNG morphology

Similar to DS, AB also generally used target-like PNG forms. Table 2 from 2.2.2 is partially duplicated here for reference, including only the persons that AB produced in her speech.

Person-number-gender	*Centamil* form	Common *koDuntamil* forms
1S	*-e:n*	*-ẽ:*
1PL.EXC and 1PL.INC	*-o:m*	*-õ:*
3F	*-a:L*	*-a:*
3M	*-a:n*	*-ã:*
3HON	*-a:r*	*-a:*
3SN	*-adhu*	
	(for class III verbs) *-iRRu* (incorporates past tense)	*-chchu* (incorporates past tense)
	-um (incorporates future tense)	Many compound verb forms exist that fuse the 3SN morpheme with preceding auxiliaries.
3PL	*-a :rgaL*	*-a:ngaL* *-a:nga* ***-a:*** *-a:L*

Table 21: Person-number-gender morphemes in Tamil (reduced for AB)

There were several instances where AB hesitated over the PNG morpheme, and it was clear that PNG was a source of uncertainty especially in the utterances where she omitted the PNG entirely, instead ending the verb at the tense morpheme, as seen in

(100), (101), and (102). Interestingly, in these cases she always ended with the past tense morpheme, though in (102) the intended meaning is the present. This makes it unclear whether the form is meant to indeed be a tense morpheme, or instead a verbal participle (which is homophonous to the past and appears in compound verb constructions without a PNG morpheme).

(100)	*Speaker utterance*	*Target form*
Transc.ta	அவல் கேட்டு	கேட்டா
Trans.la	aval ke:**TTu**	ke:TT**a:**
MB	aval ke:T-**Tu**	ke:T-T-**a:**
UR		ke:L-T-**a:L**
MG	she ask-**PST/VBPL**	ask-PST-**3F**
Transl.	"she asked" (intended)	

(101)	*Speaker utterance*	*Target form*
Transc.ta	அப்புறம் நன்ச்சி எந்து animal Fluffy	நன்ச்சா
Trans.la	appram nan**chi** endhu animal Fluffy	nanchch**a:**
MB	appram nan-**ch(i)** endhu animal Fluffy	nan-chch-**a:**
UR		ninai-thth-**a:L**
MG	after think-**PST** which animal Fluffy	think-PST-**3F**
Transl.	"then (she) thought, what (kind of) animal is Fluffy?"(intended)	

(102)	*Speaker utterance*	*Target form*
Transc.ta	பசங்கு தண்ணீ குடிக்கு- குடுச்சி	குடிக்கிறா/கிடிக்கிறாங்க
Trans.la	pasangu thanni kuDi**kku**— kuDu**chi**	kuDikkir**a:**/kuDikkir**a:nga**
MB	pasangu thanni kuDi-**kku**— kuDu-**chi**	kuDi-kkir-**a:**/**a:nga**
UR		kuDi-kkir-**a:ngaL**
MG	kids water drink-**PRS** drink-**PST**	drink-PST-**3PL**
Transl.	"(the) kids are drinking water" (intended)	

AB frequently used the following four forms to express PNG: *-e:*, *-a:*, *-um*, and *-am*. *e:* is associated with 1st person singular and *a:* is associated with every 3rd person form.[69] The last two forms, *-um* and *-am*, are possibly derived from the 3rd singular neuter future *-um*, or the 1st plural *-o:m*, both of which appear in L1 speech.

As seen in (103), AB uses *-e:(n)* in most 1st person singular utterances. The exception is shown in (104), where she instead uses *-am* with the verb *pa:Du*. Considering that she uses *-e:* in the 62 other 1st person singular utterances, (including with the verb *pa:Du*), I'm not sure why she uses *-am* in this one instance; perhaps this is tied to the general uncertainty she seemed to have with expressing PNG.

[69] The 3rd person feminine form is *-a:L*, the 3rd person honorific is *-a:r*, and the 3rd person plural is *-a:ngaL*, but as is shown in Table 21, these are all often reduced to the colloquial *-a:* in some dialects. The 3rd person masculine *-a:n* and 1st person *e:n* end with nasal vowels, but as was touched on in 3.5.2, I was unable to determine if word-final nasals were present in speakers' utterances or not, and so for this analysis I will be ignoring this feature.

(103)	*Speaker utterance*
Transc.ta	*நா கேக்கப்போறேன்*
Trans.la	na: ke:kkappo:r**e**
MB	na: ke:k-k-a-ppo:-r-**e:n**
UR	na: ke:L-k-a-po:-kiR-**e:n**
MG	I ask-MID-INF-go-PRS-**1s**
Transl.	"I'm going to ask"

(104)	*Speaker utterance*	*Target form*
Transc.ta	*நா பாட்டு பாடப்போறம்*	*பாடப்போறேன்*
Transc.la	na: pa:TTu pa:Duppo:r**am**	pa:Duppo:r**e:n**
MB	na: pa:TTu pa:D-u-ppo:-r-**am**	pa:D-u-ppo:-r-**e:n**
UR		na: pa:TTu pa:D-a-ppo:-r-**e:n**
MG	I song sing-INF-go-PRS-**NTARG.1S**	sing-INF-go-PRS-**1S**
Transl.	"I'm going to sing a song" (intended)	

When expressing the 3rd person feminine or honorific, AB used the *-am* (105), (106), (107) and *-um* (108) forms in a few initial utterances, before settling into consistently using the target-like *-a:* form as seen in (109), (110), and (111).

(105)	*Speaker utterance*	*Target form*
Transc.ta	*அப்புறம் the friend pet store போ[...] போனம்*	*போனா*
Transc.la	appram the friend pet store po[…] po:n**am**	po:n**a:**
MB	apram the friend pet store po[…] po:-n-**am**	po:-n-**a:**
UR		po:-n-**a:L**
MG	afterwards the friend pet store go […] go-PST-**PNG.UNC**	go-PST-**3F**
Transl.	"afterwards, the friend went to the pet store" (intended)	

(106)	*Speaker utterance*	*Target form*
Transc.ta	*அதுக்கு அப்புறம் bone fish and carrot வாங்கினம்*	*வாங்கினா*
Transc.la	adhukkappram avaL bone fish and carrot va:ngin**am**	va:ngin**a:**
MB	adhu-kk-appram avaL bone fish and carrot va:ng-in-**am**	va:ng-in-**a:**
UR		va:ng-in-**a:L**
MG	that-DAT-after she bone fish and carrot buy-PST-**PNG.UNC**	buy-PST-**3F**
Transl.	"after that, she bought bone, fish, and carrot" (intended)	

(107)	*Speaker utterance*	*Target form*
Transc.ta	*பாட்டி பாடு- பாடரம்*	*பாடறா*
Transc.la	pa:Ti pa:Du— pa:Dr**am**	pa:Dar**a:**
MB	pa:Ti pa:Du— pa:D-r-**am**	pa:Da-r-**a:**
UR		pa:Du-kiR-**a:L/a:r**
MG	grandmother sing— sing-PRS-**NTARG.3F/3HON**	sing-PRS-**3F/3HON**
Transl.	“grandmother is singing” (intended)	

(108)	*Speaker utterance*	*Target form*
Transc.ta	*புண்ணு lunch சாப்ரம்*	*சாப்பட்றாள்*
Transc.la	puNNu lunch sa:pr**um**	sa:ppaDr**a:**
MB	puNNu lunch sa:p-r-**um**	sa:ppaD-r-**a:**
UR		sa:ppiDu-kiR-**a:L**
MG	girl lunch eat-PRS-**NTARG.3F**	eat-PRS-**3F**
Transl.	“(the) girl is eating lunch” (intended)	

(109)	*Speaker utterance*
Transc.ta	*புண்ணு பாட்டு பாடுறா*
Trans.la	puNNu pa:TTu pa:Dur**a:**
MB	puNNu pa:TTu pa:Du-r-**a:**
UR	poNNu pa:TTu pa:Du-kiR-**a:L**
MG	girl song sing-PRS-**3F**
Transl.	“(the) girl is singing a song”

(110)	*Speaker utterance*	*Target form*
Transc.ta	*பாட்டி தண்ணி குடி- குடிச்சிறா*	*குடிக்கிறா*
Trans.la	pa:Ti thanni kuDi— kuDichir**a:**	kuDikkir**a:**
MB	pa:Ti thanni kuDi— [kuDi-ch(i)]-r-**a:**	kuDi-kkir-**a:**
UR		kuDi-kkiR-
		a:L/a:ngaL/a:r
MG	grandma water drink- [drink-PST]-NTARG.PRS-	drink-PRS-
	3F/3PL/3HON	**3F/3PL/3HON**
Transl.	"Grandma is drinking water" (intended)	

(111)	*Speaker utterance*
Transc.ta	*பாட்டி பாட்டு பாடுறா- பாடுறா*
Transc.la	pa:Ti pa:TTu pa:Dur**a:**— pa:Dur**a:**
MB	pa:Ti pa:TTu pa:Du-r-**a:**— pa:Du-r-**a:**
UR	pa:Ti pa:TTu pa:Du-kiR-**a:L/a:ngaL/a:r**—
	pa:Du-kiR-**a:L/a:ngaL/a:r**
MG	girl song sing-PRS-**3F/3PL/3HON**— sing-PRS-
	3F/3PL/3HON
Transl.	"grandmother is singing a song"

With the 3rd person plural, AB produced *-um* and *-am* each once (112), (113), in alternation with the target-like *-a:* (114) that was used in five utterances. It is possible that in this case, the *-am* and *-um* forms are derived from the 1st plural *-o:m* form, and that AB is making an analogy between the two persons since they are both plural. Interestingly, in (113), AB first uses the 2nd person plural form *e:L*, which is target-like for

her *vattara vazhakku*, before changing to *-am*, and so she is clearly drawing an association between different plural forms.

