6-2014

Windesi Wamesa Morphophonology

Emily A. Gasser
Swarthmore College, egasser1@swarthmore.edu

Windesi Wamesa Morphophonology

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Emily Anne Gasser

Dissertation Director: Dr. Claire Bowern

December 2014

Abstract

Windesi Wamesa Morphophonology

Emily Anne Gasser

2015

Wamesa [WAD] is an endangered Austronesian language spoken in the south-eastern Bird's Head of New Guinea, in the Indonesian province of West Papua. This dissertation provides a description and formal analysis of the phonology and morphology of the Windesi dialect based on the author's fieldwork with speakers of the language.

Chapter 1 provides an introduction to the language, its speakers, and the cultural, geographic, and linguistic context in which Wamesa is spoken. It also provides background on the fieldwork which forms the basis of this dissertation and the resulting corpus. Chapter 2 describes the phonology of Wamesa, including its phoneme inventory, phonotactics, and productive phonological processes, with phonetic detail. The second half of the chapter gives an account of the phonological adaptation of loan words into Wamesa. Chapter 3 gives a formal analysis of stress assignment in the language based in Optimality Theory. Chapter 4 describes the Wamesa clitics and affixes, and Chapter 5 gives an account of the three major word classes, nouns, verbs, and adjectives, as well as modes of spatial expression and a selection of other minor word classes. Chapter 6 gives a formal synchronic analysis of the infixation of verbal subject agreement affixes in Wamesa, followed by a diachronic account of how the pattern might have arisen from incremental improvements in speech production and perception.

This dissertation provides the first in-depth description of the grammar of Windesi Wamesa, as well as the first formal analysis of its structures. The data presented here will be of interest for typological and historical studies of Austronesian, particularly the understudied South Halmahera-West New Guinea subgroup to which Wamesa belongs.

In addition to enriching our understanding of this family, the dissertation presents data and analyses which will be of interest for morphological and phonological theory more narrowly.

Contents

List of Figures

List of Abbreviations

1	1st person
2	2nd person
3	3rd person
APPL	applicative
CAUS	causative
CB	Cenderawasih Bay
COMPL	completive
D	Dutch
DET	determiner
DU	dual
DUR	durative
E	English
ESS	essive
EXCL	exclusive
HUM	human
INCL	inclusive
INDEF	indefinite
INTENS	intensifier
LOC	locative
M	Malay
NEG	negative/negation
NH	non-human
PA	*pa* aspectual marker
PBY	Proto-Biak-Yapen
PCEMP	Proto-Central Eastern Malayo-Polynesian

PL	plural
PM	Papuan Malay
POSS	possessum
Q	interrogative
REDUP	reduplicant
SG	singular
SHWNG	South Halmahera-West New Guinea
TOP	topic
WAD	Wamesa
WNG	West New Guinea
WO	*wo* discourse particle

Acknowledgments

There are many people without whom this dissertation would not have been possible, and to whom I am deeply grateful. First and foremost I want to thank all of the Wamesa speakers who so generously shared their language with me: Ibu Marice Karubuy, who gave up her afternoons nearly every day while I was in Manokwari to sit with me in front of a recorder, answer all my silly questions, translate sentences, tell stories, drink tea, and laugh; Bapak Aukila and Ibu Karolina, who hosted me in Kampung Windesi and showed me their village and landscape; Bapak Marten Windesi and Tante Lorensina Biambara, who met with me in Manokwari; and all of the other speakers who contributed their time and their voices to this project.

I never would have gotten to Papua in the first place, much less been able to stay to carry out this research, were it not for the institutional support of Universitas Negeri Papua and the Center for Endangered Languages Documentation in Manokwari. The CELD staff jumped through all sorts of hoops to get me my visas and travel permits, gave me a desk to work at, connected me with housing and Wamesa speakers, shared their expertise on related languages, and showed me the best places to go for lunch near campus or an afternoon on the beach. Particular thanks to Bapak Rektor Suriel Mofu, Yusuf Sawaki, Eny Arilaha, Sutri, Manu, Kris, Willem Burung, and my student assistant Cika Tethool. My *keluarga kos* Ibu Mince, Om Jemmy, Randy, Reni, Virgin, Jhio, Rifqy, Chen, and Juen, who I lived with for my seven months in the field, deserve mention here, as do my fellow researchers in Manokwari, Freya Morigerowsky, Sonja Reisberg, Laura Arnold, and Fanny Cottet, all of whom contributed both to the research itself and to making my time in the field as enjoyable as it was. Thank you also to David Gil, who put Papua on my radar when I was looking for a field site and who suggested Wamesa as a potential language for me to study, and to David Kamholz, René van den Berg, and Theo Henning who shared valuable data and resources with me.

At Yale, I have had the privilege to work with Claire Bowern for four years as my advisor; without her support, guidance, and copious red pen this dissertation would look very different than it does. The same is true of my other committee members, Steve Anderson, Ryan Bennett, and Mary Dalrymple, for whose feedback and encouragement I am extremely grateful. On an institutional level, many thanks are due to Yale University for funding my graduate studies, to the Yale Council on Southeast Asia Studies, who provided financial support for two preliminary trips to Indonesia, and the National Science Foundation grant 1153795, which funded the bulk of my fieldwork. And finally, my endless gratitude to all of those professors, post-docs, colleagues, family and friends who pointed me towards relevant literature, helped me work through analyses, listened to me gripe, cheered me on, and occasionally provided a much-needed motivational cookie, hug, or kick in the pants: my cohort, Jason Zentz and Sabina Matyiku; Ray, Maria, Sarah, and Linda Gasser; Jen Runds, Sean Gleason, Mike Freedman, Nicole Palffy-Muhoray, E-Ching Ng, Jim Wood, Raffaella Zanuttini, Indriyo Sukmono, Andy Carruthers, Kevin Fogg, Nina Thanawala, Jaime Rosegren, and Nikhil Navkal, among many others. Thank you.

So ine maneta pasia, kawasa Wamesa.

Chapter 1

Introduction

Wamesa [WAD] is an Austronesian language of West Papua, Indonesia. Wamesa is most often referred to in the literature as Wandamen; it also appears, particularly in older sources, as Windesi, Windessi, Wendesi, Bintuni, Bentoeni, Wondama, and various combinations of these terms. The speakers with whom I have worked prefer to use the name 'Wamesa' to refer to the language as a whole and 'Wandamen' or 'Wondama' to refer specifically to the dialect spoken around Wandamen Bay (*Teluk Wondama*); I defer to them and here adopt their nomenclature. Speakers refer to their language as *kavio Wamesa* 'Wamesa language' or, in Malay, as *bahasa Wamesa*.

There are two competing folk etymologies claimed for the language name. The first recalls a traditional story of the *perang saudara*, War of Siblings, during which much blood was shed and which according to tradition caused the Wamesa people to disperse from their original homeland in the mountainous interior and spread across the current language area as far as the coasts of Cenderawasih and Bintuni Bays. Under this account, *wamesa* is supposed to derive from the Kuri words *wam* 'blood' and *sa*, which was translated for me as either 'family' or 'sheet', depending on the storyteller. The Kuri wordlist I recorded in Bintuni confirms *wams* as the word for 'blood' in that language; 'family' and ‘sheet’ were not elicited. An alternate folk etymology is that the language is named after a species of blowfish, known in Indonesian as *ikan buntal* or *ikan porobibi* and in Wamesa as *dia wamesa*, supposedly because, just as the blowfish has a large belly, the Wamesa cultural identity is likewise broad, encompassing many groups in the southeastern Bird’s Head and Yapen Island.

This dissertation, based on data collected during my fieldwork on the Windesi dialect of Wamesa, gives an overview of the phonology and morphology of the language and the interface between the two. In addition to a detailed description of the facts, some facets of the grammar are also investigated from a theoretical standpoint. I have chosen to use Optimality Theory (Prince & Smolensky 1993/2004) as the framework in which to analyze the Wamesa data, with an eye towards both discovering the best available analysis of the Wamesa patterns and also how this language can enrich our theory of language, either by offering support for previous proposals or challenging current predictions. Chapters 3 and 6 in particular explore the implications of Wamesa data for OT and the sorts of representations and constraints that are necessary to account for the attested patterns. I have also tried here to include as much illustrative data as is feasible, so that this work may be useful to those investigating other questions or using other theories.

The remainder of this chapter will discuss the speakers of Wamesa, its use and endangerment status, its relations to other languages of eastern Indonesia, and the fieldwork that underlies this research. §1.6 gives an outline of the rest of the dissertation. Grammatical description and analysis commences in Chapter 2.

1.1. Dialects, Location and Speakers

1.1.1. Geographic Position

Wamesa is traditionally spoken in the south-eastern part of the Bird's Head of West Papua, Indonesia, on the island of New Guinea.[1] The Wamesa-speaking area covers a roughly triangular territory demarcated by a spot just beyond Bintuni in the west, the base of the Wandamen Peninsula in the south-east, and Rumberpon Island in the north-east. It is bordered by the Austronesian languages Meoswar, Roon, Dusner (all Biakic), and Tandia to the east, Umar (also called Yeretuar) in the south-east, and Kuri to the south-west; as well as the Papuan languages Arandai and Kemberamo (Trans-New Guinea family) to the north-west, Meyah, Manikion, and Moskona (East Bird's Head-Sentani family) in the north, and Mer (Mairasi family) to the south (Lewis et al. 2013). This is a diverse group, and not atypical for the region; West Papua province is home to approximately 60 languages belonging to seven language families, plus three isolates (Lewis et al. 2013). Contact levels have historically been high, particularly along the coastal trading routes, with visible consequences. Wamesa has reportedly been used as a lingua franca between the different linguistic groups of the area. Bilingualism is recorded in both directions; my consultant IMK claimed to be proficient in Roon and Biak as well as Windesi Wamesa, all three remaining Dusner speakers also speak Wamesa (Dalrymple & Mofu 2012), and Umar shows a large number of loans from Wamesa (David Kamholz p.c.). Though Wamesa is unambiguously Austronesian, it shows some features which Klamer (2002b) identifies as likely the result of contact of substrate influence from Papuan languages when found in

1. A note on terminology: There are two Indonesian provinces on the island of New Guinea: West Papua (formerly West Irian Jaya), where Wamesa is spoken, and Papua (formerly Irian Jaya) to the east. I will follow local convention in using the term *Papua* to refer to the region as a whole, encompassing both provinces; *West Papua* to refer specifically to that province; and *Papua Province* to refer to that easternmost province in particular.

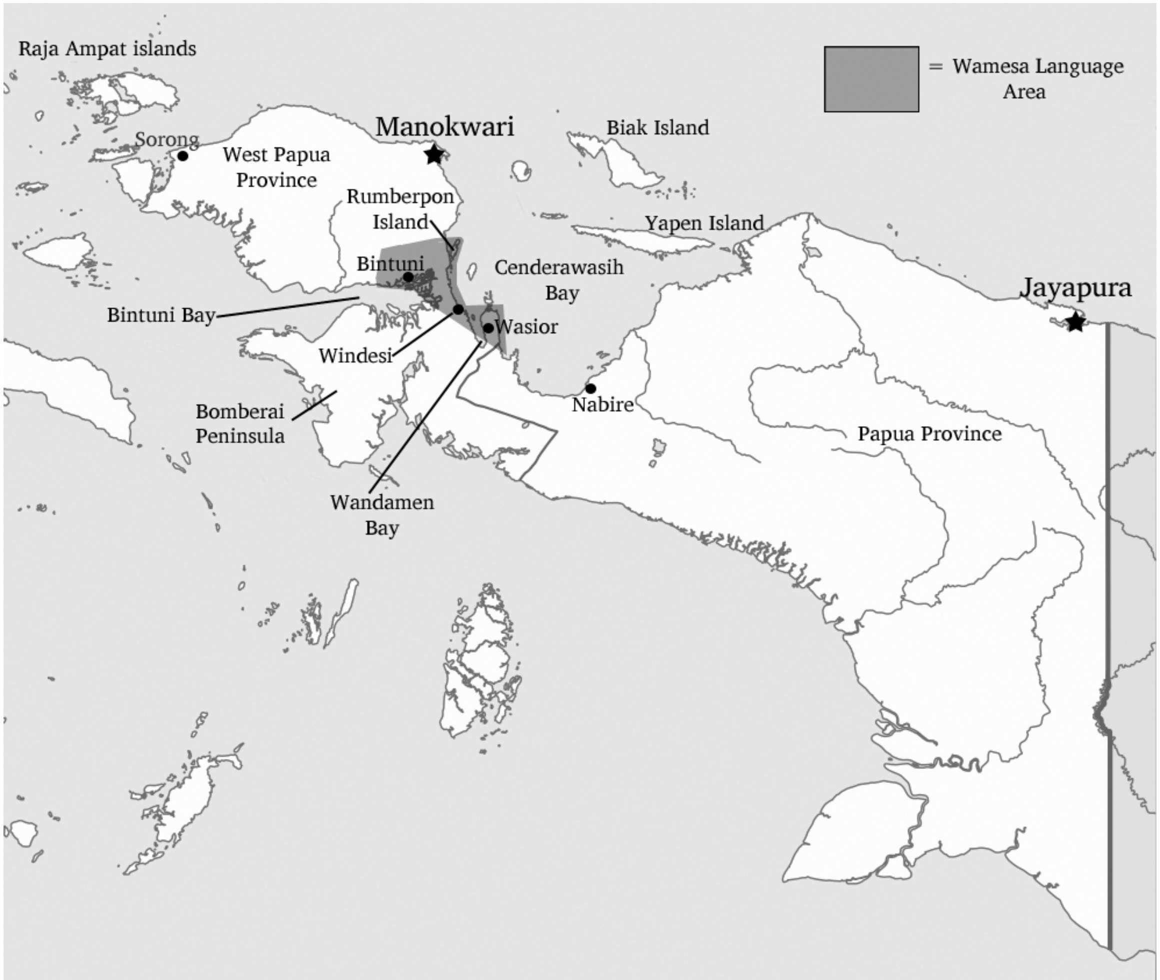


Figure 1.1: Wamesa in Geographic Context (adapted from NordNordWest/Wikipedia 2010)

the Austronesian languages of eastern Indonesia, for example phrase-final negation.[2]

1.1.2. Dialects

There are three major dialects of Wamesa: Windesi, Wandamen, and Bintuni. Windesi, the dialect described here, is a coastal dialect, spoken on Rumberpon Island and in villages on the coast of Cenderawasih Bay from the northern border of the language area south to Wandamen Bay. The Wandamen Peninsula and the coast of Wandamen Bay are home to

2. See Chapter 4 for a discussion of the negator enclitic *va* in Wamesa.

the Wandamen dialect. This is perhaps the best-documented of the dialects, having been studied by a number of early missionaries as well as a group of SIL linguists, including Theodore and Jean Henning, Rachel Flaming, Naomi Saggers, and Nitya Ongkodharma, in the 1970's and 80's. (See §1.5 for references.) While the Windesi area is reachable primarily by motor canoe or on foot (some areas are accessible via very rough roads from the regional capitol, Manokwari), the city of Wasior on the Wandamen Peninsula is visited several times a week by ferries traveling between Sorong, Manokwari, Nabire, Yapen, and Biak, and from there often on to Jayapura and major ports in western Indonesia. Wasior is also accessible via Susi Air from Manokwari. The third major dialect, Bintuni, covers the mountainous regions west of the Windesi dialect to Bintuni Bay. This inland area is perhaps the least accessible of the three, with no roads or airports. The city of Bintuni can be reached by a difficult eight-hour drive from Manokwari or by Susi Air from Sorong or Manokwari.

No cross-dialectal study has been done to determine the defining features of each dialect or the extent of the variation between them. My data contains wordlists from each of the three dialects; comparison of these shows limited lexical difference in core vocabulary, though moving beyond a Swadesh list and comparing data from Henning et al.'s (1991) Wandamen dictionary with items provided by my Windesi consultants reveals a greater degree of lexical divergence. Some examples in of lexical difference include Wandamen *atuma*, Windesi *mararea* 'child', Wandamen *anda*, Windesi *piderekari* 'mango', and Wandamen *sianggono*, Windesi *urar* 'red'. Wandamen and Windesi have identical phoneme inventories, but many instances of [ŋ] in Wandamen appear as [k] in Windesi, as in Wandamen [ŋaŋgomi], Windesi [kakomi] 'nutmeg' and Wandamen [ŋaŋgane], Windesi [kakane] 'hawk'. Windesi speakers report that certain constructions and word order choices are different in Wandamen, but I have no direct evidence to confirm this. Speakers also report salient differences in intonation contours between the dialects.

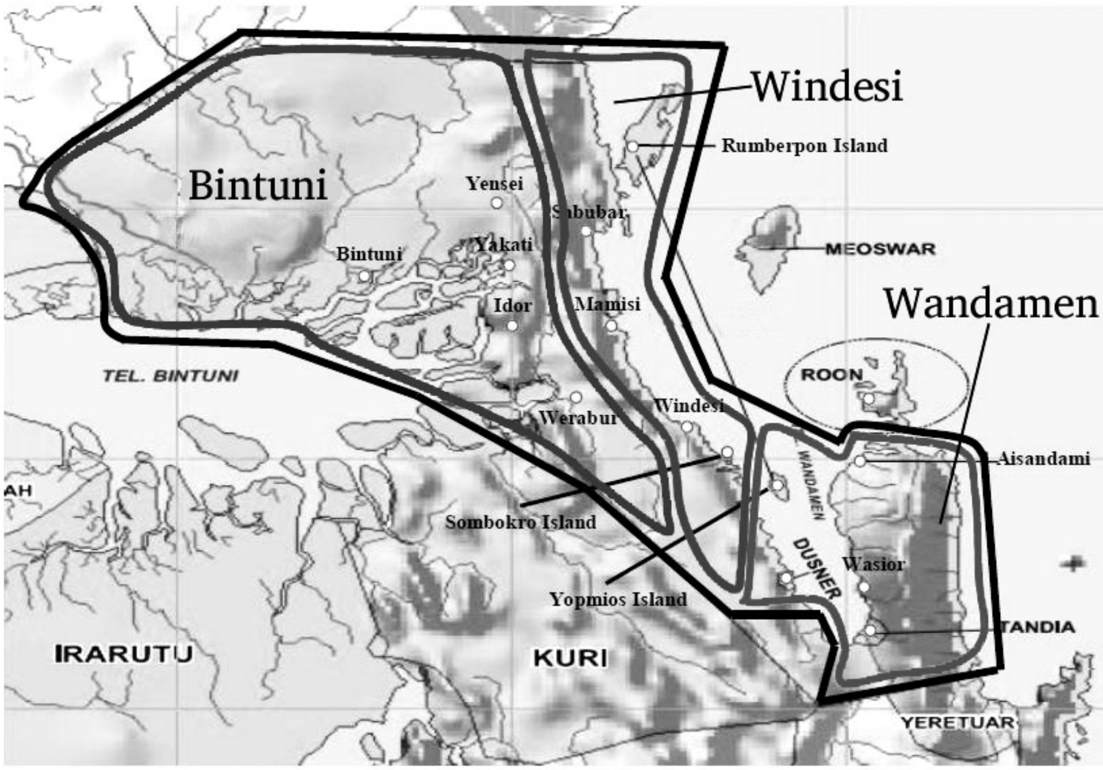


Figure 1.2: Wamesa Dialect Map (Karubuy 2011, adapted from SIL)

On one occasion TLB, a Wandamen speaker, and IMK, a Windesi speaker, participated in an interactional task together. TLB claimed that she could not understand Windesi so IMK offered to speak Wandamen during the task; the two conversed without issue though IMK was inconsistent in her use of Wandamen features. Based on this episode and my own experience with materials from the two dialects, they seem to be mutually intelligible, despite TLB's reservations. My experience with the Bintuni dialect is far more limited, but it too appears to me to be intelligible with the other two. Saggers (1979) agrees with this assessment.

1.1.3. Speakers

Depending on the source, the number of Wamesa speakers is listed as either 5,000 (Lewis et al. 2013; Ongkodharma 1985) or 8,000 (Henning et al. 1991; Flaming 1983b). All of these, however, with the exception of the Ethnologue, refer specifically to speakers in the Wasior area, who would almost exclusively speak Wandamen. Saggers (1979), who carried out her six months of fieldwork in Manokwari and the Wasior area, reports 5000 Wandamen speakers in the Wasior area and larger Papuan cities, and only approximately 500 of the Windesi and Bintuni dialects combined. Given that the population of Windesi District alone is just over 2,500 people according to the 2010 census, even if the majority of these do not speak the language (as many surely do not), 500 strikes me as an underestimate. Windesi is one of approximately 13 districts, albeit mostly sparsely populated outside of ethnically mixed urban areas, which fall within the Wamesa language area.

At least in the area around Windesi Village, people largely get by through fishing, hunting, and sago farming. Papeda and smoked fish are the staple foods, and children are adept at finding bamboo shoots in the forest or sea snails at the beach and roasting them over a fire to eat. Rice, tofu, coffee, sugar, and other staple foods of urban Papua were only available through trade to Wasior, a time-consuming and expensive trip by *jonson* (motorized outrigger canoe). At the time of my visit electricity was only available for a few hours each night, and even then only if gas was available to run the generator; there was no cell phone reception until much closer to Wasior. Discussions were underway to elevate the village's status in the regional administrative hierarchy which, if carried out, would bring cell phones and internet, expand the village significantly, and deepen its harbor in order to make it a suitable port of call for large ships; this has not yet been carried out.

The Wamesa speakers whom I got to know were all devout Protestants, though their precise denomination varied. The presence of Christianity in the area can be traced back to the Dutch missionaries of the late 19th century. The most influential missionary in Windesi District was J. A. van Balen, who arrived in Windesi in 1889 and stayed there for an extended period of time. A monument, erected in 1994, now stands at the site of his house, the school is named after him, and the church he built was under renovation at the time of my visit. Though his stay in Windesi was over 100 years ago, people still bring it up as an important event for the village, referring to it as though it were much more recent. Van Balen's Bible translation into Wamesa was published in the Netherlands in 1915, and copies are highly valued.

1.2. Wamesa Genetic Affiliations

Wamesa is a member of the South Halmahera-West New Guinea (SHWNG) branch of Eastern Malayo-Polynesian, within the Austronesian language family (Adriani & Kruyt 1914; Esser 1938; Blust 1978, 1993a). SHWNG is a sister to Oceanic, and is comprised of an estimated 45 languages spoken on the southern half of Halmahera Island and its satellite islands, along the western coast of Cenderawasih Bay on the Bird's Head of New Guinea, on the islands of Raja Ampat, Yapen, and Biak, and in the interior of the Bomberai Peninsula (van den Berg 2009). SHWNG is further divided into South Halmahera (SH) and

West New Guinea (WNG) branches (Blust 1993a; Ross 1995), though the affiliation of the Raja Ampat languages is debated (Remijsen 2001; van den Berg 2009). Within the West New Guinea group, subgrouping is much less clear, and very little comparative work has been published on these languages. Blust (1978) divides WNG, which he calls Sarera, into four branches, after Anceaux (1961): Biakic, Waropen, Moor, and Yapen. The Ethnologue (Lewis et al. 2013) takes a slightly different tack, dividing WNG into the Bomberai and Cenderawasih Bay groups, with the latter breaking down into Biakic, Iresim, Moor, Raja Ampat, Tandia, Waropen, Yapen, Yaur, and Yeretuar. The Yapen group is further subdivided into East Yapen, consisting of Wabo and Kurudu, and Yapen proper, to which belong all of the other languages of Yapen Island as well as Wamesa. Wamesa's closest relatives, then, are the Yapen languages Wooi, Ambai, Ansus, Busami, Marau, Munggui, Pom, Papuma, and Serui-Laut.

The tree in Figure 1.2 is adapted from the familial structure presented in the Ethnologue.[3] The flat structure of the Cenderawasih Bay group is due to lack of data and investigation, rather than reflecting a positive hypothesis about the actual relationships among the groups included therein. In Chapter 6 I argue for a Biak-Yapen group within Cenderawasih Bay based on the distribution of verbal infixation, adding one more layer of articulation to the tree given here.

There are very few grammatical descriptions of Cenderawasih Bay languages available. The best-documented are the Yapen language Ambai (Silzer 1983) and Biak (van den Heuvel 2006; Mofu 2009). A short sketch grammar exists of the Biakic language Dusner (Dalrymple & Mofu 2012), as well as an older grammar of Waropen (Held 1942). An in-depth study is currently under way on Wooi (Sawaki in prep), like Ambai a close relative of Wamesa, and fieldwork has been carried out by David Gil on Roon and by David

3. Some language names have been altered to reflect speaker preferences (after Kamholz p.c.); Roon has been relocated from the Yapen group to Biakic (Gil 2010).

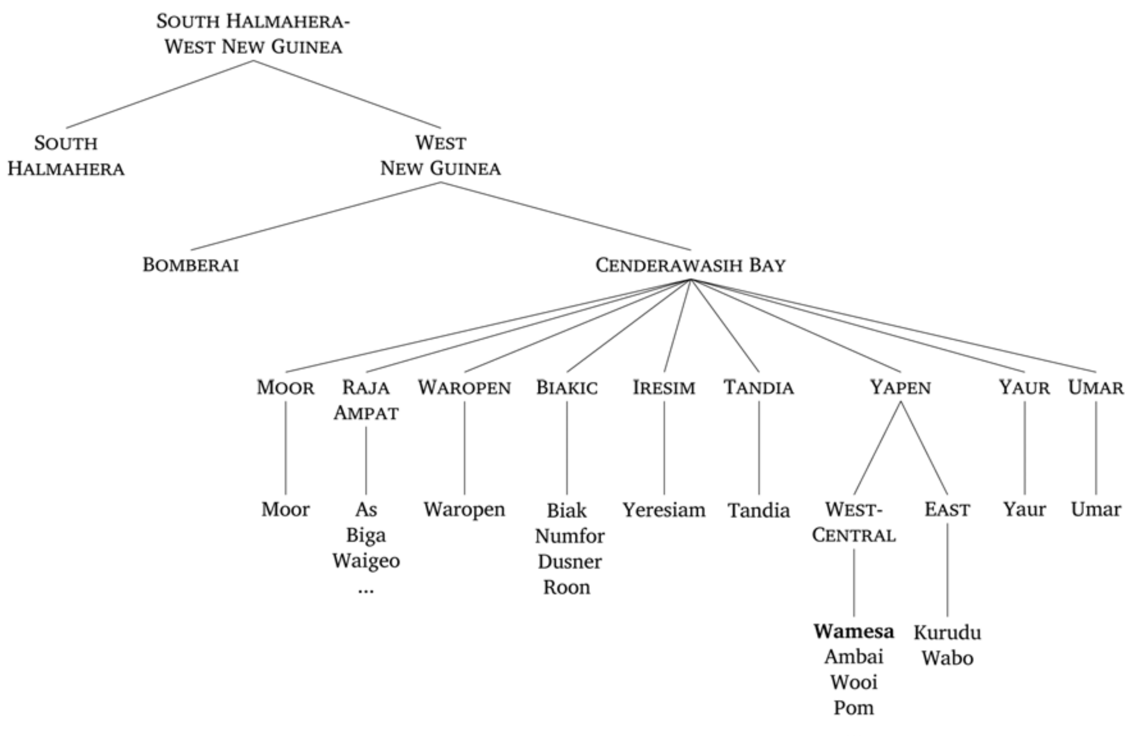


Figure 1.3: Wamesa in Genetic Context (adapted from Lewis et al. 2013)

Kamholz on Moor and neighboring languages, but no grammar has yet been published. Slump (1924-38) compiled a description of Serui-Laut written in Malay, which is available in Het Utrechts Archief in Utrech but was never published. For the remaining languages, available resources are limited to wordlists and occasional grammatical information collected by early missionaries and as part of more modern surveys (Cowan 1953; Anceaux 1961; Holle 1982; Grimes 1990; Anceaux 1992; Price & Donohue 2009, among others).

1.3. Fieldwork

The data used here comes primarily from the nearly seven months I spent working with native speakers of Wamesa over three trips in 2011-2014. The first of these trips, in June - August of 2011, lasted six weeks and was partially funded by the Yale Council for Southeast Asia Studies. This trip served as a pilot for a later extended field trip in September through December of 2012. I returned for a month in January - February 2014 to ask follow-up questions and participate in the 3rd Workshop on the Languages of Papua, hosted at the state university in Manokwari. All work subsequent to initial preliminary visit was funded by a Dissertation Research Improvement Grant from the National Science Foundation (DEL-1153795).

1.3.1. The CELD

During this these field trips I was based primarily in Manokwari, the capitol city of West Papua Province and home of the Universitas Negeri Papua (Unipa), the state university, which hosts the Center for Endangered Languages Documentation (CELD). The CELD was established at Unipa in 2009 with the goal of working to document and preserve the local languages of Papua, most of which are to some degree endangered. The CELD is locally run, and its staff has major documentation projects in progress concerning three languages: Wooi, an Austronesian language closely related to Wamesa and the subject of co-founder and director Yusuf Sawaki's doctoral dissertation, currently underway at the Australian National University; and Iha and Yali, Trans-New Guinea languages spoken in the Bomberai Peninsula and the highlands south of Jayapura, respectively. For their senior projects, linguistics majors at the university choose a local language to research, and write a thesis on some aspect of it. The center provides support and equipment for these projects and trains students in how best to go about their fieldwork, and many of the languages studied in this way are largely (if not entirely) undocumented otherwise. Finally, the center hosts international researchers, including several such as myself who are engaged in long-term projects in the area.

A core belief around which the CELD was founded, as described on their website (www.celd-papua.net), is the idea that language and culture develop from and encode a worldview and body of knowledge specific to the human experience of the society to which they belong and the environment in which that culture is situated. Preserving these languages and cultures is thus imperative, as the loss of a language entails not just the loss of a treasured piece of world heritage, but also the self esteem and identity of that society.

This outlook, combined with the reality that language shift to Papuan Malay is leading to a rapid decline in speaker numbers for many languages of Papua, provides the motivation for the CELD to welcome in and partner with outside researchers and encourage their work in the region: the more bodies actively involved in research, the more chance there is to begin to make a dent in the massive amount of work yet to be done before the languages in question disappear entirely. The center thus works to raise awareness of the plight of these languages through publication of scholarly work, conference presentations, and other outreach, both locally and internationally, to attract support for their mission of documentation. The end goal of the CELD is to build local capacity to the point that Papuan linguists can work on Papuan languages, carrying out their own high-quality documentation projects, and to support the use and propagation of indigenous languages.

This emphasis on local capacity underlies the center's four main directions of emphasis in its work: the internal documentation projects mentioned above; the training of Unipa students in fieldwork techniques; support of teachers, artists, local governments, etc. in the development and use of local language materials; and the creation of sustainable archives of linguistic and anthropological data within the CELD, so that such data is accessible to the people whom it most concerns (www.celd-papua.net). In addition to pursuing their own research in the region, outside linguists sponsored by the CELD contribute their time and expertise towards facilitating the above goals.

Collaboration with the CELD proved vital to my fieldwork. They provided me with working space in their office, access to the campus wifi network, advice and sponsorship with regards to obtaining the necessary visas and travel permits, help when I ran into challenges with unfamiliar Papuan Malay constructions or vocabulary, and practical advice about living and working in Papua, not to mention friendship. It was through them that I found a place to live, and that I was connected to my first Wamesa speakers. Their expertise on linguistic matters also proved invaluable. Because of their experience working with the languages of the area, the CELD is a rich source of specialized knowledge on those languages. If I didn't know what to make of an aspect of Wamesa or got stuck in my analysis of a challenging data set, there was always someone in the office who could explain to me how that phenomenon worked in Wooi or another closely related language, which often proved very similar to the Wamesa structure. These cases too provided insights into the sorts of things I should look for in my elicitation, which I otherwise might not think to explore.

For my part, I was able to further the Center's mission primarily through teaching and mentoring students. The CELD promotes a 'shared knowledge approach' to collaboration, in which visiting researchers benefit from their expertise as well as vice versa. In my first two visits, I helped lead the weekly Reading Circle, in which students would work their way through a linguistics textbook, do the problem sets, and discuss the material as a supplement to their coursework. I also worked directly advising students who were planning field trips to do their own documentation, or analyzing data from those trips, particularly one senior, Nova, who was working on her family's dialect of Wamesa.[4]

It was also through the CELD that I came to work with a student assistant, Cika Tethool. This relationship much more closely resembled one of mentorship than of employment, though I did pay her for her work. Working with me on my recordings provided training in transcription of an unfamiliar language and use of relevant software (in this case, Elan and Lexique Pro); accompanying me occasionally to and participating in elicitation sessions worked as a real-world field methods course. I could help Cika work out answers to questions about her own data, gathered for her senior essay on a related language, and she could help me with questions of culture and Papuan Malay. Working together on Wamesa, we each benefited from the other's expertise in ways that advanced our own research agendas.

1.3.2. Travel within Papua

Though the majority of my time was spent working with speakers in Manokwari, I was able to do some limited traveling within West Papua to visit other Wamesa-speaking areas. During my initial visit, my host sister, Juen, invited me to accompany her to Bintuni, where I was able to record wordlists from the Bintuni dialect and from Kuri, a related language spoken adjacent to the Wamesa area, and to visit a Wamesa *kampung* (neighborhood or village). During my second trip, my primary consultant, Ibu Marice Karubuy, accompanied me on a week-long trip to her native village, Windesi. Her younger brother, Pak Aukila Karubuy, is the village head, and hosted us for the duration of our stay. Dur-

4. Copies of all of my recordings (as well as, soon, a copy of this dissertation), are housed in the CELD's archive with permission of the speakers themselves, so that they may continue to be useful to students like Nova, as well as other interested community members.

ing that trip we also spent a night in the much smaller village of Sombokoro, where her mother grew up, and two nights with Ibu Marice's family in different parts of Wasior. All of our hosts were speakers of Wamesa, and though I only made recordings in Windesi and Sombokoro, all were happy to talk about my project and about their use of the language.