(112)	*Speaker utterance*	*Target form*
Transc.ta	*the* பசங்க தண்ணீ குடிச்சீறும்	குடிக்கிறா(ங்க)
Transc.la	the pasanga thani kuDichi:r**um**	kuDikkir**a:(nga)**
MB	the pasanga thani kuDi-chi:-r-**um**	kuDi-kkir-**a:(nga)**
UR		kuDi-kkiR-**a:ngaL**
MG	the kids water [drink-PST]-NTARG.PRS-**NTARG.3PL**	drink-PRS-**3PL**
Transl.	"the kids are drinking water" (intended)	

(113)	*Speaker utterance*	*Target form*
Transc.ta	பசங்கு *lunch* சாப்றேல்- சாப்றம்	சாப்பிடறா(ங்க)
Transc.la	pasangu lunch sa:pr**el**— sa:pr**am**	sa:ppiDar**a:(nga)**
MB	pasangu lunch sa:p-r-**el**— sa:p-r-**am**	sa:ppiDa-r-**a:(nga)**
UR		sa:ppiDu-kiR-**a:ngaL**
MG	kids lunch eat-PRS-**2PL**— eat-PRS-**NTARG.3PL**	eat-PRS-**3PL**
Transl.	"(the) kids are eating lunch" (intended)	

(114)	*Speaker utterance*
Transc.ta	பசங்கு *lunch* சாப்றா
Transc.la	pasangu lunch sa:pr**a:**
MB	pasangu lunch sa:p-r-**a:**
UR	pasangu lunch sa:ppiDu-kiR-**a:ngaL**
MG	kids lunch eat-PRS-**3PL**
Transl.	"(the) kids are eating lunch"

When expressing the 3rd person masculine, AB consistently used the *-a:(n)* form, as seen in (115) and (116). Recall that the target-like form is *-a:n*, and recall also that it was not possible to consistently determine if the word-final nasal vowel was present or not.

(115)	*Speaker utterance*	*Target form*
Transc.ta	பயென் குடிக்கிறா	கிடிக்கிறான்
Trans.la	payen kuDikkir**a:**	kuDikkir**a:n**
MB	payen kuDi-kkir-**a:**	kuDi-kkir-**a:n**
UR	payen kuDi-kkir-**a:n**	kuDi-kkiR-**a:n**
MG	boy drink-PRS-**3M**	drink-PRS-**3M**
Transl.	"(the) boy is drinking" (intended)	

(116)	*Speaker utterance*
Transc.ta	பையன் பாட்டு பாடுறான்
Transc.la	payen pa:TTu pa:Dur**a:n**
MB	payen pa:TTu pa:Du-r-**a:n**
MG	boy song sing-PRS-**3M**
Transl.	"(the) boy is singing a song"

There were a few instances where AB intended a 3rd person singular subject but the gender was unclear. In (117), she uses the *-a:* form associated with the 3rd person generally, but the intended gender is unclear from the morphology or the context. In (118), though from context she clearly intends the 3rd person, she uses the *-e:* form associated with the 1st person singular instead. Note that (118) was one of the earlier utterances in the session, and it is possible that AB was simply not fully comfortable speaking in Tamil yet; she became more comfortable and less hesitant with her forms as the session went on.

(117)	*Speaker utterance*	*Target form*
Transc.ta	*உரு ஆலு friend பெய்ட்டா*	*பெய்ட்டா(ன்)*
Transc.la	uru aalu friend peiTT**a:**	peiTT**a:(n)**
MB	uru aalu friend pe-i-TT-**a:**	pe-i-TT-**a:(n)**
UR		po:-i-TT-**a:L/a:n**
MG	one guy friend go-PST-leave.PST-**3F**	go-PST-leave.PST-**3F/3M**
Transl.	"a guy [went to go see (?)] their friend" (unclear intended meaning)	

(118)	*Speaker utterance*	*Target form*
Transc.ta	*அதுக்கு அப்புறம் அந்த ஆலு நன்ச்சித்தே*	*நன்ச்சா(ன்)*
Transc.la	adhukkappram andha aalu nanchiththe**:**	nanchcha**:(n)**
MB	adhu-kk-appram andha aalu nan-chi-thth-**e:**	nan-chch-**a:(n)**
UR		ninai-thth-**a:L/a:n**
MG	that-DAT-after that guy think-PST-PST-**1S**	think-PST-**3F/3M**
Transl.	"after that, that guy thought […]" (intended)	

AB consistently used target-like forms to express the 3rd person singular neuter. She used a combination of simple PNG forms (119), fused past forms (120), (121), and fused future forms (122), (123). Notably, all but one of these forms are have underlyingly irregular structures (i.e., unpredictable), either because they are fused forms ([(120)], [(121)], [(122)], [(123)]) or because they are colloquial variants of productive, regular morphemes (64). It is interesting that AB seems to be familiar and comfortable with these forms despite them being irregular in these ways. Additionally, the forms *-am* and *-um* seen in several of the examples above seem to be related to the fused future tense and 3rd singular neuter form -

um. It's possible that the 3rd singular neuter forms are less marked for AB, and so she generalizes them to the other persons as well. Perhaps these forms are commonly used and so familiar to her (indeed, *irukku* and *irukkum* would be); research looking into what are the most commonly used PNG forms, specifically in heritage Tamil households, could shed some light on this.

(119)	*Speaker utterance*
Transc.ta	*எங்க Fluffy இருக்கு*
Trans.la	engu fluffy irukku
MB	enga fluffy iru-kku
UR	enga fluffy iru-[kkiR-adhu]
MG	where fluffy be-[PRS-3N]
Transl.	"where is Fluffy"

(120)	*Speaker utterance*
Transc.ta	*cat சாப்ட்டிச்சு*
Transc.la	cat sa:pTi**chchu**
MB	cat sa:pT-**ichchu**
MG	cat eat-**PST.3SN**
Transl.	"the cat ate (it)"

(121)	*Speaker utterance*
Transc.ta	*and உரு snake வந்துச்சு*
Transc.la	and uru snake vandhu**chchu**
MB	and uru snake va-ndhu-**chchu**
UR	
MG	and a snake come-VBPL-**PST.3SN**
Transl.	"and a snake came/arrived"

(122)	*Speaker utterance*
Transc.ta	*snake-உ என்ன சாப்டும்*
Transc.la	snake-u enna sa:pD**um**
MB	snake-u enna sa:pD-**um**
UR	snake-u enna sa:ppiD-**um**
MG	snake-EUPH what eat-**FUT.3SN**
Transl.	"what does a/the snake eat?"

(123)	*Speaker utterance*
Transc.ta	*…and fish cat க்கு இருக்கும்*
Transc.la	and fish cat kku irukk**um**
MB	and fish cat kku iru-kk-**um**
MG	and fish cat-DAT be-STR-**FUT.3SN**
Transl.	"…and the fish is / will be / would be / must be for the cat"

4.5 Speakers 1 (DS) and 2 (AB) phonology

The focus of this work is morphology, and so I did not include a detailed phonetic analysis, as was discussed in 3.5.2. However, there were some interesting phonological similarities between AB and DS that I thought were worth briefly highlighting. Hopefully the patterns discussed here could be the subject of fruitful future research.

I noticed two broad patterns of variation common to both speakers: the first is a tendency towards lenition, and the second is a merging of round vowels. I look at lenition first.

4.5.1 Lenition

Lenition occurred primarily with two phonemes: the retroflex voiced stop [D] (underlyingly the voiceless /T/) and the retroflex rhotic /zh/.

In intervocalic contexts, both AB and DS would often delete /D/ in verb stems, as seen in AB's utterances in (124), (125), and (126) and DS's utterances in (127), (128), and (129).[70] They didn't do this all the time; (130) and (131) (from AB and DS respectively) show the [D] phoneme being retained in a remarkably similar phonological environment to the ones in which it was deleted. I'm not entirely sure why this is - possibly the /ya:/ in the preceding syllable of *viLaiya:Du* (that was also deleted, as it commonly is colloquially), is conditioning [D] retention for some reason? Maybe deleting two syllables (as opposed to just one) is dispreferred? Ultimately with the given limited data, I can only present this as a variably occurring process of lenition.