1.3.3. Consultants and Methods

While in Manokwari, I lived with a Manadonese family near the university. Six days a week I met with Wamesa speakers to study the language. The majority of my elicitation was done with Ibu Marice (IMK), a schoolteacher in her mid-50's. I was introduced to Ibu Marice through her son, Theo, who at the time of my first visit had just finished writing his senior thesis, (*skripsi*) on Wamesa verbal morphology, based on data from his mother (Karubuy 2011). Ibu Marice turned out to be an excellent consultant, and I continued to work with her for the duration of my three visits. In Manokwari I also worked a number of times with Lorensina Biambara (TLB), a Wandamen speaker who worked as a cleaner at the university, and her husband Marten Windesi (BMW), whose dialect is largely Windesi with some Wandamen features, as he grew up in the transitional zone between the two dialects. In Sombokoro I recorded an elderly uncle of IMK's telling a traditional flood story, once in Wamesa and then again in Papuan Malay. In Windesi I recorded frog stories from Pak Aukila (BAK) and another uncle, David Parairawai (ODP), as well as picture naming and discussion involving many other community members. In all, approximately seventeen speakers appear in my recordings to various extents.

I utilized a range of elicitation methods in my work. On first meeting a speaker, I started with a paradigm for 'to eat' and a wordlist, reduced from that constructed by Willem Burung of Unipa (Burung 2011). This list is meant to include basic vocabulary similar to that of the Swadesh 200-word list, including objects and ideas particularly relevant to languages of Papua, such terms for the sago mush *papeda*, *koteka* (penis gourds), certain familial relationships common in the area (*mama adik*, *mama tua*), cassowaries, *noken* (traditional net bags), etc. Continuing forward, I used a combination of traditional storytelling, elicited storytelling, descriptions of events, picture naming, conversations, and grammatical elicitation to gain a well-rounded picture of the language and its structures, as suggested in the fieldwork literature (Chelliah 2001; Crowley 2007; Bowern 2008, among others).

Traditional stories are useful both as a rich source of natural speech and as a repository of cultural traditions, but proved challenging to record, as only certain people have ownership of any given story, and only they can retell it. Therefore I mostly relied on more structured prompts to elicit narratives. I brought with me three of Mercer Mayer's wordless children's books (Mayer 1967; Mayer & Mayer 1971; Mayer 1969) to elicit frog stories; each of these was retold by a number of speakers, and the limited vocabulary each called for made them useful for comparison. The pear video (Chafe 1980) is similarly wordless, and like the frog stories asks the speaker to retell what they have seen in their own words, providing natural speech within predictable boundaries. In many cases I would try to elicit similar natural but predictable speech by asking speakers to describe a process, such as how they make papeda from sago, or retell an event, like our boat ride from Windesi village. These recordings were later reviewed with speakers to accurately transcribe and translate them.

Another effective prompt was asking speakers to describe or name photographs. I began in the city with photos of plastic figurines designed to elicit quantifier scope judgments (Bruening 2001); while I got no scope judgments from them the resulting descriptions nonetheless proved a rich source of other data. A strategy which proved successful in the village was for me to take photos of flora and fauna while on walks with my hosts; I later transfered these to my laptop and presented them to a group of speakers who gave me the Wamesa name for whatever creature appeared in the photo, often accompanied by a discussion of its use and where it was found, and also sometimes debate over what to call it.

Interactional data was harder to come by. An attempt at a map task (Anderson et al. 1991), where one speaker describes to another a route on a map of which each has a slightly different copy, failed entirely; the speakers simply named the items on the map. A more successful attempt came at the end of my final visit, when IMK and TLB participated in a DoBeS[5] project on phonological and gestural entrainment, for which Dr. Sonja Riesberg was collecting data at Unipa. In this task, one speaker holds a card with a geometric picture and describes it to a second speaker, who must decide which of the four very similar pictures on her card is being described. The speakers were able to perform this task fairly successfully, and were recorded both for my corpus and that of DoBeS.

Finally, my recordings include a large proportion of grammatical elicitation. These include more vocabulary lists, using Henning et al.'s (1991) Wandamen dictionary as a basis, full grammatical paradigms, translation of Papuan Malay sentences into Wamesa and Wamesa sentences into Malay, grammaticality judgments, and follow-up or clarification questions from previous sessions.

1.3.4. The Corpus

The corpus of Wamesa data collected during this time consists of 83 recording sessions, totaling just over 100 hours of audio recording[6] and approximately 45 minutes of video. The dictionary being compiled from this corpus, currently in progress, includes just under 1000 Wamesa items at time of writing, and is often referred to here. So as to be useful to Wamesa speakers, the dictionary translates headwords and example sentences into both English and Indonesian; photos are also included where relevant. Data from other sources,

5. *Dokumentation bedrohter Sprachen*; http://dobes.mpi.nl/

6. 2011: 29 hours; 2012: 58 hours; 2014: 15 hours.

described in §1.5, does not constitute part of the corpus considered here, for reasons of dialect differences and source reliability. When others' data does occasionally become useful it will be cited as such. Items from my own field notes in most cases will be cited here with the session number from which they are taken and the initial of the speaker who produced the example or judged its grammaticality.

1.4. Language Attitudes, Use, and Endangerment

Hale (1992: 1) discusses a modern paradigm of language loss in which "politically dominant languages and cultures simply overwhelm indigenous local languages and cultures, placing them in a condition which can only be described as embattled''. This accurately describes the situation of Wamesa vis a vis Papuan Malay, which developed out of Malay varieties from Ambon and Sulawesi and has been present in coastal Papua at least since the 19th century (Sawaki in prep). Papuan Malay is the *lingua franca* of the region, and its use is spreading from inter-group communications outside the home to intra-group conversations within the home as many indigenous languages fall out of use. The use of Wamesa is similarly declining, particularly in coastal and urban areas, and it should be classified as threatened. I worked with four speakers in Manokwari. One, IMK, is married to a man from Sulawesi, halfway across the Indonesian archipelago; her children have some passive competence in Wamesa but cannot be considered speakers themselves. She sometimes addresses her children in Wamesa and speaks the language with Wamesa friends and when visiting relatives, but uses Malay for most daily interactions. A second speaker, BMK, similarly uses mostly Malay in the home; his son, YK, identifies as Wamesa but does not speak or understand it. The remaining two speakers, BMW and TLB, are married to each other, and use a mixture of Wamesa and Papuan Malay in the home. In Windesi Village, adults conversed in Wamesa and Malay in roughly equal proportions, mixing varieties over the course of a conversation. Code switching both across and within utterances. Young adults and children, however, spoke only Papuan Malay, with some passive competence in Wamesa. This is the same situation I encountered in the towns of Wasior and Bintuni, both situated within the traditional Wamesa-speaking area. Schooling, governmental functions and all mass media use varieties which fall somewhere on the spectrum between local Papuan Malay and the national language, Indonesian, making that the more 'useful' language, and necessary for life outside the village. I was told that children in villages farther inland still learn Wamesa as a native language, though I was unable to travel to observe this. Though the college-aged children of Wamesa speakers whom I met all showed interest in their parents' language – two were students in the linguistics department at the state university writing their *skripsi* on aspects of Wamesa; another, a police officer, proudly volunteered the handful of words he knew – none could speak it themselves.

Despite this, the Wamesa speakers I interacted with showed a great deal of pride in their language. I was without exception warmly welcomed by speakers, both those with whom I worked in depth and those whom I met in passing in the villages and in Bintuni and Wasior. When the subject of my documentation project inevitably came up, it too was consistently received with enthusiasm, usually immediately followed by the speaker trying to teach me a few words. The speakers I interacted with were genuinely proud of their language and excited to teach it to me, and nothing elicited so much delighted laughter in the village as the spectacle of my trying to speak it. Similar support was also given to the idea of my sharing my knowledge of Wamesa with others at home. While the nature of linguistic analysis and academic publication were not necessarily fully grasped, speakers understood that I was writing my dissertation (*skripsi S3*) on the language and would be writing articles and teaching others about it, and encouraged me in doing so. There were several aspects to their reaction. One was the prestige gained by having an outsider, particularly an American, spend time in the village and show interest in the language. That I found the Wamesa language interesting enough to travel halfway around the world to spend several months studying it and then write a book on what I had learned conferred particular prestige on my hosts. There was also a religious aspect to their enthusiasm, as Europeans are still associated with the Dutch missionaries of the late 1800's, who are still revered figures. IMK, described her reasons for devoting so much time to working with me by saying: "My principle of language is this: God gave it, we share it. We can't hide what we've received. That's not good. We have to share our knowledge with people."[7] BAK too expressed what he saw as religious/spiritual benefits incurred by my visit. By sharing their language with me, and encouraging me to share it with the world at large, speakers are, as they see it, doing a spiritual good deed.

What I found was that people enjoyed speaking Wamesa, and that it served as a signifier of an identity and culture of which they were equally proud. This impression was echoed in an email I received from Theodore Henning, who worked extensively with speakers in Wandamen Bay until about 20 years ago. Henning (p.c.) characterized the language as "the one overriding unifier for the various clan groups… who call themselves 'Wandamen'," and expressed surprise that intergenerational transmission had declined so drastically. I would argue that despite the lack of child acquisition, the language maintains its unifying symbolic value. This surely contributed to the pleasure many speakers seemed to derive from teaching it to me, either through direct elicitation/instruction or through storytelling, an act allowed them to share the language with a wider audience, often including non-fluent children looking on.

It bears mentioning that while speakers were proud of the existence of my project and excited to teach me their language, they showed no interest in guiding the research or becoming collaborators rather than teachers, and as there was no sense among speak-

7. *"Saya punya princip bahasa ini: Tuhan kasih, kita membagi. Tidak boleh sembunyi yang dapat. Tidak bagus. Harus membagi kepintaran buat orang."*

ers that the language might be endangered, there was no demand for a maintenance or revitalization program. Though I would have been happy to train consultants to take on these more-involved roles, I did not feel it was my place to push it on those who weren't interested. This is one major contrast between the goals of the speech community and those of the CELD, where training formed a major part of my contribution.

1.5. Previous Literature

1.5.1. Existing Sources

The existing research on Wamesa is largely lexicographic. Many early wordlists come from Dutch missionaries and colonial administrators posted to the region in the late 19th and early 20th centuries. Early handwritten notes exist in the KITLV archives in Leiden, though these can be difficult to decipher. The earliest published source is Bink's (1891) *Lijstje van telwoorden en eenige zelfstandige naamwoorden enz van Wandamen*, a word list collected by the author, location and dialect unspecified. This material is included in Anceaux (1992), discussed below.

The next major publication chronologically is Holle (1982). This multivolume work includes wordlists from numerous languages throughout the archipelago, two of which correspond to dialects of Wamesa. These two lists were collected in 1895 and 1906, respectively; the first was originally published as (1915b). Each includes between 700 and 800 lexical items, with minimal notes on pronunciation and some explication of meaning. The two lists resemble one another closely; the majority of discrepancies may well be due simply to differences in transcription conventions.

Van Balen also produced a bible translation into Wamesa (van Balen 1915a). Copies of this still exist in Wamesa households - I was given a copy by one consultant - though its language differs somewhat from the language as it is currently used, in both lexical content and application of phonological processes such as cluster reduction. A Wamesa hymnbook was also in use in the village, possibly the same one, published in 1941 by the Protestant Mission in Miei, referred to by Cowan (1955) in his source list.

Anceaux (1992) compiles wordlists from 39 languages of the area, including Wamesa.[8] This work too consists purely of lexical data. The Wamesa material is compiled from twenty published and unpublished lists collected in various locations by linguists including Anceaux himself, as well as by local functionaries. Anceaux (1961) includes a very similar wordlist, along with a few verbal and possessive paradigms for each variety. Anceaux here uses this data to infer genetic relationships between the languages. These conclusions largely coincide with more recent publications on the topic (i.e. Greenhill et al. 2008; Lewis et al. 2013), though the exact set of languages used varies between sources, as do some of the low-level relationships, such as the exact placement of Roon and Moor in relation to the other languages of the area.

Most recently, Henning et al. (1991), published by SIL International, is an extensive glossary of the Wandamen dialect with translations into English and Standard Indonesian. SIL has also put out a book of *Wandamen Conversations* (Ramar et al. 1983), which is aimed at language learners and gives a series of short narratives and conversations in the same three languages.

Wamesa has also been given mention in a number of other surveys of the area, including Cowan (1953), which lists and groups 24 languages of what was the Dutch New Guinea; Siltzer & Heikkinen (1984), a much more extensive index of the languages of Papua with classification, maps, speaker counts, and references; and Price & Donohue (2009), an SIL report on the sociolinguistic position of Ansus and other Yapen languages.

Little linguistic analysis of Wamesa has been published to complement our knowl-

8. The full work is many volumes long, and includes data from languages across the Indonesian archipelago.

edge of its lexicon. Ramar et al. (1983) includes an introductory chapter giving a brief descriptive overview of the phonology, verbal morphology, and word order, without any theoretical account. The sections on pronunciation and verb conjugation are reproduced in Henning et al. (1991).

Kamma (n.d.) produced what is likely the first grammatical description of Wamesa, written in Dutch roughly mid-century. This work is unpublished, and available in Het Utrechts Archief.

H.K.J. Cowan's (1955) 18-page article *Notes on Windesi Grammar* is more detailed than Henning et al., but drawn from second-hand information, a combination of wordlists and translated texts produced between 1915 and 1953 by a range of linguists, Dutch civil servants, and missionaries. The author did not collect any data on the language himself. Cowan's description of the phoneme inventory is based entirely on the spellings of words in the source texts, and is incomplete; he omits /r/ entirely and decides that /β/ is a "non-distinctive variant" (Cowan 1955: 44) (i.e. non-contrastive allophone) of /b/ or /w/. Cowan goes on to give a brief treatment of other aspects of the grammar, but again this is descriptive rather than analytical. The language described by Cowan is quite different from that which I encountered; whether this is due to language change, dialect differences,[9] problems in Cowan's original sources, his interpretation of the material, or some combination of factors, is unclear. Cowan's accounts of the verb conjugation paradigms require three verbal classes where a simple cluster reduction rule would do; he retains the prefix final C (almost always as *n*) where in my data cluster simplification has it disappear; and while he posits *r*-initial verbs as a separate class, he fails to notice that /r/ becomes [nd] in all derived clusters and that /β/ and /k/ behave similarly. (See §2.3.2 for more detailed discussion of cluster reduction.) His data always fails to reduce clusters at mor-

9. Cowan refers to the language as 'Windesi' in his work, but not all of his sources name the dialect from which they are drawn.

pheme boundaries, and also preserves geminates and morpheme-internal clusters which are simplified in my data, as in his *siniontu* 'person' and *antum* 'child', in my data *sinitu* and *atuma*,[10] and his *menne-pasiat* 'your people', *me-ne=pa-sia* in my data (from underlying /met-ne=pa-sia/, 2PL-have=DET-3PL.HUM). Many morphological features described by Cowan do not appear in my data, or appear with different meanings. The applicative morpheme *it-* (see §4.4.2), for example, is listed as *in-*/*rin-*/*tin-*, with the latter two forms incorporating the final consonant of the plural and dual subject agreement prefixes and the final nasal presumably coming from its form before *r*-initial roots. Cowan lists it as an imperfective aspect marker, and while that is one of its features, he neglects to mention its other aspectual meanings, or its far more prominent instrumental argument-adding function.

Though it is far more accurate than Cowan's sketch, Saggers's (1979) unpublished masters thesis on the Wandamen dialect also presents data which contradicts that in my recordings. Again, whether this is due to dialect differences, language change over time, or differences of analysis is unclear. For example, Saggers decomposes the forms *nini* 'this' and *nina* 'here' into bimorphemic *ni* + *ni*/*na* constructions, where I analyze them as monomorphemic. She interprets the topic-marking clitic *=ma* as a non-conjugating copula, and finds a transitivizing suffix *-rV* not present in my corpus. And she overcorrects for Cowan's mis-apprehension of the form of the applicative and causative prefixes: while she correctly lists the applicative as *it-* rather than Cowan's *in-*, she lists causative *on-* as *ot-*. Like Cowan, Saggers also lists the applicative as a simple imperfective marker. Saggers' thesis is rather more detailed than Cowan's sketch, but is entirely descriptive; she provides no theoretical account or analysis of the patterns she describes, other than a few basic syntactic trees.

10. This form is from the Wandamen dialect; Windesi uses *mararea*.

van den Berg (2009) compares the possessive constructions of eleven SHWNG languages, including Wamesa. His forms very closely match those produced by my consultants, though my speakers were far more permissive in their use of alienable possession for items traditionally inalienably possessed. Silzer's (1983) Ambai grammar makes use of some Wandamen lexical and morphological data from Ongkodharma, Flaming, and Saggers, along with some forms from other Cenderawasih Bay languages, as a comparison with Ambai and to explore possible diachronic scenarios. And Blust (1978) makes use of Wamesa data from Anceaux to argue for his subgrouping of the SHWNG languages, though he mildly misinterprets a few data points, positing for example [siŋ̯a] (his [siña]) rather than [sinia] as the pronunciation of *sinia* 'mother'.

Rachel Flaming of SIL has published two papers on the Wandamen dialect of Wamesa, 'Cohesion in Wandamen Narrative' (Flaming 1983a) and 'Wandamen Kinship Terms' (Flaming 1983b). These represent probably the most in-depth analysis of aspects of the language produced so far, though both are very limited in their scope. Ongkodharma (1985), researched as part of the same project which produced Flaming's articles and the SIL dictionary and conversation book, is more anthropological in nature, and describes traditional belief systems of the Wamesa people, since largely displaced by Christianity. Henning (2014) describes a poetic register of the language, particularly as it was used in funeral rites, which likewise has been largely eliminated since the spread of Christianity.

1.5.2. Unpublished Student Research

In addition to the sources mentioned above, some documentation and analysis of Wamesa has been carried out by undergraduate students studying linguistics at Unipa as senior thesis projects (*skripsi*). The two most recent *skripsi*, submitted in July 2011, are a paper on verbal morphology in the Windesi dialect by Theopilus Karobuy and one on the Wamesa pronomial system by Yesra Kandami. These are available in hard copy at the university.

A third *skripsi* is currently underway by Novalia Refwalu, on Wamesa comparative constructions.

1.6. Outline of this Dissertation

This dissertation has twin goals: description of the morphology and phonology of Wamesa, and theoretical analysis of some of its more interesting features. Description of the language begins in Chapter 2, which covers the phonemic inventory with phonetic detail, allophonic alternations, and other phonological processes such as cluster reduction. Chapter 3 gives an OT account of the stress system of Wamesa, including both regular stress assignment and a pattern of antepenultimate stress shift occuring with enclitics. Chapter 4 discusses Wamesa's bound morphology; word classes of free morphemes are discussed in Chapter 5. Chapter 6 gives an OT account of the verbal infixation pattern found on consonant-initial roots, and discusses how the distribution of similar infixing morphology in related languages can be used as a subgrouping argument to further articulate the Cenderawasih Bay group.

Chapter 2

Segmental Phonology

2.1. The Phonemes of Wamesa

2.1.1. Consonants

The Wamesa phoneme inventory includes eleven primary consonants and three marginal consonants. Three places of articulation are used contrastively: labial, coronal, and velar. Native words make use of oral and nasal stops, fricatives, and a tapped or trilled rhotic; laterals and affricates appear only in loans. Geminates do not appear in Wamesa; any which arise through morphological processes are reduced to singletons.[1]

	Bilabial	Alveo-dental	Velar
Nasal	m	n	ŋ
Plosive	p b	t d	k (g)
Fricative	β	s	
Affricate		(dʒ)	
Tap/Trill		r	
Lateral		(l)	

Figure 2.1: The Wamesa Consonant Inventory

1. The excptions are /r-r/, which surfaces as [nd], and word-initial /i-i/, which surfaces as [ji]; see §2.3.3 and §2.3.6 for discussion.

Minimal and near-minimal pairs illustrating these contrasts are given below in (2.1), with the segment in question in the environment /C$_{[\text{labial}]}$a_a/ where possible.

(2.1) *Wamesa Consonants*

p	/mapar/	*mapar*	'valley'
b	/baba/	*baba*	'big'
t	/βata/	*vata*	'good, true'
d	/padamara/	*padamara*	'lamp'
k	/makarabat/	*makarabat*	'eel'
g	/maŋgar/	*manggar*	'yell'
m	/mamara/	*mamara*	'clear'
n	/manau/	*manau*	'already'
ŋ	/waŋgar/	*wanggar*	'rat'
β	/βaβa/	*vava*	'under'
s	/masabu/	*masabu*	'broken, cracked'
r	/marapa rau/	*marapa rau*	'paddy oat leaf' (*sayur melinjo*; *Gnetum gnemon*)

Stops

Wamesa stop consonants occur at three places of articulation: bilabial, alveolar, and velar. The alveolar place of articulation for stops might be better described as alveo-dental; no palatographic data has yet been gathered for Wamesa but acoustically [t] and [d] sound quite dental in their pronunciation.

Voiceless stops can occur word-initially, intervocalically, and word-finally. Examples of each of the three voiceless stop phonemes in each of these positions are given in (2.2).

(2.2) *Distribution of Voiceless Stops*

	/p/	/t/	/k/
Initial:	[**p**are] ‘nipa palm’	[**t**abura] ‘conch shell’	[**k**akuna] ‘caterpillar’
Medial:	[sa**p**ami] ‘grasshopper’	[ta**t**ar] ‘sin’	[a**k**anak] ‘breadfruit’
Final:	[matitiota**p**] ‘destroy’	[suba**t**] ‘mud’	[sara**k**] ‘bracelet’

The voiceless stops are unaspirated, with a mean VOT of 19ms for [p] and [t] and 24ms for [k], below the threshold of perceptibility. VOT was measured in word-initial and -medial segments, from the release to the onset of periodic voicing of the following vowel.[2] In fast or casual speech, intervocalic stops sometimes lenite to fricatives, with clearly audible frication rather than the clean stops of more carefully articulated tokens. An example of this is given in Figure 2.2, where intervocalic /t/ surfaces as [θ] in the word *pibata* ‘turtle’. Spectrograms in this dissertation are usually labeled with a broad phonetic transcription; here the fricative is labeled as such though it does not contrast with the stop [t].

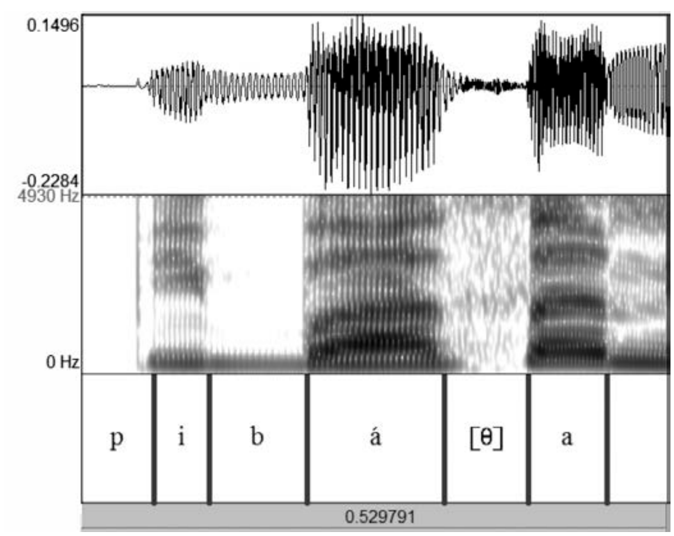


Figure 2.2: Lenition of Intervocalic /t/ with visible frication

2. N = 20 for each phoneme, evenly divided between word-initial and word-medial tokens.

Voicing is contrastive; one voiced and one voiceless stop occurs at each place of articulation. The contrast between [p] and [b], [t] and [d], and [k] and [g] is a true voicing distinction. Figure 2.3 contrasts the spectrograms for [t] and [d] intervocalically and [p] and [b] in word-initial position. The voicing bar, clearly visible for the voiced segments, is absent for the voiceless ones.

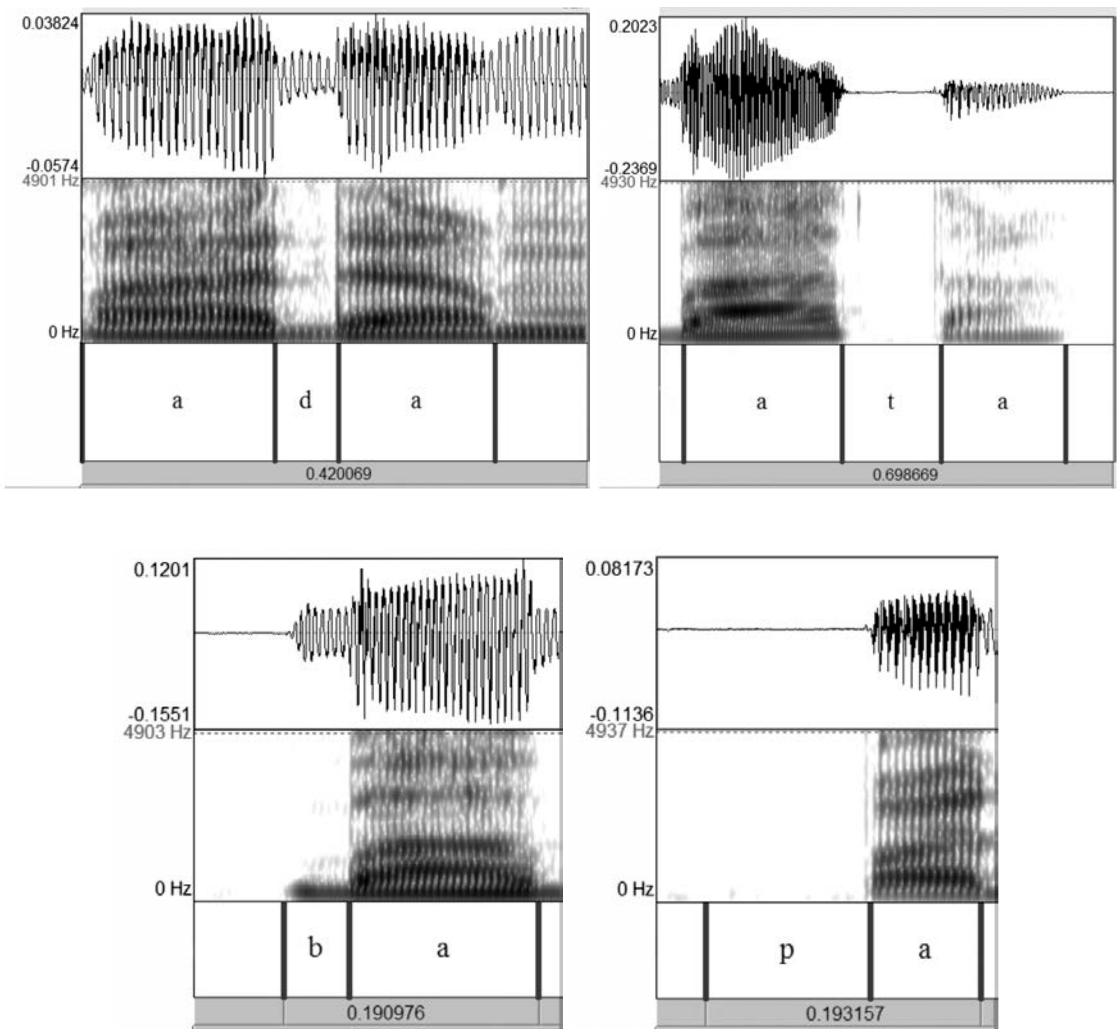


Figure 2.3: Voicing Contrast

Voiced stops are somewhat more restricted in their distribution than their voiceless counterparts. All three surface as the second member of homorganic NC clusters, and /b/

and /d/ occur word-initially and intervocalically as well. There is one instance of word-final /b/, in *Yob*, the name of an island near Windesi village, and none of final /d/. Except in loanwords such as *gomo* 'breadfruit' (from Malay) and *moga* 'crow', /g/ (source unknown, also present in Umar) only ever surfaces as part of an /ŋg/ cluster. Examples of voiced stops in each of their possible positions within the word are given in (2.3).

(2.3) *Distribution of Voiced Stops*

	/b/	/d/	/g/
Initial:	[**b**arimu] 'k.o. taro'	[**d**iru] 'night'	-
Intervocalic:	[sa**b**a] 'sago spine'	[do**d**eso] 'spike, arrow'	-
Cluster:	[kam**b**arai] 'not want, dislike'	[man**d**akiri] 'sea cucumber'	[aŋ**g**ara] 'lime'
Final:	[jo**b**] 'island near Wasior'	-	-

Nasals

As with stops, Wamesa contrasts bilabial, alveolar, and velar nasals. The distribution of the nasals is analogous to that of the voiced stops. The two nasals frontmost in the vocal tract, /m/ and /n/, are free in their distribution; they appear word-initially, intervocalically, as the first member of homorganic NC clusters, and occasionally also word-finally, though this last position is rare. The velar /ŋ/, however, may only occur root-initially or as part of a word-medial /ŋg/ cluster. Zuraw (2010) notes that stem-initial /ŋ/ is more marked cross-linguistically than /n/, which is in turn more marked than initial /m/; she proposes a family of markedness constraints set in a stringency hierarchy to account for this. Her predictions are well borne out by the Wamesa lexicon: There are 94 *m*-initial roots so far entered in my dictionary, compared to 23 *n*-initial and four *ŋ*-initial roots.

Note that there are cases of word-medial /ŋV/ sequences; these occur when a verbal agreement or other prefix is attached to a /ŋ/-initial verb root, as in [i-ŋaŋgau] '1sg is confused', where the morpheme boundary falls directly after the initial vowel. While there are exceptions, most words with /ŋ/ take the shape (ŋ)V_{α}ŋgV_{α}X or VNVŋgVX, where N is any non-velar nasal and X represents an unspecified amount of additional segmental material. Some examples are given in (2.4) - (2.5), and Figure 2.4 gives a spectrogram of an /ŋg/ cluster. This distribution may be the result of a historical pattern of reduplication; more research is needed to confirm this.

(2.4) (ŋ)V_{α}ŋgV_{α}X:

a. *nginggisi* [ŋiŋgisi] 'cricket'

b. *anggadi* [aŋgadi] 'coconut'

(2.5) VNVŋgVX:

a. *amunggeri* [amuŋgeri] 'sago grub'

b. *anangganai* [anaŋganai] 'bait'

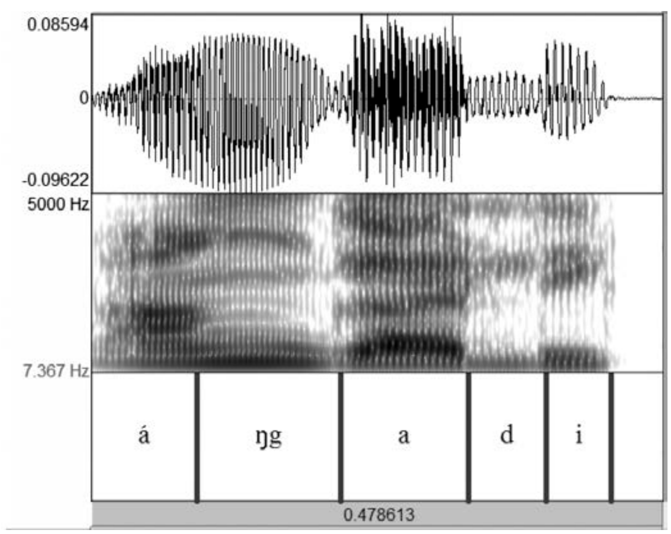


Figure 2.4: Intervocalic /ŋg/ Cluster

Velar nasals are overall rare in the language. Henning et al.'s (1991) dictionary of the Wondama dialect lists only seven /ŋ/-initial roots, compared to 39 /n/-initial and roughly

160 /m/-initial roots. It is even rarer in the Windesi dialect, where most instances of [ŋg] in Wondama appear instead as [k], as in Won. [ŋaŋgomi], Win. [kakomi] ‘nutmeg’. The segments /m/ and /n/ are each ten times as frequent in my dictionary as /ŋ/ is, each making up 10% of total consonant tokens to /ŋ/’s 1%.

Fricatives

Wamesa has two phonemic fricatives, bilabial /β/ and alveolar /s/. Both of these can appear word-initially and intervocalically; with one marginal exception, only /s/ appears word-finally.

The voiced bilabial fricative /β/ has an extremely variable realization, surfacing also as [w], [v], and occasionally [b], all potentially between tokens of the same word. Many older sources, such as van Balen (1915a) and Cowan (1955), are more or less consistent in transcribing it as *b*, failing to distinguish it from the voiced stop /b/ with which it contrasts. In my recordings it is more often resembles [w]; there are a number of words which are transcribed with [w] in my early notes which, when my ear became more acclimated to the sound of Wamesa, I later recognized as containing [β] instead. This is particularly true intervocalically and in fast speech, when target undershoot is more likely; in order to produce the required turbulent airstream, fricatives require a more precise positioning of the articulators than, for example, stops (Ladefoged & Maddieson 1996: 137), and in fast speech this target is more likely to be missed with audible consequences (see also e.g. Lavoie 2001).

The Rhotic

Wamesa has a single rhotic, the alveolar /r/. This segment is realized alternately as a tap or a trill, with the two possibilities in free variation. It is the most frequent consonant in the language, comprising 23% of non-glide consonant tokens in the 954 entries so far included in my dictionary. It can appear word-initially, as in *ravinie* ‘evening’, intervocalically, as in *mamara* ‘clear, clean’, and finally, as in *wamar* ‘Papuan (Blythe’s) hornbill’.