[70] In this section, AB's examples are in blue and DS's are in purple to increase clarity.

(124)	*Speaker utterance*	*Target form*
Transc.ta	நா பாட்டு பாடினேன் oh பாட்டு பாறேன்	பாடுறேன்
Transc.la	na: pa:TTu pa:Dine:n oh pa:TTu **pa:**re:n	**pa:D(u)**re:n
MB	na: pa:TTu pa:D-in-e:n oh pa:TTu pa:-r-e:n	**pa:D(u)**-r-e:n
UR		**pa:Du**-kiR-e:n
MG	I song sing-PST-1S oh song sing-PRS-1S	sing-PRS-1S
Transl.	"I sang a song – oh – I'm singing a song"	

(125)	*Speaker utterance*	*Target form*
Transc.ta	இப்போ நா சாப்றே	சாப்பிடுறேன்
Transc.la	ippo na: **sa:p**re	**sa:ppiD**re:n
MB	ippo na: **sa:p**-r-e	**sa:ppiD**-r-e:n
UR		**sa:ppiDu**-kiR-e:n
MG	now I eat-PRS-1S	eat-PRS-1S
Transl.	"I'm eating now"	

(126)	*Speaker utterance*	*Target form*
Transc.ta	இப்போ நந்தறே	நடக்கிறேன்
Trans.la	ippo **na**ndhare:	**naDa**kkare:n
MB	ippo [**na**-ndh]-a-r-e:	**naDa**-kkar-e:n
MG	now [walk-PST]-EUPH-NTARG.PRS-1S	walk-PRS-1S
Transl.	"I'm walking now"	

(127)	*Speaker utterance*	*Target form*
Transc.ta	*dog-ஓட food-உ, bone சாப்புவோம்*	*சாப்பிடும்*
Trans.la	dog-o:Da food-u bone **sa:ppu**vo:m	**sa:pp(i)D**um
MB	dog-o:Da food-u bone **sa:ppu**-vo:m	**sa:pp(i)D**-um
MG	dog-SOC food-EUPH bone eat-FUT-1PL	eat-FUT.3SN
Transl.	'a dog's food; (it) eats/will eat bone(s)'	

(128)	*Speaker utterance*	*Target form*
Transc.ta	snake-உ என்ன சாப்புவோன்	*சாப்பிடும்*
Trans.la	snake-u yenna **sha:ppu**vo:n	**sa:pp(i)D**um
MB	snake-u yenna **sha:ppu**-vo:n	**sa:pp(i)D**-um
MG	snake-EUPH what eat-FUT.3SN	eat-FUT.3SN
Transl.	'what does/will a snake eat'	

(129)	*Speaker utterance*	*Target form*
Transc.ta	night ல நா நறுந்துவோம்	*நடப்பேன்*
Trans.la	night la na: **naRu**ndhuvo:m	**naDa**ppe:n
MB	night la na: **naRu**-ndhu-v-o:m	**naDa**-pp-e:n
MG	night LOC I [walk-PST]-NTARG.FUT-1PL	walk-FUT-1S
Transl.	'I'll walk at night'	

(130)	*Speaker utterance*
Transc.ta	பசங்க *game* வளாடறா
Transc.la	pasangu game **vaLa:D(a)**ra:
MB	pasangu game **vaLa:D(a)**-r-a:
UR	pasanga game **viLaiya:Du**-kiR-a:L
MG	children game play-PRS-3PL
Transl.	"(the) children are playing a game"

(131)	*Speaker utterance*
Transc.ta	பாட்டி விளையாடுறா
Trans.la	pa:TTi **vaLa:D**Ra:
MB	pa:TTi **vaLa:D**-R-a:
UR	pa:TTi **viLaiya:Du**-kiR-a:L
MG	grandmother play-PRS-3F
Transl.	'grandmother plays'

The other phoneme subject to frequent lenition was the retroflex rhotic /zh/. Recall from 2.4.1 that this is one of the last phonemes to be acquired. The two verb stems containing this phoneme that were part of the data elicitation were *azhu* 'cry' and *vizhu* 'fall'. DS and AB used similar pronunciation strategies for both these words. For instance, AB and DS both replaced the /zh/ in *azhu* with a different phoneme at least once and deleted the /zh/ in *vizhu* at least once.

AB replaced the /zh/ with /r/ in some utterances, as shown in (132), but accurately produced it using the /zh/ in others (133). She continued to go back and forth with these two pronunciations with subsequent utterances. It is interesting that in (132), rather than

deleting the /zh/, AB in fact switches the /zh/ and /r/ phonemes from their expected positions in an apparent instance of metathesis.

(132)	*Speaker utterance*	*Target form*
Transc.ta	இப்போ அரழேன்	அழறேன்
Trans.la	ippo ara**zh**e:n	a**zh**a**r**e:n
MB	ippo ara-**zh**-e:n	a**zh**a-**r**-e:n
UR	ippo a**zh**u-ki**R**-e:n	a**zh**u-ki**R**-e:n
MG	now cry-PRS-1S	cry-PRS-1S
Transl.	"I'm crying now"	

(133)	*Speaker utterance*	*Target form*
Transc.ta	இப்போ அழினே	அழுதேன்
Trans.la	ippo **azh**-in-e:	**azhu**dhe:n
MB	ippo **azh**-in-e:	**azhu**-dh-e:n
MG	I cry-NTARG.PST-1S	cry-PST-1S
Transl.	"I cried" (intended)	

DS lenited the /zh/ to a /y/ and then to an /r/ in her first utterance of the verb in (134). In her next utterance of the verb, she hesitated with the /zh/ phoneme but ultimately produced it in (135). In all subsequent utterances of *azhu* she used the /zh/ phoneme.

(134)	*Speaker utterance*	*Target form*
Transc.ta	*so இப்பொவே அய- அரேரன் நா*	*அழறேன்*
Transc.la	so ippove: **aya**— **ara**re:n na:	**azha**re:n
MB	so ippove: **aya**— **ara**-r-e:n na:	**azha**-r-e:n
UR		**azha**-R-e:n na:
MG	so now.itself cry— cry-PRS-1S I	cry-PRS-1S
Transl.	'so I'm crying now itself'	

(135)	*Speaker utterance*	*Target form*
Transc.ta	*morning ல அழ்—ழ--தான்*	*அழுதேன்*
Trans.la	morning la na: **azh—zhu**—dha:n	**azhu**dhe:n
MB	morning la **azh—zhu**-dh-a:n	**azhu**-dh-e:n
MG	morning LOC cry-PST-3M	cry-PST-1S
Transl.	'I cried in the morning' (intended)	

With the verb *vizhu,* rather than replacing the /zh/, DS and AB delete the phoneme in at least one of their utterances each.

DS initially deletes the /zh/ entirely when she first produces it in (136), but in all subsequent utterances she produces the /zh/, as exemplified in (137).

(136)	*Speaker utterance*	*Target form*
Transc.ta	*uh இப்பொவே வீந்துட்டேன் uh வீந்த்- yeah*	*விழுந்துட்டேன் uh விழுந்த்—*
Transc.la	uh ippove: **vi:**ndhuTTe:n uh **vi:**ndh— yeah	**vi:zhu**-ndhuTT-e:n uh **vi:zhu**ndh—
MB	uh ippove: **vi:**-ndhu-TT-e:n uh **vi:**-ndh— yeah	**vi:zhu**-ndhu-TT-e:n uh **vi:zhu**-ndh—
MG	uh now fall-VBPL-leave.PST-1S uh fall-VBPL yeah	fall-VBPL-leave.PST-1S uh fall-VBPL
Transl.	'I fell now itself'	

(137)	*Speaker utterance*
Transc.ta	*அப்ப morning ல நா விழுந்துட்டேன்*
Trans.la	appa morning la na: **vizh**ndhuTTe:n
MB	appa morning la na: **vizh**-ndhu-TT-e:n
MG	then morning LOC I fall-VBPL-leave.PST-1S
Transl.	'then, in the morning I fell' (implication of 'I finished falling' or 'I went and fell')

AB, on the other hand, consistently deletes the /zh/ phoneme when producing *vizhu*, as exemplified in (138). In AB's case, it is instructive to compare her speech to that of baseline speaker RB, her mother. RB consistently lenited the underlying /zh/ phoneme to the colloquially used variant /y/. I hypothesize that her daughter AB is simply taking this process of lenition one step further to delete the phoneme entirely. What remains to be studied is why AB consistently deleted /zh/ with *vizhu* but not with *azhu* (a possible explanation is /i/

being particularly incompatible with /zh/ in terms of tongue configuration), and more broadly, what phonological environments condition lenition and deletion in her speech.