Laterals and Affricates

There are no laterals or affricates in the native Wamesa vocabulary, but /l/ and /dʒ/ do appear in some loanwords, such as *apel* ‘apple’ and *gaja* ‘elephant’. In established loans, /l/ is nativized to /r/ and /dʒ/ to /di/. See §2.4 for further discussion of loanword phonology.

2.1.2. Vowels

Figure 2.5 gives the Wamesa vowel phonemes. Wamesa has a symmetrical five-vowel system. Vowel length is not phonemic, nor is nasalization. Minimal and near-minimal pairs are presented in (2.6).

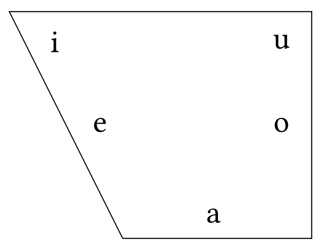


Figure 2.5: The Wamesa Vowel Inventory

(2.6)	a	/ra/	*ra*	‘go’
	e	/re/	*re*	‘eye’
	i	/ri/	*ri*	‘traditional dance’
	o	/ron/	*ron*	‘ironwood’
	u	/ru/	*ru*	‘head’

The low vowel /a/ is by far the most frequent vowel by lexical type; it comprises 23% of phoneme tokens in the dictionary and 43% of vowel tokens. The next most common vowel is /i/, comprising 23% of vowel tokens, followed by /u/ and /e/ at 12% and /o/ at 9%. These counts do include some compounds but do not include affixed forms. Preliminary statistical analysis shows little evidence for any covert vowel harmony in the lexicon.

Figure 2.6 gives spectrograms for each of the five Wamesa vowels, taken from interconsonantal stressed position.

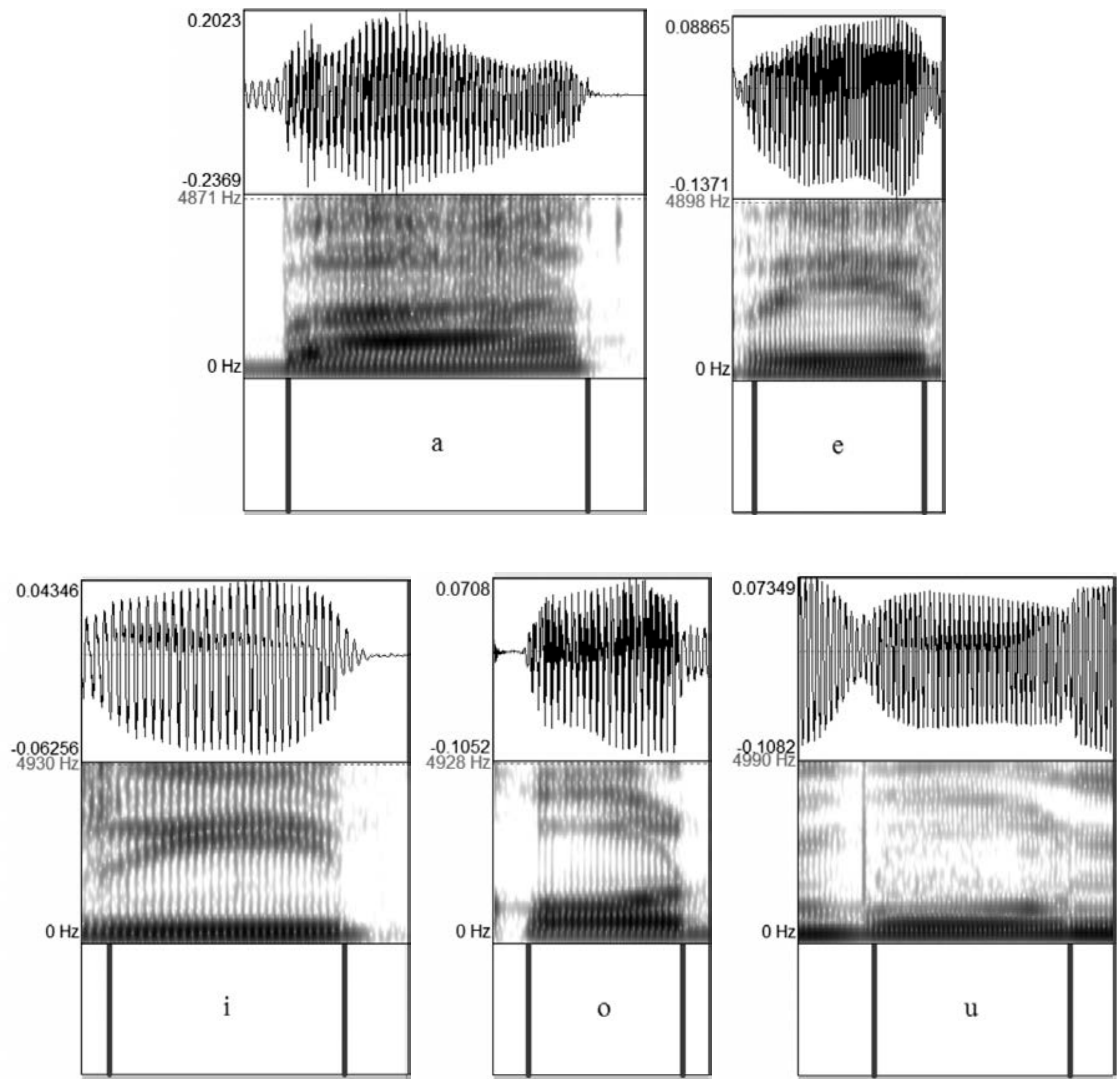


Figure 2.6: Wamesa Vowels

Figure (2.7) plots the F1 and F2 vowel means and 1 standard deviation from the mean of the vowels as pronounced in stress-bearing position. Measurements were taken from the midpoint of each vowel token as produced during a frog story narration by a single speaker (IMK) using Praat (Boersma & Weenink 2013) and plotted using NORM (Thomas & Kendall 2007). Ten tokens were measured of each vowel. Each of the vowels measured here is in stress-bearing position; the vowels are somewhat centralized when unstressed.

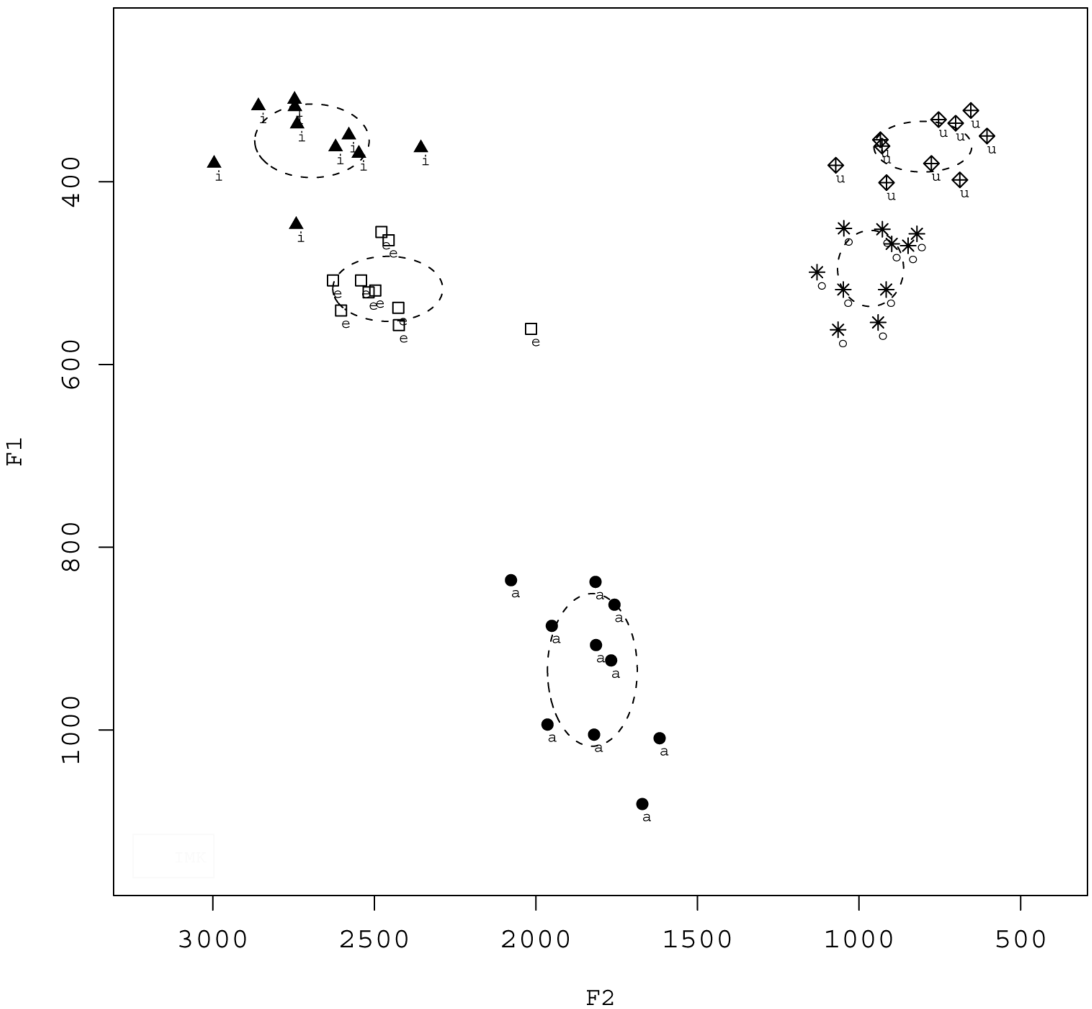


Figure 2.7: Vowel Plot

2.1.3. Orthography

Wamesa orthography is based on that used for Indonesian. In most cases, the orthographic representation of a segment is the same as its IPA symbol. The exceptions are the use of orthographic *v* for [β], *y* for [j], *j* for [dʒ], and the digraph *ng* for [ŋ]. Following Henning et al. (1991), I will write underlying high vowels as glides word-initially and intervocalically, where they obligatorily surface as such, but not in environments where reduction is optional (see §2.3.6).

2.2. Phonotactics

2.2.1. CC sequences

Consonant clusters in Wamesa are limited to heterosyllabic homorganic NC pairs, in which the second member must be a voiced stop, and consonant-glide or glide-consonant sequences. Only the NC clusters occur underlyingly, and only they are created by the lexical phonology of the language, though post-lexical or phonetic processes do create consonant + glide sequences. All surface glides are the result of reduction of an underlying high vowel, and where this takes place adjacent to a tautomorphemic consonant it is a highly variable, post-lexical process. See §2.3.6 for discussion. NC clusters never occur initially or finally in a word; they must span a syllable boundary. Other CC clusters, where they are created by the morphology, are simplified by deleting the first C of the pair. See §2.3.2 for details.

2.2.2. Vowel Sequences

Wamesa allows sequences of two or more adjacent vowels, as in roots such as *kiai* 'digit', *ariou* 'flower', *v<i>ui* '3sg writes', and *awawai* 'patrol', from underlying [auauai] with intervocalic high vowel reduction. In most pairs of adjacent vowels, at least one is [+high].

The only mid + mid vowel sequences attested in the language occur occross a morpheme boundary, created by adding the essive *ve-* prefix to an *o*-initial root, which are rare.[3] Only one /ae/ sequence appears in my data, in *adia kaesa* 'coals', and no /ao/ sequences do. Sequences of [e] plus [u] are also banned, and surface instead as [iu] when created by the morphology.[4] Because of these restrictions, longer sequences of vowels always consist of high vowels alternating with vowels of any height. No high-mid-low sequences are attested in my data, though this may be a statistical accident, as they do not violate any of the above pairwise phonotactic constraints.

2.2.3. Syllable Structure and Word Shape

Syllables in Wamesa consist minimally of a single vowel, and maximally include one onset and one coda segment. The Wamesa syllable template is given in (2.7). One example of each syllable type is given in (2.8); more extensive examples follow in (2.12).

(2.7) (C)V(C)

(2.8) V: *i* '3rd person plural pronoun'

CV: *mu* 'k.o. small anchovy'

VC: *at* 'four'

CVC: *sis* 'cloth

Consonant clusters must span a syllable boundary and therefore never occur word-initially or -finally; complex onsets and codas are not allowed. This may not be universally true across dialects. Saggers (1979: 9) lists three Wondama forms beginning in /mb-/: *mbot* 'round, circle', *mborov* 'celebrate traditionally', and *mba* 'west'. Saggers may have misinterpreted agressive pre-voicing of an initial [b] on these forms as a nasal, though

3. Only eight /o/-initial roots appear in my data, one of which, *onta* 'camel', is a Malay loan.

4. See §2.3.1 for discussion.

she includes a number of correctly-transcribed *b*-initial forms in her examples. Henning et al. (1991), also describing the Wondama dialect, does not include any of these forms; 'circle' is given as *wawarira*, while the other two meanings are omitted entirely from the vocabulary. They list a single cluster-initial form, *mbekua* 'large wild bat'. When the three forms listed by Saggers were presented to speaker IMK, she accepted *mborov* and *mba*, but alternated between pronouncing the former as [mboroβ] and [βoroβ], and accepted *mba* only in the context of 'west wind' rather than 'west' more generally. If such forms are present in Wamesa they are probably derived from *β*-initial forms (see §2.3.3 for β/mb alternations), and are at best marginal in the language.

Any consonant may appear in the syllable onset, and any segment but /g/ may appear word-initially.

(2.9) *Segments in Onset Position*

/p/:	[pipi]	'money'
/t/:	[titiet]	'bitter'
/k/:	[kakuomi]	'nutmeg fruit'
/b/:	[babua]	'roof'
/d/:	[dire]	'edge'
/g/:	[biaŋgaina]	'large sea turtle'
/m/:	[mamei]	'fish hook'
/n/:	[niniai]	'floor'
/ŋ/:	[ŋaŋgate]	'bother'
/β/:	[βaβi]	'woman'
/s/:	[sasi]	'salt'
/r/:	[ruru]	'lake'

Vowel-initial roots, by contrast, make up just under a quarter of the recorded lexicon. Any vowel may begin a root, but just under three-fourths of V-initial roots begin with /a/.

(2.10) *Word-Initial Vowels*

/a/:	[amoi]	'aunt'
/e/:	[eŋgerek]	'morinda tree'
/i/:	[imboni]	'beetle'
/o/:	[okatuma]	'momentarily'
/u/:	[uβi]	'yam'

The only licit clusters in Wamesa consist of a nasal followed by a homorganic voiced stop, which means that word-internally, only nasals may occupy coda position, a common restriction cross-linguistically. Word-finally, however, coda segments will not create a cluster,[5] and this restriction is relaxed. While vowel-final words far outnumber consonant-final ones - just over 10% of the words in my dictionary are C-final, and roughly 10% of those are identified as loans - words are attested in my data with final /β, k, m, n, ŋ, p, r, s, t/, plus one place name, *Yob*, ending in /b/. The only instance of final /β/ comes from the word *mborov*, just discussed. The only members of the Wamesa phoneme inventory which do not appear word-finally are /g/, which only surfaces following /ŋ/ in native words, /ŋ/ itself, and /d/. Saggers also notes the absence of final /d/ in her data, though, as she says, this may be a statistical accident rather than a true restriction in the grammar.

5. Consonant sequences may be formed across a word boundary, which is allowed, though see §2.3.7 for on strategy to avoid even these.

(2.11) *Root-Final Consonants*

/p/:	[komap]	'amaranth'
/t/:	[suomabut]	'large edible forest rat'
/k/:	[komok]	'cheap, not good'
/m/:	[tenam]	'live'
/n/:	[sawan]	'husband'
/β/:	[mboroβ]	'celebrate traditionally'
/s/:	[mas]	'hot'
/r/:	[mor]	'seed'

The most frequent syllable type in Wamesa is CV, though CVC syllables do exist, as described above. Long sequences of vowels can occur underlyingly, as in /auauai/ 'to patrol', but intervocalic high vowels will always reduce to a glide,[6] yielding CVCV strings in most of these cases - /auauai/ surfaces as [awawai].[7] The longest vowel sequences which surface intact in my data are three segments long, both in monomorphemic words such as *nioi* 'knife' and *niau* 'cat', and as created by infixation of the subject agreement marker, as in *viui* /β<i>ui/ '3sg writes'. The forms in (2.12) show examples of the four licit syllable types in monomorphemic words.

(2.12) a. V:

[**i**] '3sg pronoun'

[**a**.ri] 'church'

[ra.ri.**a**] 'day'

[mi.mi.**o**.ta.ri] 'snakehead fish'

6. See §2.3.6.

7. This form is never actually pronounced in isolation; it must occur with a verbal agreement prefix, yielding for example [jawawai] '1sg patrols'.

b. CV:

[**nu**] ‘island’

[**ba**.**ta**] ‘large wave’

[a.**pa**.**ra**.**pi**.**ri**] ‘gnat’

c. VC:

[**at**] ‘four’

[ti.ti.**et**] ‘bitter’

d. CVC:

[**ron**] ‘ironwood tree’

[**waŋ**.**gar**] ‘rat’

[ko.**rom**.bo.wi] ‘cowrie shell’

The majority of Wamesa roots are disyllabic. Monosyllabic words do exist, as do longer forms, as seen above. The longest attested monomorphemic words in my data are five syllables long. When verbal morphology is added, three- and four-syllable words become very common. Some examples are given in (2.13).

(2.13) a. *nu* ‘island’

b. *anda* ‘crown of thorns sea star’

c. *papano* ‘soft coral’

d. *ikerat* /i-kerat/ ‘1sg screams’

e. *paramera* ‘New Guinea rosewood’

f. *ikasio* /i-kasio/ ‘1sg is angry’

g. *piderekari* ‘mango’

h. *siverawana* /si-ve-rawana/ ‘they (non-human) are blue’

i. *diaŋgariria* /dia-C-kariria/ ‘crocodile’

j. *setikaviora* /set-it-kavio=ra/ 'they use (it) to talk to over there'

k. *amambekomamasare* /ama-ve-komamasare/ 'we (excl) are being funny'

2.3. Phonological and Post-Lexical Processes

Wamesa has a number of phonological and post-lexical processes which cause the underlying and surface forms of its words to differ. The following section discusses the raising of /e/ to [i] before /u/; derived cluster simplification; neutralization of /β/, /r/, and /k/ to NC sequences in clusters; deletion of initial-syllable /a/ in verbs; reduplication, which unlike the others has limited productivity; the reduction of high vowels to glides, which may happen both morpheme-internally and across morpheme boundaries, and both obligatorily and as an optional post-lexical process, depending on the environment; the use of optional paragogic *-e*; and reductions in CC and VV sequences in fast and casual speech. These are the processes which involve full deletion or epenthesis, or cause allophones of the underlying segments to surface which take the form of other, contrasting segments, i.e. [i] for /e/ or /mb/ for /β/, rather than non-contrastive phonetic variations of the same sound. The one exception is high vowel reduction, which affects the syllable structure of the surface form of the word.

2.3.1. Mid-Vowel Raising

The sequence [eu] never surfaces in Wamesa. Other mid + high vowel combinations are allowed in the language; [ei], [oi], and [ou] are all robustly attested both within individual morphemes and across morpheme boundaries. Underlying /eu/ sequences are created when the essive prefix *ve-* is added to an *u*-initial root such as *urar* 'red'. The result of this combination is that the mid vowel /e/ raises to be realized as [i].

(2.14) a. /β**e**-**u**rar/ ⟶ [β**iu**rar] 'which is red'

b. /β**e**-**u**nu/ ⟶ [β**iu**nu] 'who drinks'

This raising of /e/ to [i] creates homophony between forms with both the 3rd-person singular agreement affix and the essive *ve-*, and those with only *ve-*. In the former case, the agreement marker surfaces as [-i-], infixed after the initial [β] of the essive. The /e/ of the essive raises to [i] adjacent to root-initial /u/, and the two resulting adjacent [i]s merge into a single vowel. Compare the forms in (2.15) below to those above in (2.14).

(2.15) a. /β<**i**>**e**-**u**rar/ ⟶ [β**iu**rar] ‘3sg is red’

b. /β<**i**>**e**-**u**nu/ ⟶ [β**iu**nu] ‘a drink; that which he drinks’

The mirror-image restriction does not hold. There are examples attested in my corpus of [ue] sequences both tautomorphemically, as in (2.16a), and across a morpheme boundary. The latter occur when the 2nd-person singular agreement affix is added to a root whose first vowel is /e/, whether prefixed to an /e/-initial root as in (2.16b) or infixed to a C-initial one with /e/ in the first syllable nucleus, as in (2.16c).

(2.16) a. [s**ue**pe] ‘cave’

[rau b**ue**ma] ‘edible hibiscus leaf’

[kat**ue**rinei] ‘just now’

b. /b**u**-**e**na/ ⟶ [b**ue**na] ‘2sg sleeps’

c. /t<**u**>**e**nam/ ⟶ [t**ue**nam] ‘2sg lives’

2.3.2. Cluster Simplification

As noted above, Wamesa disallows most consonant clusters; the only acceptable CC sequences consist of a nasal followed by a homorganic voiced stop. In most cases, illegal clusters created at a morpheme boundary are simplified through deletion. Exceptions to this generalization, where the second member of the underlying cluster is /β/, /r/, or /k/, are discussed in §2.3.3. In all other cases, the first of the two adjacent consonants is deleted. Wilson (2000); McCarthy (2008); Steriade (2009), and others have posited that deletion of the first member of a CC sequence rather than the second is universal in cluster reduction, and attempted to explain why; in Wamesa, because these clusters are formed exclusively over a prefix-root boundary,[8] it is enough to invoke positional faithfulness (Beckman 1998) and say that the root-initial segment is preserved at the expense of the affix-final one.

Examples of this are given in (2.17) - (2.19). The underlying form of the affix, with the final /t/ or /r/, can be seen on vowel-initial roots, shown here for contrast using the verb *awer* 'hunt'. Deletion of the affix-final consonant can be seen on C-initial roots, here exemplified by *samuai* 'get, gather, collect'.

(2.17) a. /su**r**-awer/ ⟶ [su**r**awer] '3du hunt'

b. /su**r**-**s**amuai/ ⟶ [su**s**amuai] '3du gather'

(2.18) a. /ta**t**-awer/ ⟶ [ta**t**awer] '1pl.incl hunt'

b. /ta**t**-**s**amuai/ ⟶ [ta**s**amuai] '1pl.incl gather'

(2.19) a. /bu-i**t**-awer/ ⟶ [bui**t**awer] '2sg use (it) to hunt'

b. /bu-i**t**-**s**amuai/ ⟶ [bui**s**amuai] '2sg use (it) to gather'

This process can in certain cases lead to the apparent disappearance of a morpheme from the surface form of a word. When the applicative marker /it-/ is added to a consonant-initial verb root whose the initial C is other than /β/, /r/, /k/, the /t/ is deleted as the initial member of an illegal cluster. If the subject of the verb is 3rd person singular, the /i/ of the verbal agreement prefix will coalesce with the /i/ of the applicative. With a verb such as *mun* 'kill', the 3rd person singular applicative form surfaces as [dimun] '3sg-uses to kill', from underlying /di-it-mun/. The presence of the applicative can be deduced from the fact

8. It is plausible that such clusters could be formed between roots in a compound word; however no examples of a C-final root as the first member of a compound exist in my data. Because of the low frequency of both the *-si* suffix and C-final verbs, there are no identified instances in my data of both co-occuring.

that the agreement marker surfaces here as a prefix [di-] rather than an infix [-i-] as is usual with C-initial roots. Without the applicative, this verb would surface as [m<i>un] '3sg-kills'.

2.3.3. v/r/k Splitting

When a consonant cluster is formed over a morpheme boundary with /β/, /r/, or /k/ as the second member,[9] simplification does not occur as just described. Instead, the sequence surfaces as a homorganic NC cluster at the place of articulation of the second member. This occurs regardless of the identity of first member of the cluster, even if it is identical to the second member; /r-r/ sequences are common with subject agreement prefixes in the dual, and always surface as [nd]. Compare the examples in (2.20) - (2.22) to those in (2.17) - (2.19) above.

(2.20) a. /su**r**-**β**o/ ⟶ [su**mb**o] '3du paddle'

b. /su**r**-**r**a/ ⟶ [su**nd**a] '3du go'

c. /su**r**-**k**utu/ ⟶ [su**ŋg**utu] '3du cut'

(2.21) a. /ta**t**-**β**o/ ⟶ [ta**mb**o] '1pl.incl paddle'

b. /ta**t**-**r**a/ ⟶ [ta**nd**a] '1pl.incl go'

c. /ta**t**-**k**utu/ ⟶ [ta**ŋg**utu] '1pl.incl cut'

(2.22) a. /bu-i**t**-**β**o/ ⟶ [bui**mb**o] '2sg use (it) to paddle'

b. /bu-i**t**-**r**a/ ⟶ [bui**nd**a] '2sg use (it) to go'

c. /bu-i**t**-**k**utu/ ⟶ [bui**ŋg**utu] '2sg use (it) to cut'

This same pattern is active throughout the phonology. In addition to occurring at an affix-root boundary, it also takes place between affixes, specifically with the essive prefix

9. Because /β/ is written as *v*, I refer to this as v/r/k splitting.

/βe-/. As shown in example (2.23), /βe-/ is realized as [mbe-] when preceded by a plural or dual agreement affix.

(2.23) a. /su**r**-**β**e-marisiani/ ⟶ [su**mb**emarisiani] '3du are spicy'

b. /ta**t**-**β**e-marisiani/ ⟶ [ta**mb**emarisiani] '1pl.incl are spicy'

There are a limited number of roots beginning with a voiced stop in the Wamesa lexicon, but two examples exist in my data in which a *b*- or *d*-initial root takes plural agreement morphology. Roots beginning with /g/ are impossible in the native lexicon. In these two cases, the preceding consonant nasalizes rather than deleting, leading to the same homorganic NC clusters seen with /β, r, k/.

(2.24) a. /se**t**-**b**aba/ ⟶ [se**mb**aba] '3pl.hum are big' [S76 IMK]

b. //se**t**-**d**eriasi/ ⟶ [se**nd**eriasi] '3pl.hum are close' [S56 IMK]

Splitting can also be seen in compounds. There is a compounding morpheme in Wamesa, similar to the semantically empty *-s-* and *-en-* interfixes seen in German compounds such as *Liebe-s-brief* 'love letter' and *Schwan-en-gesang* 'swansong' (Haspelmath & Sims 2010: 139; 191), which conveys no semantic content and whose distribution is unpredictable. The examples in (2.25) show a selection of forms in which the compounding morpheme is evident.

(2.25) a. /a-**C**-**v**esie/ ⟶ [a**mb**esie] 'delicious' (lit. 'eat well')

b. /dia-**C**-**k**ariria/ ⟶ [dia**ng**ariria] 'crocodile' (lit. 'evil fish')

c. ? /na-**C**-**r**au/ ⟶ [na**nd**au] 'thatched roof' (lit. '??-leaf')

d. /pi-**C**-**β**ara-**k**iai/ ⟶ [pi**mb**ara**k**iai] 'ring' (lit. 'hand digit thing')

Whether the compounding morpheme can synchronically be considered a separate morpheme or if it only appears in fossilized forms from an earlier stage of the language is unclear. The only evidence for its existence comes from some compounds whose second member begins with /β/, /r/, or /k/; in the above forms and others like them these surface having undergone splitting even though the preceding member of the compound is vowel-final. All that can be said about the shape of this morpheme is that it consists of a single consonant, as it does not appear in my data between vowel-final and vowel-initial compound elements where it might surface intact; if it occurs underlyingly between V-final and other C-initial items it deletes and no evidence remains in the surface form of it ever having existed.

There are no identified instances of the compounding morpheme occurring before /r/ in my corpus, except very tentatively as a possible source for the word *nandau* 'roof thatched from leaves', as *rau* in Wamesa means 'leaf'.[10] This form is given in (2.25c), though the identity and meaning of the putative first element of the compound is unknown. *Pimbarakiai* 'ring' in (2.25d) is notable in that splitting of /β/ gives evidence for the morpheme's presence between the first and second members of the compound, but not between the second and third, where the /k/ would be expected to surface as [ŋg].

Cowan (1955); Saggers (1979) and Henning et al. (1991) all describe v/r/k splitting as regular across the language, and for the most part that is true in my data as well. The one exception is with /k/: splitting of /k/ is irregular and unpredictable, varying between speakers, lexical items, and elicitation sessions. For example, in the verb paradigms elicited on 7/14/11 with speaker IMK, the verb *kubira* 'to bathe' appears unsplit in the 3rd person plural as [se**k**ubira] rather than [se**ŋg**ubira], and similarly throughout the rest of the paradigm. Another speaker, BAK, in a later elicitation reliably produced the form with splitting as [se**ŋg**ubira]. The verb *kavio* 'speak' was produced as [se**ŋg**avio] in the

10. This analysis is somewhat complicated by the fact that the word for 'hair' in Wamesa is *ru nandau*, literally 'head thatching'; the metaphor of using a phrase like 'head leaf' for hair is reconstructable at least to Proto-Central-Eastern Malayo-Polynesian (Blust 1993a), so this may be inherited rather than an instance of the synchronic alternation.

3rd person plural by both of these speakers. This may be a symptom of language endangerment; as Wamesa is spoken less and less in daily life its less-regular patterns are falling out of use. Splitting of /k/ may be particularly vulnerable compared to the other segments, as the resulting [ŋg] cluster is far more limited in its distribution than the analogous [mb] and [nd] clusters. If /k/ is removed from the list of splitting consonants, the remaining two do form a natural class, comprised of all of the Wamesa voiced continuants, perhaps contributing some additional pressure in support of the change.

Why these three segments should undergo splitting is puzzling; this appears to be a 'crazy rule' in the sense of Bach & Harms (1972). The voiced bilabial fricative, alveolar tap/trill, and voiceless velar stop do not form any sort of natural class; the most specific feature shared by all three is their status as obstruents, a class which includes a number of other non-splitting sounds in Wamesa. There is no obvious phonological reason why any process should target these three segments to the exclusion of all others, or why any of these in particular should trigger splitting rather than consonant deletion as occurs in all other derived clusters. Particularly puzzling from a synchronic point of view is that /rr/ clusters should surface as [nd], as in example (2.20b) above, rather than simply [r]; other instances of derived geminates, as when a *t*-final plural agreement prefix attaches to a *t*-initial stem, simplify into singletons under the normal rules of cluster reduction. There are no instances in my data of underlying /kk/ or /ββ/ clusters, which would only arise via compounding, but presumably they would surface as [ŋg] and [mb] respectively.

Neither is there any apparent phonetic motivation for splitting. The set /β, r, k/ are just as difficult to link phonetically as they are phonologically. One feature which they may all have in common is the involvement of a dorsal component; /k/ is primarily dorsal, /r/ has been shown to involve a secondary dorsal gesture in many languages (Browman & Goldstein 1995; Gick 2003; Proctor 2009), and the bilabial status of /β/ means that any dorsal component would not be interfered with by the primary articulatory gesture. However, no articulatory studies have yet been done on Wamesa to confirm or disprove the presence of a dorsal component in the latter two segments, and even if it were present the question arises of why other bilabial sounds, such as /w/, /m/, /p/, and /b/, do not participate in splitting. And, as with the phonological aspect, there is no clear phonetic reason why any of these should be realized as NC clusters rather than triggering deletion in clusters.

This leaves us with a historical motivation for splitting. It is clear from looking at other related languages that the roots of this process arose in previous stages of the language, at least as far back as the common ancestor of the Biakic and Yapen groups. Ambai, one of Wamesa's closest relatives, has a process resembling that found in Wamesa, where /p/, /w/, and /r/ become the stops [p, b, d] following a nasal, while /n/ assimilates to the place of articulation of a following consonant (Silzer 1983: 51-52). Biak is far more permissive than Wamesa in terms of its consonant clusters,[11] but it too has restrictions regarding /β, r, k/ after nasals: Biak /mr/ surfaces as [mbr], /nr/ as /nd/ or /ndr/ depending on lexical class, /Nβ/ as [Nb], and /Nk/ as /ŋk/ (van den Heuvel 2006: 54-57). Sequences of /rr/ over a word boundary also surface variably as a geminate [rː] or as [rd]. These patterns are not by any means identical to that found in Wamesa, but they share enough similarities to suggest a common origin somewhere within the West New Guinea or Cenderawasih Bay subgroups.

As with many other synchronically "crazy" patterns, I would tentatively suggest that Wamesa v/r/k splitting is the result of rule inversion (Vennemann 1972; McCarthy 1991; Blevins 1997; Garrett & Blevins 2009). This is what happens when a diachronic change takes place turning phoneme A to phoneme B in environment X, which is a subset of all of the environments in which A occurs. If environment X encompasses the majority of the

11. See van den Heuvel (2006: 37-43) for the allowable combinations, many of which arose from the deletion of /a/ in the first syllable of certain words in Proto-Biakic.

instances of A, or the more basic ones - Vennemann suggests for example all singular but no plural noun forms, or all present indicative verb forms but not past or subjunctive ones - B may be reanalyzed by speakers as the underlying form, changing to A in the complement of the environments described by X. In the case of Wamesa, this would look something like a scenario in which Proto-Eastern Malayo-Polynesian /*b, *d, *g/ became /β, r, k/ in all environments except after nasals (or perhaps all Cs) in Proto-Cenderawasih Bay. If this change applied across morpheme boundaries as well as morpheme-internally, it would lead to roots beginning with [β, r, k] word-initially and post-vocalically, but with the earlier [b, d, g] retained post-consonantally, where the change was blocked. If nasalization of the first member of this cluster followed, the result is exactly the alternation we see in the synchronic language. Slightly different paths of development would lead to the more-complex patterns found in Biak, Ambai, etc.