(138)	*Speaker utterance*	*Target form*
Transc.ta	நா வீந்துறே	விழுறேன்
Trans.la	na: **vi:**ndhure:	**vizhu**re:n
MB	na: [**vi:**-ndh]-u-r-e:	**vizhu**-r-e:n
UR		**vizhu**-kiR-e:n
MG	I [fall-PST]-EUPH-PRS-1S	fall-PRS-1S
Transl.	"I'm falling" (intended)	

In sum, [D] was deleted in several utterances by both speakers, while /zh/ was either reduced to the alveolar glide /y/ or alveolar rhotic /r/, or deleted entirely. Both [D] and /zh/ are voiced retroflex sounds – possibly the retroflexion is challenging and the voicing makes them more vulnerable to deletion (e.g. as seen in Spanish [González, 2002]? Note that the underlying voiceless counterpart of [D] is /T/, which generally always appears as part of a geminate, thereby possibly making it less vulnerable to deletion as evidenced in DS's faithful utterance in (139) and AB's faithful utterance in (140). Ultimately, with the data on hand, there is only the evidence to demonstrate the variability in production, but not to strongly argue for a phonological reason why.

(139)	*Speaker utterance*
Transc.ta	அப்ப *morning* ல நா சாப்ட்டேன்
Trans.la	appa morning la na: **sa:pTT**e:n
MB	appa morning la na: **sa:pT-T**-e:n
UR	appo: morning la na: sa:ppi**T-T**-e:n
MG	then morning LOC I eat-PST-1S

Transl.	'then, I ate in the morning'

(140)	*Speaker utterance*
Transc.ta	*cat சாப்பிச்சு*
Transc.la	cat **sa:ppTT**ichchu
MB	cat **sa:ppTT**-ichchu
MG	cat eat-PST.3SN
Transl.	"the cat ate (it)"

4.5.2 Round vowel merger

Recall from 2.2.5 that underlying /u/ in the first syllable of some words often surfaces as [o]. The conditioning environment for this in L1 speech is when the second syllable of the target word has a low vowel (/a/, /a:/, or /ai/) and there's only one singleton consonant (so excluding geminates) between the two vowels. In DS and AB's speech, however, this vowel alternation appeared outside of this phonological environment, suggesting that perhaps their grammars contain a more generalized / productive u~o merger. This would need to be tested explicitly; as of now there is simply evidence to show that the variation in production exists.

AB regularly produced the form *puNNu* to express the target form *poNNu* 'girl'. I strongly suspect this is an analogy to the colloquial variant *uNNu* for *oNNu*[71] 'one', as well as evidence of a u~o merger.

Recall the discussion in 4.2 regarding how DS frequently produced the form *sa:ppivo:m* when I believe she intended the form *sa:ppiDum* 'it eats'. I presented two

[71] This form is technically also a colloquial form. The underlying *centamil* form is *onRu*.

hypotheses for this form, one of them being phonological. In brief, I posited that DS has a u~o merger and has replaced the *u* in *sa:ppiDum* with an *o*.

Since I did not do a phonetic analysis of DS and AB's speech, it is difficult to posit any strong claims about this u~o merger; however, I believe there is sufficient evidence to motivate future research that could investigate this possibility in more depth. This research would of course have to consider the presence (or absence) of this merger across different dialects.

5 Discussion

I conclude by summarizing and comparing the results from DS and AB (5.1), contextualizing these results in light of the trends observed in heritage languages cross-linguistically (5.2), discussing the implications of this work for heritage language pedagogy (5.3) and heritage Tamil educators (5.4), and then concluding with some final thoughts (5.5). Tamil teachers who are interested in the implications of this work for their classroom (or anyone interested in Tamil language learning and teaching in general) can skip directly to 5.4 as this section is self-contained and written specifically with a bilingual Tamil-English audience in mind.

5.1 Summary and comparison of heritage speakers' results

The results from chapter 4 clearly demonstrate that the two heritage speakers I worked with, DS and AB, both have deep knowledge of the language; their grammars are clearly natural language grammars and well-formed, proving that heritage Tamil is indeed a legitimate variety of Tamil. Both AB and DS 1) mostly produced target-like forms, and crucially, 2) any non-target-like morphology they produced was clearly a result of overgeneralizing patterns they'd observed in the speech of fluent L1 speakers. Though I deliberately didn't focus on phonology for this study, I nevertheless did notice a tendency towards phonological reduction in both speakers' speech, (again, a very natural phenomenon in language generally) specifically through the processes of lenition and vowel mergers. This may be a promising area for a future study looking specifically at heritage Tamil phonetics and phonology.

The two speakers used similar morphological strategies. With respect to PNG agreement, by and large DS and AB used target-like morphemes. Though it wasn't always clear whether the final nasal vowel was present, they generally used the target-like *-e:n* and *-a:n* forms for the 1st singular and 3rd masculine, and the target-like *-a:* for the 3rd feminine, honorific, and plural. Both speakers produced a wide range of target-like forms of the 3rd singular neuter,

which we would normally expect to be tricky because many of those forms are morphologically complex and/or phonologically irregular. Both speakers also produced some non-target forms that seemed to resemble either the 1st person plural *-o:m* or the 3rd singular neuter future *-um*. DS produced these forms exclusively in the future tense, while AB produced these largely with the non-masculine 3rd persons. I provided morphological and phonological hypotheses on why they may be producing these forms, but ultimately more detailed research specifically looking at the full range of PNG forms is required to understand this.

The most interesting similarities between DS and AB emerged with how they marked tense. As with PNG morphology, most of their forms were target-like. Where they weren't, however, what they were instead doing was lexicalizing morphophonologically unpredictable material and regularizing the conjugation of the resulting forms via the strategies of tense-stacking, compound verb constructions, and overgeneralization. The use of each of these strategies once again establishes that though their grammars are not the same as that of L1 speakers, they are nevertheless still natural and well-formed.

Tense-stacking is the strategy where the speaker lexicalizes any tense allomorphs that they find phonologically hard to predict (i.e. because they are an "exception") by incorporating them into the verb stem, and then "stacks" a tense allomorph that they find to be more salient and/or less marked on top, like so:

[Verb + target-like tense allomorph]$_{verb\ stem}$ *+ generalized tense allomorph*

DS used this the most with future tense utterances (likely because the future tense morpheme tends to be used least frequently of the three tenses), and arguably also with some present tense utterances. AB used this strategy with all three tenses. I discuss this strategy as it appears more generally in language learning in 5.2.

Compound verb constructions are built with a primary lexical verb followed by one or more auxiliaries. AB used these primarily with future tense utterances, and DS used them

mainly with past tense utterances, with both of them generally creating structures that matched those used by L1 speakers.

In addition to these strategies, AB also occasionally used the strategy of overgeneralization by replacing a target-like tense allomorph with a tense allomorph that was less marked in her mental grammar. This strategy is distinct from tense-stacking, in that the target-like tense allomorph is not included as part of the verb at all. AB only used this strategy with past tense utterances, which makes sense considering that past tense allomorphy is the most varied and least predictable. Note that is also clear evidence of learning and language knowledge; the fact that AB is overgeneralizing allomorphs in this way at all means she is noticing important patterns in the language.

Notably, the first two strategies described above appear to make the resultant form morphologically *more* complex. Tense-stacking means producing more tense morphemes than is required, and CVCs contain two verbs instead of just one. However, what DS and AB show is that maybe morphologically *large* (in terms of number of morphemes) should not necessarily be interpreted as morphologically *complex*. A simple (i.e. non-compound) verb phrase with a singular tense allomorph that is phonologically unpredictable may be morphologically small, but complex if you don't know what the target allomorph is. On the other hand, a compound verb phrase that marks tense using an auxiliary that maintains the same phonological form across all verbs is morphologically *large*, but less *complex* to learn and produce. When thinking about heritage language acquisition, it may be more useful to think about complexity from the perspective of predictable, productive, consistently reproduceable patterns that avoid "exceptions", even when the pattern may require more morphemes than a smaller, inconsistent pattern.

There seems to be some evidence that these processes of regularization observed in AB and DS's speech may be paralleling ongoing language change in L1 speakers, as evidenced in comparing some *koDuntamil* forms with their Old Tamil and/or *centamil* equivalents (note

that because of its formal nature, *centamil* is more resistant to language change than *koDuntamil*). The Old Tamil *marave:n* 'I will not forget' uses the now mostly-defunct negative indicative (in this case, phonologically null) to express negation. In modern Tamil, however, with the exception of the of the 3rd neuter which still uses a negative indicative morpheme, the negative indicative is instead expressed using the periphrastic expression (recall this is another name for a CVC), *marakka ma:TTe:n* using the negative auxiliary *ma:TTe:n* (Lehmann, 1989). Perhaps this change was motivated by the Old Tamil negative indicative often being phonologically null, and thus ambiguous, and therefore resulting in an "unpredictable" pattern. And so even though the modern periphrastic expression is longer and contains more morphemes, it is overall "simpler" because it reduces ambiguity. This parallels the CVC strategy used by DS and AB to express the past and future tenses respectively.

Tense-stacking as a strategy also seems to be a means of language change, as seen in the evolution of the *centamil* paradigm *ezhundhen, ezhukiRe:n, ezhuve:n* ('I got/woke up,' 'I am getting/waking up', 'I will get/wake up') into the modern *koDuntamil* paradigm *ezhundhen, ezhundhukkiRe:n, ezhundhuppe:n*[72] that lexicalizes the past tense *-ndh-* and "stacks" the present *-kkiR-* and future *-pp-* suffixes on top for the present and future forms. This process appears to be identical to the tense-stacking strategy used by DS with the present and future tenses, and AB with all three tenses; thus, it is possible that the patterns seen in their grammars are representative of, or at least related to, ongoing language changes in progress.

Finally, Schiffman et al. (2009) notes that many verbs that belonged to certain verb classes in *centamil* have over the years, in *koDuntamil*, shifted to being part of different classes. Indeed, a comparative study of language change in Tamil verbal morphology and verbal morphology in heritage Tamil speakers could be a very interesting area of future research.

[72] எழுந்தேன், எழுகிறேன், எழுவேன் > எழுந்தேன், எழுந்துக்கிறேன், எழுந்துப்பேன்

5.2 Results in the context of cross-linguistic heritage language grammars

Overall, DS and AB's approach to verbal morphology can be described using two principles: they lexicalize irregular/unpredictable morphophonological patterns, and they strongly prefer regular, perceptually salient patterns. The combination of these two principles is most apparent with the tense-stacking strategy, but the desire for regular patterns is also evident in the other strategies they used to mark tense, such as creating CVCs or overgeneralizing unmarked morphemes. These preferences are attested among heritage speakers cross-linguistically, and can be summed up as "[reducing] paradigmatic irregularities" (Putnam et al., 2021). Heritage speakers show a strong preference for one-to-one meaning mappings and avoiding ambiguity. The strategies they use to achieve this are largely influenced by perceptual salience and frequency (which, crucially, are not the same thing [Polinsky, 2018]). In this section I discuss how the conclusions from this study compare with what has been observed of the morphology of other heritage languages and their speakers, beginning with tense-stacking.

The phenomenon that I termed tense-stacking in this study, when applied more broadly to a language's entire morphology (rather than just tense), is referred to in the literature as "overmarking" (Montrul, 2022; Putnam et al., 2021). Overmarking is itself a type of overregularization, which is when speakers simplify a paradigm by "regularizing" it, thus creating forms a fluent speaker would consider non-target-like. Overmarking goes one step further, by incorporating the "irregular" morphemes into the word, but also including the "regular" form as well. An example in English would be the form "childrens" in the sentence "the childrens are playing", which applies the regular/unmarked plural morpheme '-s' on top of the "irregular" plural '-en'. Overmarking is attested across heritage languages (Montrul, 2022; Putnam et al., 2021). Putnam et al. (2021) cite Tóth's 2007 study on heritage Hungarian, which observed overmarking specifically in the context of verb classes, similar to what was seen in this study for Tamil. Polinsky (2018) also notes several studies that describe

heritage speakers using overmarking. Specifically, she cites Choi's 2003 study on heritage Korean, where the author noted that speakers would reanalyse irregular verbal morphology (in this case, denoting register) as part of the stem, just as the Tamil speakers in this study did with tense.

Polinsky (2018) describes how it is a combination of perceptually salient (which she defines as "perceptually noticeable…phonetically heavier") and high frequency elements that get overgeneralized in heritage speech, and that it is as yet unclear which has greater influence in any given context. She also points out that the majority of these studies are *production* studies; if it emerged that heritage speakers not only produced overmarking, but were also more accepting of *hearing* overmarking, it would indicate a change in their grammar itself, rather than just being a product of their preference for perceptual salience, avoiding ambiguity, and high frequency. Ultimately, more comprehension studies are required to assess this.

It is observed that heritage speakers of agglutinative languages, in particular, have a preference for analytical and periphrastic structures, which seems to relate to their preference for one-to-one meaning mappings (Montrul, 2022; Putnam et al., 2021). This preference may likely be one of the motivators behind the CVC strategy that both DS and AB used. (Recall that another term used for CVCs in Tamil is "periphrastic expression"). It's possible that DS and AB prefer using CVCs, even though they require more morphemes than a simple, non-compound verb phrase, because a CVC has a clearer one-to-one meaning mapping and thus reduces ambiguity. For example, though the CVC in (141) has more morphemes than its simple, non-compound equivalent, the marker of the future tense in the simple phrase is a bit less clear as it is fused with the PNG agreement marker, compared to the clear demarcation of the future construction in the CVC (using the *po* auxiliary).

(141)	Simple, non-compound	CVC
Transc.ta	மழை பெய்யும்	மழை பெய்யப்போகிறது
Trans.la	mazhai peyy**um**	mazhai peyya**po:**giRadhu
MB	mazhai peyy-**um**	mazhai peyy-a-**po:**-giR-adhu
UR	mazhai peyy-**um**	mazhai peyy-a-**po:**-kiR-adhu
MG	rain rain-**3N.FUT**	rain rain-INF-**go**-PRS-3N
Transl.	'it's going to rain'	

There is significant evidence of the desire to reduce ambiguity in comprehension experiments as well. Polinsky (2018) notes several examples where heritage speakers reanalyse a single-valued, privative contrast into a multi-valued, equipollent contrast. For instance, she describes how with L1 speakers of Slavic languages, telicity is encoded with a privative contrast (where the perfective is specified for "completed" events but the imperfective is not specified and so can refer to completed or incomplete events), but heritage speakers instead interpret it as an equipollent contrast (where the perfective is specified for completed events and the imperfective is specified for incomplete). I noticed something that may have been similar occurring with DS and AB in the picture matching task I attempted that looked at the perfective morpheme *viDu*. As discussed in 3.4, I ultimately decided not to include the results of this task in the analysis because of compounding limitations, but this cross-linguistic evidence suggests that it would be worthwhile to redo the task (after addressing its limitations), to see if heritage Tamil speakers also follow this pattern.

The speakers I worked with showed a lot of variation in the patterns they used; either this is stable variation, or instead this indicates a changing grammar and is evidence that a heritage language grammar does not stop evolving after puberty. DS and AB would often use different strategies to express the same meanings, or use different morphological strategies

in the same context. Many of their target-like utterances involved complex morphological structures that they would then avoid producing in other situations. It is highly likely that many of their utterances were memorized "chunks". On the flip side, their consistent use of overmarking shows a profound understanding of morphological patterns in the language. This combination of memorized, target-like utterances and pattern-based, non-target-like utterances is strongly reminiscent of the well-documented "U-shape" learning curve seen in L1 acquisition. It has been well described in L1 acquisition how children will often start off producing many target-like forms that are simply memorized input, but that as they start to learn more of the patterns of the language, they overgeneralize some of these and so their accuracy reduces. Eventually they learn when to apply which patterns, and so the percentage of target-forms increases again, thus creating a "U-shape" over time to represent the percentage of target-like utterances. I wonder if there is something similar happening with heritage speakers but over a longer time frame, where a lot of their target-like forms are memorized chunks, while their non-target forms are evidence of recognizing patterns in the language and generalizing them. Possibly the distribution of which forms are the result of recognized patterns and which are memorized follows the Tolerance Principle (Yang, 2016), i.e. it is optimized based on how productive consistently following a given pattern is vs how productive simply memorizing chunks is. While it is of course possible that the variation in the forms produced and the strategies used by DS and AB is just stable variation that is being influenced by other sociolinguistic or linguistic factors, it also seems possible that the process of heritage language acquisition continues well into adolescence. It would be interesting to do a long-term study to see if and when heritage speakers' grammars stabilize, and how. In the absence of this, conducting a similar study but with older speakers would also be a valuable comparison.

I feel it important to reaffirm that the results of this study are not meant to espouse a "deficit approach" to describing a heritage language grammar, or perpetuate the

monolingual bias that is prevalent in the field of second language acquisition (Ortega, 2019). Indeed, the strategies used by DS and AB are clear indicators of language learning and knowledge. They both are evidently cognizant of grammatical patterns in the language, and it is because of their familiarity with these patterns that they are able to produce the forms that they do, whether they be target-like or not. They recognize "exceptions" to these patterns and come up with strategies to regularize them, demonstrating that (of course), heritage grammars are natural, well-formed grammars in their own right. I end this section with a quote from Putnam et al. (2021), who reiterate, "An important point that cannot be overstated is that the restructuring of heritage morphology does not equate with language loss. Rather, we observe the establishment of new, innovative patterns."