Evidence for this is suggestive but not yet conclusive. A comparison of 210-item wordlists downloaded from the Austronesian Basic Vocabulary Database (Greenhill et al. 2008) for Proto-Central-Eastern Malayo-Polynesian (PCEMP) and Wamesa[12] reveals that Wamesa /β/ does indeed derive from PCEMP **b* as well as and non-initial **p*;[13] WAD /r/ is traceable back to PCEMP **l*, initial **d*, and **ʀ*. The origin of WAD /k/ is less clear, as there are fewer remaining cognate forms, but descent from PCEMP **k*, at least in initial position, appears likely. Surviving cognates are rare enough that it is impossible to determine from this dataset the origin of WAD /d/ and /g/. Modern /b/ appears to descend from PCEMP **b*. The two roots used above in (2.24), *baba* 'big' and *deriasi* 'close (to)', are both clearly more recent additions to the lexicon; the PCEMP forms are listed as **ʀaya* for 'big' and **hazani* and **raŋi* for 'near'. Examples of forms undergoing the above-described sound

12. The PCEMP data comes originally from Blust (1993a); the Wamesa data comes from unspecified materials collected by van Balen and Anceaux, as well as my field notes.

13. Initial **p* reliably deletes, as in PMP **piliq*, WAD *iri* 'choose'.

changes are given in (2.26).

(2.26) *PCEMP/WAD Correspondences*[14]

Reflex	PCEMP	WAD	Gloss
/β/	**qa**p**uk*	*wa**v**u*	'dust'
	***b**a**b**aw*	***v**a**v**a*	'above'
/r/	*bə**l**i*	*vo**r**i*	'buy'
	***d**aun*	***r**au*	'leaf'
	ta⊠aq	*ta**r**a*	'cut, hack'
/k/	***k**aʀaw*	***k**ara*	'scratch'
	***k**udug*	***k**uruya*	'thunder'
/b/	***b**uaq*	***b**uo*	'fruit'
	***b**aʀəq*	***b**ar**b**ara*	'swell'

Based on the other languages of the West New Guinea subgroup for which wordlists are available on the ABVD,[15] modern /r/ had arisen by the time of Proto-Cenderawasih Bay - it appears in the CB languages but not those of Raja Ampat. The same is true of modern /k/. WAD /β/ is cognate with /w/, /β/, and /b/ in other Cenderawasih Bay languages according to these lists, but what that actually means is unclear; in many Wamesa sources, including those used to compile the ABVD, /β/ is often written unpredictably as either *v* or *b*; what sounds those symbols actually represent in a given word of in Ambai, Numfor, etc., and how consistent the original transcribers were, is impossible to determine from this data.

14. The ABVD gives the form for 'dust' as *wabu*, while it appears as *wavu* in my data and in Henning et al.'s (1991) dictionary; this is a case of early sources mis-transcribing /β/ as *b*. The forms *tara* 'cut, hack' and *barbara* 'swell' are not attested in my data but are present in Henning et al.. *Kara* 'scratch' is given in the ABVD list but is not in my data or in Henning et al..

15. Yapen branch: Ambai, Marau; Biakic: Numfor; other Cenderawasih Bay: Mor, Waropen; Raja Ampat: As, Biga, Minyaifuin

What may have happened was that earlier /*b, *d, *g/ became /β, r, k/ word-initially in certain environments. If that change failed to happen after a prefix-final consonant, it would produce alternations with the newly-/β, r, k/-initial stems. It is possible that, at the alveolar place of articulation, splitting was initially restricted to forms derived from earlier *d and later spread by analogy to those descended from *l and *ʀ. Again, this data is suggestive but not conclusive. Further research, likely involving further fieldwork on related languages, is required to definitively discern the path of change which led to Wamesa v/r/k splitting. Until that is possible, rule inversion seems the most likely explanation.

2.3.4. Verbal [a] Deletion

There are two environments in which the first vowel of a verb root may fail to surface. As discussed in §2.3.6, when the root vowel is adjacent to an identical affixal vowel, only one surfaces, as Wamesa does not have long vowels.

The second vowel deletion environment is found in verbs with an unstressed /a/ in the initial syllable. In these verbs, the /a/ deletes when adjacent to the /u/ or /i/ of the second- or third-person singular subject agreement affix. This can occur both when the /a/ is root-initial, adjacent to the agreement prefix, or in C-initial roots with the agreement infix.

(2.27) a. /b**u**-**a**diáwa/ ⟶ [b**u**diawa] ‘2sg hears’

b. /d**i**-**a**diawa/ ⟶ [d**i**diáwa] ‘3sg hears’

(2.28) a. /m<**u**>**a**tai/ ⟶ [m**u**tái] ‘2sg is afraid’

b. /m<**i**>**a**tai/ ⟶ [m**i**tái] ‘3sg is afraid’

Unstressed /a/ will fail to delete in verb roots of the shape (C)a, such as *ra* ‘go’, *sa* ‘ascend’, and *a* ‘eat’. In all of these cases, the affixal vowel bears stress, but the root vowel remains present in the surface realization.

(2.29) a. /d**i**-**a**/ ⟶ [d**ía**] '3sg eats'

b. /k<**i**>**a**/ ⟶ [k**ía**] '3sg throws'

Stressed /a/ never deletes in this environment. Secondary stress is not sufficient to stop *a*-deletion; verbs with secondary stress on the [a] with other person/number combinations do show deletion in the 2nd- and 3rd-person singular. See Chapter 3 for a discussion of stress assignment and underlying stress diacritics.

(2.30) a. /t<**i**>**á**u/ ⟶ [t**iá**u] '3sg falls'

b. /d**i**-**á**pe/ ⟶ [d**iá**pe] '3sg reads'

(2.31) a. /k<**i**>**a**rakuái/ ⟶ [k**ì**rakuái] '3sg is strong' (cf [ikàrakuái] '1sg is strong')

This data contradicts that reported by Saggers (1979) and Henning et al. (1991). Saggers reports that in words longer than two syllables, /a/ raises to [e] after the same affixal /u/ and /i/ that cause deletion in my data. Henning et al. report two lexically-determined classes of words, one of which undergoes /a/-to-[e] raising in these environments and one of which leaves the /a/ unchanged. This appears to be a dialectal difference between the Windesi and Wondama varieties of Wamesa.

A related pattern is also found elsewhere in the family. In Ambai, a more complicated interaction between affixal and stem vowels takes place, which, depending on the stem shape, can result in deletion of either or neither vowel, or assimilation and coalescence. As in Wamesa, the second- and third-person singular agreement affixes are /bu-/ ~ /-u-/ and /di-/ ~ /-i-/ respectively, though they do sometimes interact with the root-initial consonant as well. Unstressed /a/ will coalesce with the affixal vowel to surface as mid when the vowel of the following syllable is high, and unstressed initial-syllable vowels delete entirely in this environment when they agree in height with the vowel of the following syllable (Silzer 1983: 152-157). The table in (2.32) gives examples of each of these changes, taken from Silzer; see the above-cited pages for his explanation of these changes, including consonant mutations.

(2.32) *Ambai Vowel Interactions*

a) No change:

/di-baur/	[b<i>aur]	'3sg splits'
/di-narir/	[n<i>arir]	'3sg makes'

c) Loss of affixal V:

/bu-ena/	[bena]	'2sg sleeps'
/di-tanam/	[sanam]	'3sg plants'

b) Mutual assimilation:

/bu-sansun/	[wonsun]	'2sg is clothed'
/di-kasou/	[kesou]	'3sg is angry'

d) Stem vowel deletion:

/bu-matai/	[mutai]	'2sg is afraid'
/di-matai/	[mitai]	'3sg is afraid'

2.3.5. Reduplication

Reduplication is extremely widespread and often highly productive throughout Austronesian (Klamer 2002a). Biak, for example, has an extensive and complex set of full and partial reduplicative processes, with iterative, durative, nominalizing, and other functions (van den Heuvel 2006: §7). Wamesa is somewhat typologically unusual for the family in that its reduplication has limited productivity. Of the nearly 1000 entries added so far to my Wamesa dictionary, only eight have been identified as being synchronically reduplicable.[16] All of my attempts to apply reduplication to words other than those already demonstrated by speakers were rejected as ungrammatical under any reading.

Reduplication in Wamesa is partial, not total, and copies the first or second syllable of the root. It has an intensifying effect, for example changing *masabu* 'broken, cracked' to *masasabu* 'smashed, shattered' and *kasio* 'angry' into *kasisio* 'furious'. Reduplication is attested on adjectives, verbs, and adverbs. The examples in (2.33) - (2.35) give instances of reduplication as applied to these three word classes. Example (2.36) shows ungrammatical reduplication of the verb *pera* 'cut', attempted by me and rejected by IMK.

(2.33) a. *I-ne sasu=ne-i m<i>**ra**ba.*
1sg-have clothing=det-sg 3sg-ripped
'My clothing is ripped.' [S19 IMK]

b. *I-ne sasu=ne-i m<i>**rara**ba*
1sg-have clothing=det-sg 3sg-redup-ripped
'My clothing is ripped to shreds.' [S19 IMK]

(2.34) a. *K<i>**vi**o.*
3sg-talk.
'He talks'

b. *K<i>**vivi**o.*
3sg-redup-talk.
'He chatters on'

(2.35) a. *I-ra **sa**ira.*
1sg-go quickly
'I go quickly'

b. *I-ra **sasa**ira.*
1sg-go redup-quickly
'I go very quickly'

16. An additional two bird names, *kowokowo* and *kumukumu*, are clearly reduplicated but not synchronically decomposable.

(2.36) a. *Yau i-**ma**se.*
I 1sg-hot
'I'm hot.' [S31 IMK]

b. * *Yau i-**mama**se.*
I 1sg-redup-hot
'I'm very hot.' [S31 IMK]

2.3.6. High Vowel Reduction

The high vowels /i/ and /u/ alternate with the glides [j] and [w] respectively, surfacing in reduced form in non-stressed position when adjacent to another vowel. This reduction is obligatory intervocalically and word-initially before a vowel, when the resulting glide can fill a vacant onset position, but is optional when adjacent to a consonant or word-finally.

Ladefoged & Johnson (2011: 232-3) differentiate glides from their corresponding vowels on the basis of their ability to form a syllable nucleus. They define a semivowel or glide as a 'non-syllabic vocoid', where a vocoid is a sound with no obstruction of airflow, while vowels are 'syllabic vocoids'. Ladefoged & Maddieson (1996: 323) add that glides are produced with a narrower constriction in the vocal tract than their corresponding vowel. In production, this difference is instantiated as one of intensity and phonotactic possibilities: the glides are quieter than their full vowel counterparts, and only appear adjacent to a full vowel, whereas full vowels are relatively louder and may appear between two consonants or a consonant and a word boundary.

High vowels which are marked for stress (see Chapter 3) will never surface as glides even when in potentially reducible position adjacent to another vowel, as in [ˈdi.a] 'fish'. This can result in sequences of vowels where glides would otherwise be expected to surface. Forms such as *viúi* '3sg-write' will often surface with three consecutive high vowels, because the /u/, as a stressed vowel, cannot reduce, and neither instance of /i/ is intervocalic or word-initial, making reduction optional. I analyze the vowel allophone as underlying because in environments where the realization is variable, it is the full vowel which will appear in careful speech. In fast or casual speech, the segment is more likely to surface as a glide. Both pronunciations are acceptable, and a single speaker will use both interchangeably.

Obligatory Gliding

Reduction of high vowels to glides is obligatory when followed by another vowel and preceded by either a vowel or a word boundary - in other words, when the resulting glide alone constitutes the onset of the syllable headed by the following vowel. Stated in rule form:

$$\breve{V}_{[high]} \longrightarrow G \,/\, (\#|V)\, \_\, V.$$

Examples are given in (2.37).

(2.37) a. /**i**-ase/ ⟶ [**j**ase] ‘1sg-swim’

b. /ka**i**obi/ ⟶ [ka**j**obi] ‘cuttlefish’

c. /**u**is/ ⟶ [**w**is] ‘mountain’

d. /madia**u**i/ ⟶ [madia**w**i] ‘younger sibling’

The processes described by these two rules ensure that, wherever possible, Wamesa syllables have onsets. That is, they are evidence for an active Onset constraint (Itô 1986, 1989; Prince & Smolensky 1993/2004) in the language. Ranked above a basic faithfulness constraint, alongside Dep and Max, this gives us the attested surface forms for word-initial high vowels, as in (2.38a), and intervocalic ones, as in (2.38b).

(2.38) *Onset-Induced High Vowel Reduction* a.

/i-ase/			Dep	Max	Onset	Faith(vowel)
a.	☞	ja.se				*
b.		i.a.se			*!* W	L
c.		ʔi.a.se	*! W			L
d.		se		*!* W		L

b.

/kaiobi/			Dep	Max	Onset	Faith(vowel)
a.	☞	ka.jo.bi				*
b.		ka.i.o.bi			*!* W	L
c.		ka.ʔi.ʔo.bi	*!* W			L
d.		ka.bi		*!* W		L

High vowels in other environments, even when vowel-adjacent, do not undergo obligatory gliding.[17] Examples are given in (2.39), where high vowels in word-final position or adjacent to a consonant do not reduce.

(2.39) a. /b**u**-ase/ ⟶ [b**u**ase]

b. /mad**i**au**i**/⟶ [mad**i**aw**i**] 'younger sibling'

c. /kaβ**u**β**ui**/ ⟶ [kaβ**u**β**ui**] 'bamboo shoot'

d. /k**i**a**i**/ ⟶ [k**i**a**i**] 'finger, toe'

To account for this, we must include two additional constraints: *ComplexOnset and *Coda (Prince & Smolensky 1993/2004; McCarthy & Prince 1993/2001). These prevent

17. They may undergo optional post-lexical gliding, discussed in the following section.

vowel reduction when doing so would produce a consonant cluster consisting of C+glide or vice versa. The tableaux in (2.38) demonstrated that Onset must be ranked above the constraint enforcing faithfulness to the vowel; candidates (c) - (f) below demonstrate that all other constraints shown here must outrank Onset.

(2.40) *Non-Reduction of High Vowels in Other Positions*

/madiaui/	*Coda	*ComplexOns	Dep	Max	Onset	Faith(vowel)
a. ☞ ma.di.a.wi					*	*
b. ma.di.a.u.i					*!** W	L
c. ma.dja.wi		*! W			L	** W
d. mad.ja.wi	*! W				L	** W
e. ma.di.ʔa.ʔu.ʔi			*!**W		L	L
f. ma.di				*!** W	L	L

Optional Gliding

Unstressed high vowels which fall between a consonant and another vowel, in the environment C_V, or word-finally following a vowel, in the environment V_#, sometimes also reduce to a glide. This can occur both within a morpheme and across a morpheme boundary, leading to alternations such as [siβererei] ~ [siβererej] 'star' and [kiopa] ~ [kjopa] '3sg jumps', from /k<i>opa/. Optional gliding is variable, and is far more common in fast and casual speech than in slow, careful speech and citation forms. It is also gradient. There is no discrete boundary between the realizations of high vowels and glides; they simply fall at different points on a continuum of intensity. In cases of optionality the realization of these phonemes can be anywhere on that continuum, rather than falling unambiguously into one category or the other; segments whose status as vowel or glide is obligatorily defined by the phonology is much more clearly auditorily distinct.

Under an explicitly stratal theory such as Lexical Phonology (Kiparsky 1982), this version of gliding would be considerd a post-lexical rule or part of the phonetic implementation. Kiparsky (1985: 86) tentatively suggests that rules of exactly this type - gradient versions of categorical processes - should in fact be part of the post-lexical phonological grammar. More recently, Coetzee & Pater (2011) term this stage the "late phonology", which is categorized by non-categorical changes which are sensitive to factors such as speech rate and which takes as its input whole utterances.

Evidence that optional gliding must take place extremely late comes from its lack of interaction with stress placement. Whether a optionally reducible vowel is in fact reduced or not has no impact on the location of stress in the form. Because stress placement relies on foot structure (see Chapter 3), and gliding removes a syllable, we would expect there to be an interaction if this were a process of the lexical phonology. Furthermore, because stress placement is affected by clash arising between words, optional gliding must occur quite late for it to fail to have an effect on this. Depending on how many strata one is willing to admit into the model, optional gliding could be placed as a feature of either the (very) late post-lexical phonology or simply in the phonetic implementation. What is clear is that it does not occur as part of the lexical, or even phrasal, phonology.

Glide + High Vowel Sequences

There are no examples of monomorphemic [wu] or [yi] sequences in the Wamesa lexicon, which suggests that no underlying /uu/ or /ii/ sequences exist.[18] The addition of subject

18. It does not entail their absence, as underlying /uu/ and /ii/ in environments without obligatory gliding would simply surface as [u] and [i]. This is however unlikely, as Wamesa learners would have no reason to posit a double underlying vowel rather than a single one.

agreement affixes to a verb, predicative adjective, or inalienably possessed noun[19] will sometimes create a sequence of two identical underlying high vowels across a morpheme boundary. The realization of this sequence depends on the same deletion processes discussed above.