5.3 Implications for heritage language pedagogy

Once we noticed the pattern of tense-stacking in DS and AB, my mother and I couldn't stop seeing the same pattern in the speech of our students. Not only that, we would both notice overmarking in my own speech all the time. In one instance, I tried to say that a room felt very cold and produced the utterance *'chilliNDa: irukku'*, where *chilliNDa:* was intended to mean "cold". My parents pointed out that the correct form would be *'chillin.'*[73] Upon reflection, I realized that my form was a result of overmarking: *-a:* is the unmarked adverb morpheme, seen in words like *melisa:* 'thin', *su:Da:* 'hot', *kuLLamma:* 'short' or even code-mixed loans like 'cute-*a:'*. I had noticed this pattern and mistakenly thought that every adverb needed to carry this morpheme, and so "stacked" it on top of the "irregular" adverb marker *-in* that already appeared on the stem. [74]

The above anecdotal evidence highlighted to me that overmarking isn't just something that happens with tense morphology, or indeed just with verbal morphology – it can occur

[73] I'm fairly certain this is a nativized loan of the English word "chill".
[74] The change of *n* to *ND* in my form is a misapplied morphophonological pattern (that I don't go into here) that was triggered by the *r* in the following word *irukku*.

anywhere in the grammar of a language.[75] Thus, heritage language educators may find it valuable to anticipate areas where students may be overgeneralizing and/or overmarking by assessing what parts of their language's grammar have irregular grammatical paradigms, and systematically observing how students produce these forms. It is important to observe the full range of the paradigm; in our class, we tended to largely speak in the present tense, where tense-stacking is less evident. Only by deliberately eliciting past and future tense utterances as part of this study did the pattern of tense-stacking become clear to us.

The next section summarizes the content of this section and distils it into specific, actionable advice for heritage Tamil educators on how the results of this study can be applied in the classroom. As the intended audience are Tamil-English bilinguals, I use Tamil script and vocabulary and leave it untranslated for concision. Non-Tamil readers are advised to skip ahead to 5.5.

5.4 Takeaways for heritage Tamil educators

This section provides some suggestions on how the results of this study can be applied in your classroom.

Examine how your students construct basic verbs. Specifically, get them to say sentences in *இறந்தக்காலம்*, *நிகழ்காலம்*, and *எதிர்காலம்*. Get them to say sentences that agree with all the different possible grammatical subjects (i.e. *அவள்*, *அவன்*, *நாங்கள்*, *அவர்கள்*), not just *நான்*, *நீ*, and *நீங்கள்*. Students may struggle more with *எதிர்காலம்* and less frequently used subjects like *அவர்கள்*, so practice these more. The memory boards in Appendix A are one way of practicing these in a game format.

The biggest struggle as a heritage Tamil teacher is figuring out what your students already know, and what your students might struggle with. Here are some things that might be challenges:

[75] In fact, Putnam et al. (2021) argue that nominal morphology tends to be more affected than verbal morphology among heritage languages.

- Anything without a clear English equivalent. E.g. consider how the *நான்* in *நான் பாடுகிறேன்* and the *எனக்கு* in *எனக்கு வாழைப்பழம் பிடிக்கும்* both translate to "I" in English. Students are often confused about where to use *எனக்கு*, so try and reinforce this with them.
- Anything that doesn't have a clear, obvious pattern. E.g. consider how *இறந்தக்காலம்* is marked with -*ந்த-* in *விழுந்தேன்*, -*இன்-* in *பாடினேன்*, and *த்த* in *கொடுத்தேன்*, and knowing which one to use just depends on the verb. Our students often think the sounds that form the *இறந்தக்காலம்* are part of the main verb itself because these sounds don't "pattern" nicely. They'll then stack -*இன்-* on top to *explicitly* mark the *இறந்தக்காலம்*. For example they will say something like *கேட்டினேன்* instead of *கேட்டேன்*. Another area this might happen is with adjective marking, where they may feel like adding -*ஆ* after every adjective to match the pattern in words like *நல்லா* and *சூடா*. Do take the time to think of other parts of the grammar that have irregular patterns.
- Don't underestimate how challenging it is to make the connection between how a word sounds in spoken Tamil vs written Tamil. A lot of the grammatical patterns in Tamil are way more transparent in written Tamil than they are in spoken, but if your students are mostly only exposed to spoken Tamil, it's definitely worth it to take the time to explicitly teach the connection between the two forms.

In general, whenever you notice your students making "mistakes", it's probably because they've actually noticed a pattern in the language but are just applying it differently than we might normally expect. Once you figure out what this pattern they're using is and why, then it becomes much easier to meet them where they're at, which makes it easier to teach them in a way they'll understand and retain.

Finally, of course it goes without saying that there's a huge range of *வட்டார வழக்குகள்*, so some things a student says or some "different-sounding" pronunciations might just be what they are used to their parents saying at home, which of course doesn't make it wrong or something to be "corrected". Encouraging students to use their native *வட்டார வழக்கு* will help them recognize grammatical patterns on their own.

5.5 Closing thoughts

AS, the baseline speaker who is DS's mother, notes that DS has *கேள்வி ஞானம்*[76] : roughly this translates to "heard knowledge" or "knowledge gained from hearing and listening". The full import of this expression is somewhat lost in translation, so I want to emphasize that the value of this type of knowledge cannot be understated. So many of my teachers growing up, from music teachers to language teachers, would always impress upon me that there is only so much that can be explicitly taught – the foundation of any aural/oral skill is built on *கேள்வி ஞானம்*. As such I think it is important to acknowledge, and celebrate, the *கேள்வி ஞானம்* that heritage speakers like DS and AB possess. I believe that any heritage speaker who has been or is exposed to Tamil in their life, whether it's from hearing it consistently at home, talking on the phone to their relatives back in the homeland, watching Tamil movies, or even just hearing a few phrases here and there that their parents might have used when they were growing up, possesses some amount of *கேள்வி ஞானம்*. More importantly, anyone coming back to their heritage language, at any age, has the ability to acquire *கேள்வி ஞானம்*. I hope the findings from this study go some way to supporting learners in that journey.

[76] *ke:Lvi nja:nam*

References

Annamalai, E. (2011a). *Social dimensions of modern Tamil* (1st ed.). Cre-A:

Annamalai, E. (2011b). Social variation. In *Social dimensions of modern Tamil* (pp. 63–69). Cre-A:

Annamalai, E. (2017). Grammar Aid. In *சிகாகோ தமிழ் (Chicago Tamil) A course on the Tamil language* (p. section 14). University of Chicago. https://tamilcourse.uchicago.edu/node/122

Annamalai, E. (2019). Modern Tamil. In S. B. Steever (Ed.), *The Dravidian languages* (2nd ed.). Routledge.

Annamalai, E., & Asher, R. E. (2002). *Colloquial Tamil: The Complete Course for Beginners*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/uvic/detail.action?docID=2194949

Arden, A. H. (1942). *A progressive grammar of common Tamil* (A. C. Clayton, Ed.; 5th ed.). Christian Literature Society.

Arden, A. H. (with University of California Libraries). (1910). *A progressive grammar of common Tamil*. Madras : Society for Promoting Christian Knowledge. http://archive.org/details/progressivegramm00ardeiala

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Defining an "ideal" heritage speaker: Theoretical and methodological challenges. Reply to peer commentaries. *Theoretical Linguistics*, *39*, 259–294.

Berman, R. (2004). Between emergence and mastery: The long developmental route of language acquisition. In R. Berman (Ed.), *Language Development across Childhood and Adolescence* (pp. 9–34). John Benjamins.

Boon, E., & Polinsky, M. (2015). From Silence to Voice: Empowering Heritage Language Speakers in the 21st Century. *Informes Del Observatorio / Observatorio Reports*. https://doi.org/10.15427/OR007-01/2015EN

Bright, W. (1990a). *Language variation in South Asia*. Oxford University Press.

Bright, W. (1990b). Preface. In *Language variation in South Asia* (pp. vii–xi). Oxford University Press.

Bright, W., & Lindenfeld, J. (1990). Complex verb forms in colloquial Tamil. In *Language Variation in South Asia* (pp. 44–63). Oxford University Press.

Burton, S., & Matthewson, L. (2015). Targeted Construction Storyboards in Semantic Fieldwork. In L. Matthewson & M. R. Bochnak (Eds.), *Methodologies in Semantic Fieldwork*. Oxford Scholarship Online. 10.1093/acprof:oso/9780190212339.001.0001

Choi, H.-W. (2003). Paradigm leveling in American Korean. *Language Research*, *39*(1), 183–204.

Correa, M. (2011). Advocating for critical pedagogical approaches to teaching Spanish as a heritage language: Some considerations. *Foreign Language Annals*, *44*(2), 308–320. https://doi.org/10.1111/j.1944-9720.2011.01132.x

Cummins, J., & Swain, M. (2014). *Bilingualism in Education: Aspects of theory, research and practice*. Routledge. https://doi.org/10.4324/9781315835877

Dominguez, L., & Arche, M. J. (2021). The 'comparative logic' and why we need to explain interlanguage grammars. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.717635

Ducar, C. M. (2008). Student Voices: The Missing Link in the Spanish Heritage Language Debate. *Foreign Language Annals*, *41*(3), 415–433. https://doi.org/10.1111/j.1944-9720.2008.tb03305.x

Duff, P. A., Liu, Y., & Li, D. (2017). Chinese Heritage Language Learning: Negotiating Identities, Ideologies, and Institutionalization. In *The Routledge Handbook of Heritage Language Education*. Routledge.

Eisenbeiss, S. (2010). Production methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 11–34). John Benjamins Publishing Company.

Ellis, R., & Shintani, N. (2013). *Exploring Language Pedagogy through Second Language Acquisition Research* (1st ed.). Routledge.

Gardner, S. E. B. (2012). The four R's of leadership in Indigenous language revitalization. In C. Kenny (Ed.), *Living Indigenous leadership: Native narratives on building strong communities* (1st ed.). UBC Press.

Gnanaseelan, J. (2022). Early childhood language development: A case study of a Sri Lankan Tamil child's progress in language acquisition. *Sri Lanka Journal of Social Sciences*, *45*(2), 213–229. https://dx.doi.org/10.4038/sljss.v45i2.8300

González, C. (2002). Phonetic variation in voiced obstruents in North-Central Peninsular Spanish. *Journal of the International Phonetic Association*, *32*, 17–31. https://doi.org/10.1017/S0025100302000129

Jee, M. J. (2016). Exploring Korean heritage language learners' anxiety: 'We are not afraid of Korean!' *Journal of Multilingual and Multicultural Development*, *37*(1), 56–74. https://doi.org/10.1080/01434632.2015.1029933

Jee, M. J. (2020). Heritage language anxiety and major language anxiety experienced by Korean immigrants in Australia. *International Journal of Bilingual Education and Bilingualism*, *25*(5), 1–17. https://doi.org/10.1080/13670050.2020.1799321

Krishnamurti, B. (Ed.). (2003). The verb. In *The Dravidian Languages* (pp. 277–387). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511486876.009

Lakshmanan, U. (2000). The acquisition of relative clauses by Tamil children. *Journal of Child Language*, *27*(3), 587–617. Cambridge Core. https://doi.org/10.1017/S0305000900004293

Lakshmanan, U. (2006). Assessing Linguistic Competence: Verbal Inflection in Child Tamil. *Language Assessment Quarterly*, *3*(2), 171–205. https://doi.org/10.1207/s15434311laq0302_5

Lakshmi Bai, B. (2000). Sounds and words in early language acquisition: A bilingual account. In *Sounds and words in early language acquisition: A bilingual account*. Indian Institute of Advanced Study. https://search.library.uvic.ca/permalink/01VIC_INST/71bv5b/alma9913405643807291

Lakshmi Bai, B. (2004). Acquisition of dative subject in Tamil. In P. Bhaskararao & K. V. Subbarao (Eds.), *Non-nominative ubjects* (Vol. 1). John Benjamins Publishing Company.

Lardiere, D. (2005). *On Morphological Competence* (L. Dekydtspotter, Ed.; pp. 178–192). Cascadilla Proceedings Project.

Lehmann, T. (1989). *A grammar of modern Tamil*. Pondicherry Institute of Linguistics and Culture.

Mayberry, R. (2007). When timing is everything: Age of first-language acquisition effects on second-language learning. *Applied Psycholinguistics*, *28*, 537–549. https://doi.org/10.1017/S0142716407070294

Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9781139030502

Montrul, S. (Ed.). (2022). Structural Changes in Heritage Language Grammars. In *Native Speakers, Interrupted: Differential Object Marking and Language Change in Heritage Languages* (pp. 39–81). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781316459690.004

Montrul, S., & Polinsky, M. (Eds.). (2021). *The Cambridge Handbook of Heritage Languages and Linguistics*. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781108766340

Montrul, S., & Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition*, *20*, 109–132.

Nijhawan, S. (2011). "I Got the Point Across and That is What Counts". Transcultural Versus (?) Linguistic Competence in Language Teaching. *Journal of the National Council of Less Commonly Taught Languages*, *9*(1), 59–81.

Nijhawan, S. (2017). The Hindi-Urdu Heritage Language Stream: Institutional and Pedagogical Challenges. In *The Routledge Handbook of Heritage Language Education*. Routledge.

Ortega, L. (2019). SLA and the study of equitable multilingualism. *The Modern Language Journal*, *103*, 23–38.

Paramasivam, K. (1979). *Effectivity and causativity in Tamil*. Dravidian Linguistics Association.
Patience, M. (2022). *Articulatory Difficulty in L2 Spanish*. University of Toronto.
Polinsky, M. (2008). Heritage language narratives. In D. M. Brinton, O. Kagan, & S. Bauckus (Eds.), *Heritage language education: A new field emerging* (pp. 149–164). Routledge.
Polinsky, M. (2011). Reanalysis in adult heritage language: A case for attrition. *Studies in Second Language Acquisition*, *33*, 305–328.
Polinsky, M. (Ed.). (2018a). Morphology and Morphosyntax. In *Heritage Languages and their Speakers* (pp. 164–221). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781107252349.006
Polinsky, M. (Ed.). (2018b). Who Are These Speakers, Where Do They Come From, and How Did They Get to Be the Way They Are? In *Heritage Languages and their Speakers* (pp. 1–37). Cambridge University Press. https://doi.org/10.1017/9781107252349.002
Polinsky, M., & Kagan, O. (2007). Heritage Languages: In the 'Wild' and in the Classroom. *Language and Linguistics Compass*, *1*(5), 368–395. https://doi.org/10.1111/j.1749-818X.2007.00022.x
Putnam, M. T., Schwarz, L., & Hoffman, A. D. (2021). Morphology of heritage languages. In S. Montrul & M. Polinsky (Eds.), *The Cambridge Handbook of Heritage Languages and Linguistics* (pp. 613–643). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781108766340.028
Raghavendra, P., & Leonard, L. B. (1989). The acquisition of agglutinating languages: Converging evidence from Tamil. *Journal of Child Language*, *16*(2), 313–322. Cambridge Core. https://doi.org/10.1017/S0305000900010436
Renganathan, V. (2015). *Heritage language environment (HLE) and its implications upon language curriculum: A case of Tamil heritage students and designing a differentiated curriculum*. Indian Languages in Diasporas: Retention and Transmission, University of Pennsylvania, USA.
Sanchez-Munoz, A. (2016). Heritage language healing? Learners' attitudes and damage control in a heritage language classroom. In *Studies in Bilingualism Advances in Spanish as a Heritage Language* (pp. 205–218). https://doi.org/10.1075/sibil.49.11san
Sarma, V. M. (2013). Issues in the acquisition of Tamil verb morphology. In *South and Southeast Asian Psycholinguistics* (pp. 110–123). Cambridge University Press.
Schiffman, H. F., Renganathan, V., Christdas, P., Palanisamy, S., Baba, T., Karunakaran, K., Krishnamurti, G., Pradeep, C. M., & Radhakrishnan, S. (Eds.). (2009). *An English dictionary of the Tamil verb*. South Asia Studies Dept. University of Pennsylvania. https://www.tamilverb.com/
Sevinç, Y., & Backus, A. (2019). Anxiety, language use and linguistic competence in an immigrant context: A vicious circle? *International Journal of Bilingual Education and Bilingualism*, *22*(6), 706–724. https://doi.org/10.1080/13670050.2017.1306021
Slobin, D. I., & Berman, R. A. (Eds.). (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Psychology Press. https://doi.org/10.4324/9780203773512
Subrahmanyam, P. S. (1971). *Dravidian Verb Morphology*.
Taku, N. (2020, October 27). India's engineers have thrived in Silicon Valley. So has its caste system. *The Washington Post*. https://www.washingtonpost.com/technology/2020/10/27/indian-caste-bias-silicon-valley/
Tóth, G. (2007). *Linguistic Interference and First-Language Attrition. German and Hungarian in the San Francisco Bay Area.* Peter Lang Verlag. https://www.peterlang.com/document/1103927
Vaidyanathan, R. (1988). Development of forms and functions of interrogatives in children: A longitudinal study in Tamil. *Journal of Child Language*, *15*(3), 533–549. Cambridge Core. https://doi.org/10.1017/S0305000900012551
Vaidyanathan, R. (1991). Development of forms and functions of negation in the early stages of language acquisition: A study in Tamil. *Journal of Child Language*, *18*(1), 51–66. https://doi.org/10.1017/S0305000900013295

Valdés, G. (2000). *The teaching of heritage languages: An introduction for Slavic teaching professionals* (O. Kagan & B. Rifkin, Eds.). Slavica.

Yang, C. (2016). *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. The MIT Press.

Appendix A: Memory game boards

Figure 3: PNG agreement memory board

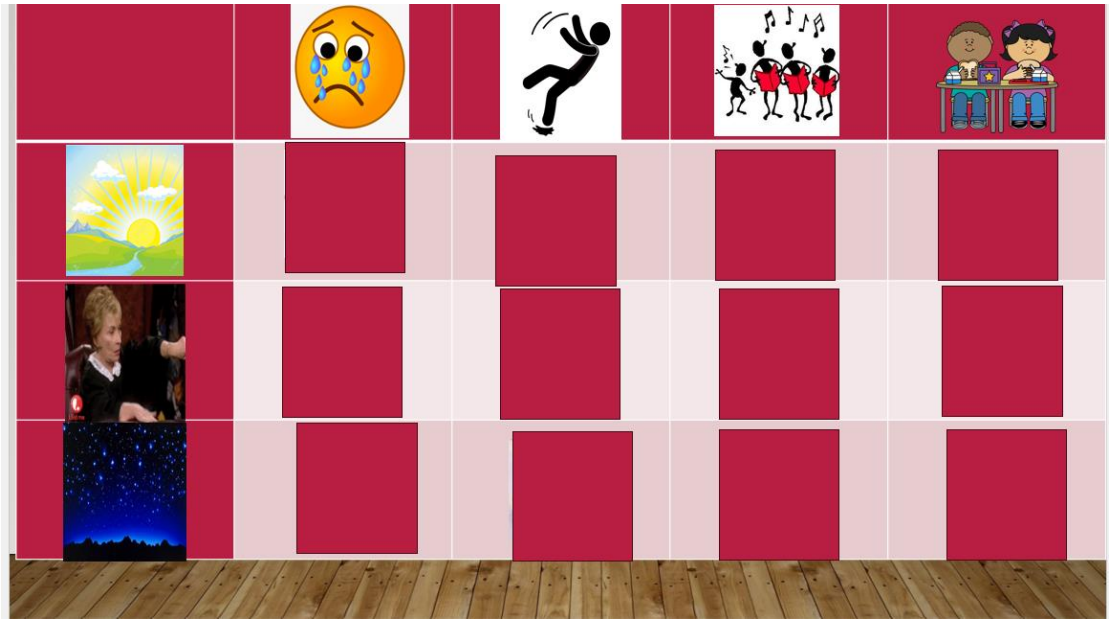

Figure 4: Tense memory board (1)

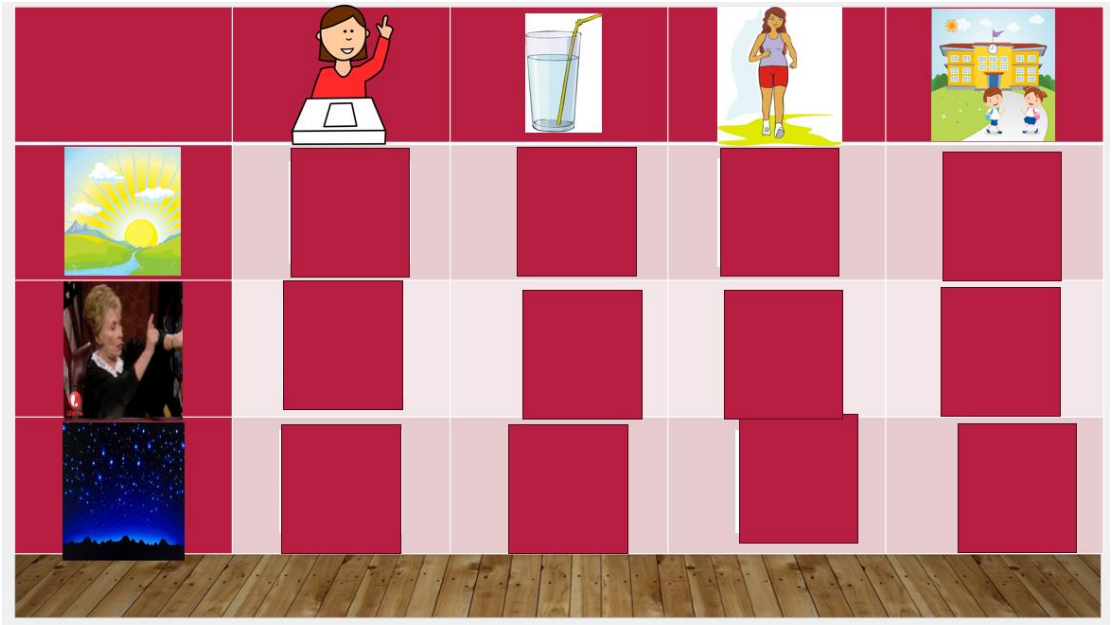

Figure 5: Tense memory board (2)

Appendix B: Tamil – IPA – Romanization Key

IPA			a	a:	ɪ	i	u	u:	e	e:	a͡i	o	o:	o͡w
	English equivalent*		a	A or a:	i	I or i:	u	U or u:	e	E or e:	ai	o	O or o:	ow
		ஃ	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
k	k (g)	க்	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கௌ
ŋ	ng	ங்	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙௌ
t͡ʃ	ch (j)	ச்	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சௌ
ɲ	nj	ஞ்	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞௌ
ʈ	T (D)	ட்	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டௌ
ɳ	N	ண்	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணௌ
θ	th (dh)	த்	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தௌ
n̪	n**	ந்	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நௌ
p	p (b)	ப்	ப	பா	பி	பீ	பு	பூ	பெ	பே	பை	பொ	போ	பௌ
m	m	ம்	ம	மா	மி	மீ	மு	மூ	மெ	மே	மை	மொ	மோ	மௌ
j	y	ய்	ய	யா	யி	யீ	யு	யூ	யெ	யே	யை	யொ	யோ	யௌ
ɹ	r	ர்	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ	ரௌ
l	l	ல்	ல	லா	லி	லீ	லு	லூ	லெ	லே	லை	லொ	லோ	லௌ
v	v	வ்	வ	வா	வி	வீ	வு	வூ	வெ	வே	வை	வொ	வோ	வௌ
ɻ	zh	ழ்	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழௌ
ɭ	L	ள்	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளை	ளொ	ளோ	ளௌ
r	R	ற்	ற	றா	றி	றீ	று	றூ	றெ	றே	றை	றொ	றோ	றௌ
n	n	ன்	ன	னா	னி	னீ	னு	னூ	னெ	னே	னை	னொ	னோ	னௌ

Table 22: Tamil-IPA-Romanization key

*Letters in parentheses are voiced allophones; voicing in Tamil is conditioned by the environment (i.e. there is no phonological voicing distinction) and the underlying form is understood to be the voiceless allophone

** ந is a dental nasal and used to be phonologically distinct from ன, an alveolar nasal but the two have merged in modern spoken Tamil. The distinction is only maintained in written Tamil.

Appendix C: Compound verb forms – aspectual auxiliaries

As described in 2.2.4, all aspectual auxiliaries must attach to a verbal participle, the form of which is determined by which verb class it belongs to. Adapted from Krishnamurti (2003) and Lehmann (1989).

Auxiliary stem(s)	Interpretation	Verb Class
***iru* (be)**	Perfect	class VII
***koL* (hold) + *iru* (be)**	Progressive	The verbal participle of *koL* is *koNDu*, and *iru* is class VII.
***viDu* (leave)**	Perfective (thoroughly, definitely; completed event; sometimes implies the event was unexpected)	class IV
***Aku* (become)**	Completion of "affective" verbs	class III

Table 23: Aspectual auxiliaries

Appendix D: Compound verb forms – modal auxiliaries

As described in 2.2.4, all modal auxiliaries must attach to an infinitive. Adapted from Krishnamurti (2003) and Lehmann (1989).

Auxiliary stem(s)	Interpretation	Notes
vENDum ***vENDAm***	Must/want/need Don't have to/don't want/don't need	
kUDAdhu	Shouldn't do/mustn't do	The positive version *kUDam* isn't really used in colloquial speech to my knowledge. In my dialect, I would instead use the hortative.
muDiyum ***muDiyAdhu***	Can do (ability) Can't do (ability)	
***po* (go)**	Intention to do something/prediction (similar to a purposive)	There are subtle differences in how these are interpreted that condition which is most appropriate to use in a given situation, but to my understanding they're largely pragmatic.
***va* (come)**	Intention to do something/prediction (similar to a purposive)	
***iru* (be)**	Intention to perform action (literally "in the state of doing the action")	
***paar* (see)**	Attemptive (trying to do something)	

Table 24: Modal auxiliaries