There exists no prefix in Wamesa which would create a /#uu/, /uuV/ or /Vuu/ sequence, putting the middle /u/ in a position of obligatory reduction to a glide, and therefore no instances of [wu] in the language. Instead, addition of a 2nd person singular agreement affix /bu-/ creates either a /bu-V/ sequence on vowel-initial roots or /C<u>V/ with infixation on consonant-initial roots. If the V of the root is an /u/, it deletes and only a single short [u] surfaces.

(2.41) a. /b**u**-**u**nu/ ⟶ [b**u**nu] '2sg-drink'

b. /r<**u**>**u**ti/ ⟶ [r**u**ti] '2sg-grasp'

Most instances of /i/ + /i/ sequences created over a morpheme boundary similarly surface as simply a short [i], as when the third person singular verbal agreement marker /di-/ is prefixed to an /i/-initial verb stem or the applicative prefix, or when its infixal allomorph [-i-], which appears with C-initial verb roots, appears on a form whose first vowel is /i/. The verb form in (2.42b), for example, can never surface as *[djitane], with glide formation rather than deletion of one of the /i/s, just as the forms in (2.41) above will never be realized as *[bwunu] or *[rwuti].

(2.42) a. /r<**i**>**i**na/ ⟶ [r**i**na] '3sg knows'

b. /d**i**-**i**t-ane/ ⟶ [d**i**tane] '3sg uses (it) to eat'

When an /ii/ sequence is formed word-initially, however, it will surface as [ji]. This occurs when the first person singular verbal agreement prefix /i-/ is attached to vowel-

19. See §4.4.

initial verb roots or the applicative prefix /it-/. Because no other affix may precede the verbal agreement prefixes, this derived /i-i/ sequence is always word-initial and therefore in an obligatory gliding environment. Rather than deleting, the prefixal /i/ reduces to /j/, creating surface [ji] sequences. Vowel-initial words normally have no on-glide.

(2.43) a. /i-iri/ ⟶ [jiri] 'I choose'

b. /i-it-ane/ ⟶ [jitane] 'I use (it) to eat'

This output follows straightforwardly from the constraints already laid out in (2.40) above.

(2.44) *Derivation of [ji] Sequences*

/i-iri/	*Coda	*ComplexOns	Dep	Max	Onset	Faith(vowel)
a. ☞ jiri						*
b. iri				*! W	* W	L
c. iʔiri			*! W		* W	L

2.3.7. Paragogic [e]

Wamesa does have C-final words, but they are dispreferred in the language. As such, a final [-e] can sporadically be added to underlyingly C-final words. In my data, appearance of this [-e] is entirely unpredictable, and unrelated to speech rate, speaker, position in the phrase or utterance, etc. In elicitation it is common for a speaker to repeat the citation form for a word several times both with and without paragogic -*e*, as in "*Suomuse. Suomus. Suomuse.*" for 'shark'. This is unlikely to be an instance of deletion or devoicing of an underlying final /e/, as many words, such as *kake* 'green', *apose* 'nutmeg', and *vape* 'but', do surface reliably with what I analyze as an underlying final [e]. Paragogic [e] can be said to attach to C-final roots; a devoicing analysis would be hard-pressed to define the appropriate environment.

Addition of this vowel both removes a *Coda violation and avoids the creation of clusters between words which, while tolerated in that environment, would be disallowed elsewhere. Epenthetic vowels cannot be used word-internally to improve syllable well-formedness or break up illegal clusters; paragogic *-e* is the only such example.

2.3.8. Reductions in Fast Speech

It is well known that in fast or casual speech, coarticulation of adjacent segments often increases, leading to reduction and/or assimilation of those segments. This is in many cases attributable to increased gestural overlap, in the sense of Browman & Goldstein (1989, 1992). There are three ways in which this tendency manifests itself in Wamesa: NC clusters may reduce to a single nasal, certain VV sequences coalesce into a single vowel, and unstressed vowel-adjacent high vowels may reduce to a glide. This third process was discussed above in §2.3.6; the first two will be described here.

The first pattern is straightforward: any of the three homorganic NC clusters present in Wamesa - /mb/, /nd/, and /ŋg/ - may merge in fast speech into a single segment [m], [n], or [ŋ]. Articulatorily, this can be attributed to a lag in the raising of the velum between the nasal and oral segments, leading to nasalization of both. A common example is given below in (2.45).

(2.45) /a**nd**i/ ⟶ [a**n**i] 'him/her'

The attested instances of vowel coalescence are rather more restricted; not every VV sequence may be simplified in fast speech. The two that can are /ua/, which are realized as [o], and /ai/, realized as [e]. In both cases the input sequence consists of one high and one low vowel. The output vowel is mid in height, and gets its backness value from the underlying high vowel. Examples are given in (2.46). Stressed vowels are less prone to reduction in this context, but not immune; in *muandu* below the stress falls on the /a/, which nonetheless merges with /u/.

(2.46) a. /βara-ki**ai**-dir/ ⟶ [βaraki**e**dir] 'fingernail'

b. /m**ua**ndu/ ⟶ [m**o**ndu] 'two'

These two types of reduction may take place individually or together. Thus the word *muandu* 'two' may surface variably as [muandu], [mondu], or [monu]. The fourth option, [muanu], is not attested in my recordings; whether this is due to an implicational relationship between the reduction processes or simply chance is unclear. In all cases the occurrence of reduction is highly variable and in no way obligatory.

It should be noted that based on my limited data from the Bintuni dialect, the two assimilation processes described in this section do appear to be lexicalized for some speakers of that variety, at least in a subset of the vocabulary. In my elicitation session with speaker BMK, *monu* was given as the careful pronunciation for 'two', in contrast with Windesi *muandu*, as well as [n] where Windesi has [nd] in the dual pronouns. Other examples of NC clusters, [ua], and [ai] sequences do appear in that wordlist, however, and speaker IAK gave non-reduced forms in my visit with her. Therefore assimilation cannot be a universal process in that dialect, and may be restricted to a limited area or group of villages rather than all Bintuni speakers.

2.4. Loanword Phonology

2.4.1. History of Language Contact

Wamesa shows a large number of loan words from Indonesian (BI) and Papuan Malay (PM), as well as from Dutch (D). All Wamesa speakers with whom I came into contact were also fluent speakers of Papuan Malay. Papuan Malay, which developed out of Malay varieties from Ambon and Sulawesi, has been present in coastal Papua at least since the establishment of the Dutch colonial apparatus in the 19th century (Donohue & Sawaki 2007), and less intense contact with those source varieties likely took place before that date by way of eastern Indonesian trade routes that moved spices, tortoiseshell, and slaves from Sulawesi in the west as far east as the Papuan Bird's Head and Biak (Klamer 2002b; Saragih 2012). Here the term 'Malay' is used to refer to the continuum of lects including Papuan Malay and local Indonesian, in which most daily communication takes place in urban Papua, as well as the Standard Indonesian encountered mainly through mass media. For the reasons outlined below, Malay of one form or another is considered to be the immediate source language for all loans into Wamesa discussed here unless stated otherwise. Original sources of the loan into Malay are given in parentheses.

The Dutch sent their first expedition to the East Indies in 1598, and first laid claim to West Papua 230 years later in 1828 (Tadmor 2007). The first Dutch missionary, J. A. van Balen, arrived in the village of Windesi in 1889. Most of the current generation of Wamesa adults do not speak Dutch, though some better-educated members of their parents' and grandparents' generations did.[20] Though many loans ultimately of Dutch origin appear in Wamesa, all but one of those attested so far in my data are present also in Papuan Malay (though not all in Standard Indonesian). While it can be difficult to determine based on phonological form whether originally Dutch loans into Wamesa were borrowed directly or whether they came via PM, the social and historical facts suggest PM as an intermediary step for most items. There may well be some examples, particularly related to Christianity, the primary religion in coastal areas, which came directly from Dutch, likely through missionaries to the area. However, contact was limited even here, as most conversions were carried out by Indonesian *gurus* from Ambon and Manado who were not

20. Mary Dalrymple (p.c.) points out that a prominent Wamesa-speaking man from Dusner village, currently in his 60's, was educated in Dutch as well.

Dutch speakers. The same is true of most colonial administrative functions in the province (van den Heuvel 2007). Those few Dutch missionaries who did spend time in the region (van Balen, Bink, Kamma) generally used Malay as a contact language, again minimizing the direct influence of Dutch on local languages such as Wamesa (David Kamholz p.c.). Overall, spoken Dutch gained little foothold at all in Indonesia (Tadmor 2007), with less than 1% of Indonesians achieving fluency even at the height of Dutch language instruction in the early 20th century before World War II. What speakers there were were largely clustered in economic and political hubs such as Jakarta in the west of the country, not in remote provinces such as Papua (de Vries 1988).

Contact with English is extremely limited, as the British were never a major presence in West Papua. Most exposure these days is through mass media, where it is embedded in an Indonesian language context, and available primarily for those urban (and a far more limited number of rural) speakers with access to a television. The only confirmed loans directly from English in my data are those which came directly from me, namely my name and the names of places I have lived. Other words of English origin, such as [es krim] 'ice cream' and [dius] 'juice' are present in Malay as well.

Standard Indonesian is a relatively new import into Papua, appearing only after the Indonesian annexation of the territory and the beginning of the transmigration program in the last 50 years, which relocated large numbers of people from western Indonesia to Papua (Donohue & Sawaki 2007). Standard Indonesian is rarely encountered even in urban centers, much less in rural areas; most "Indonesian" conversation is conducted using a non-standard basilectal variety with more or fewer Papuan Malay vs. standard features depending on the level of formality. (For further discussion on the relationship between Standard Indonesian, Papuan Malay, and the continuum of lects between them, see for example Donohue & Sawaki 2007; Saragih 2012; Fields 2010, and Kluge 2014.) I assume that loans of Malay origin likely came from Papuan Malay rather than Standard Indonesian, and where pronunciations differ between the two languages I will take PM to be the input, not BI.

Wamesa also has a long history of contact with neighboring Austronesian and Non-Austronesian languages, such as Dusner, Biak, Moor, Kuri, Ambai, Iresim, and others. Unfortunately these languages are by and large severely under-documented, and the details of the contact histories unknown. Possibly because of its relatively large territory, Wamesa has a history of use as a lingua franca between various groups of the region and is learned as a second language (T. Karubuy p.c.); for example, residents of Dusner village are all speakers of Wamesa, though their own language has only three remaining fluent speakers (Dalrymple & Mofu 2012). The related language Umar (also referred to as Yeretuar) is reported to be heavily influenced by loans from Wamesa (David Kamholz p.c.), but only one possible loan from Umar to Wamesa has been tentatively identified. More work is required in this area to fully flesh out what is certainly a rich history of regional contact.

2.4.2. Theories of Loan Adaptation

Malay words generally undergo little alteration when borrowed into Wamesa. This is in part a consequence of the large degree of overlap between the PM and WAD phoneme inventories: of the segments found in PM, five consonants, [l], [dʒ], [tʃ], [ɲ], [h], and one vowel, [ə] do not also appear in the native WAD lexicon.[21] Those segments which appear in both languages are borrowed intact, without modification. Malay allows complex syllable onsets which are not licit in Wamesa (Kluge 2014); these may be preserved or broken up with an epenthetic vowel, as described in §2.4.5.

Those Malay phonemes which do not occur in native WAD vocabulary also often re-

21. Schwa is marginal in some varieties of Papuan Malay, and not present in others (Donohue & Sawaki 2007; Fields 2010; Kluge 2014, Cika Tethool and Eny Arilaha p.c.).

main unchanged in loan words, even when those words bear Wamesa morphology, as in [iveolaraga] 'I exercise', from WAD *i-ve-* '1sg-ESSIVE' and Malay [olaraga] 'exercise', with the non-native [l] and otherwise-unattested intervocalic [g] left unchanged. This can be attributed to the bilingualism of Wamesa speakers, the fact that those living in the city predominantly use PM in their everyday life, and the high degree of code-switching between WAD and PM even for rural speakers. As speakers are fully fluent in PM, these segments pose no additional challenge for them to produce or perceive, and rapid transitions between PM and WAD are considered the norm.

Shana Poplack writes extensively on the difficulties of distinguishing loans, particularly nonce loans, from single-word code switches in bilingual speech (Poplack & Sankoff 1984; Poplack et al. 1987; Poplack & Meechan 1998). A single item from one language used during discourse in another may constitute a loan, a code switch, the result of incomplete acquisition of the discourse language, or a momentary lapse best classified as interference. As the recipient language in this case is the native language we may rule out incomplete acquisition when considering Malay words in Wamesa discourse. For some speakers, however, particularly those in urban settings whose day-to-day life is carried out largely in Malay, interference may well be a factor. While acknowledging that there is no unequivocal way to distinguish loans from other types of lexical intrusions, Poplack and her co-authors posit several criteria for determining an item's status, pulled from the work of Hasselmo (1970); Fries & Pike (1949); Murphy (1974); Bloomfield (1933a); Mackey (1970), and others, including frequency of use (those items used more frequently and by more speakers are more likely to be true loans), displacement of native synonyms (those which have replaced native items are more likely to be loans), morphophonemic and syntactic integration (those forms which are better integrated into the structures of the recipient language are more likely to be loans), and acceptability (if an item is judged to be an acceptable term for an item or concept by native speakers, it is more likely to be a loan). Haspelmath (2009) adds that if an item is used by monolingual speakers it is likely to be a loan, though this test is inapplicable here as none of the speakers with whom I interacted were monolingual. As Poplack and Sankoff point out, these criteria are far from deterministic; many true loans will fail to fulfill some points, and others will be fulfilled by words better classified as code switches, interference, etc. Poplack & Sankoff (1984) point out as well that an item's form and its use by speakers will change over time as it becomes more fully incorporated into the lexicon, so that a recent loan may fulfill fewer of these criteria than a longer-established one.

While it is impossible to say for sure based on the data so far available, I argue that the forms used as examples here can indeed be classified as loans, though some are nonce loans, imported on the fly when the speakers were asked to describe objects or phenomena for which no Wamesa word exists (for example lions and ice cream), or when the native item did not immediately come to mind (such as the word for flower). The forms cited here were not flagged in the discourse with hesitations, false starts, or other lexical means (see i.e. Poplack et al. 1987 for further discussion of such strategies in Finnish-English bilinguals). Some were online loans, used in a single discussion by a single speaker, but others recurred across elicitations, for example *kodo* 'frog' (five speakers), *keranjang* 'hamper, basket' (two speakers), and *skop* 'shovel' (two speakers).

These three forms, despite their wide distribution amongst speakers, fare differently on other criteria. *Kodo* was used consistently in all cases when a speaker needed to refer to a frog (generally in frog story elicitations), and its phonological form was consistent across utterances. Further probing revealed that in addition to serving as a general term, *kodo* also refers to a specific type of green, inedible forest frog, while another word, *rosua*, is used for a type of brown, edible frog.[22] *Karanjang* and *skop* both had more variable pro-

22. This latter species reportedly tastes like chicken.

nunciations, discussed in more detail below, indicating less complete integration into the Wamesa lexicon. In some utterances the consonant clusters - [ndʒ] and [sk] - and non-native segments - [nʒ] - were retained; in others they were replaces by native sequences. The word *skop* was consistently chosen in reference to a shovel in a frog story; *karanjang* alternated with *nawa*, a native word, and *ember*, a loan from Dutch via PM meaning 'bucket'.

Morphological integration is harder to judge, as Wamesa has little non-verbal morphology (no case marking or noun class markers, for example) and very few examples of borrowed verbs. In a list of 64 non-native lexical items collected from the corpus, only two, *jaga* 'guard' and *olahraga* 'exercise', are verbs and three, *murah* 'cheap', *mahal* 'expensive', and *loba* 'greedy', are adjectives, which take verbal agreement morphology when used predicatively. The remaining 60, including *helem* 'helmet', *botol* 'bottle', and *toples* 'jar', are nouns.[23] The majority of these refer to non-native material culture items, such as forks, candles, and camels, though not all; the list includes terms for native flora and fauna such as the Victoria crowned pigeon and breadfruit, for pre-existing items such as baskets, and for non-culturally-specific concepts like the adjectives listed above.

The verbs and predicative adjectives all bear the prefix *ve-* between the root and the subject agreement marker, shown in (2.47).

(2.47) a. *Sasu sama toini v<i>e-**mahal**?*
clothing buttocks which 3SG-ESS-expensive
'Which pants are expensive?' [S29 IMK]

b. *I-ve-**olaraga**.*
1SG-ESS-exercise
'I exercise.' [S78 IMK]

As described in §4.4.4, *ve-* serves several functions in the language, one of which is

23. This is a more extreme example of the pattern found in Haspelmath and Tadmor's (2009) World Loan Word Database, in which borrowed nouns outnumber verbs two to one.

to mark a predicate as individual-level rather than stage-level, and therefore can appear non-obligatorily on all Wamesa verbs and adjectives. Whether it is used here to mark such a property or simply to avoid attaching the agreement affixes, particularly in their infixed forms, to a non-native root is not clear. That the forms are marked for the person and number of their subject, which does not occur in Malay, argues in favor of their being loans rather than code-switches; that they are not considered appropriate for direct affixing by the agreement marker suggests that they are not fully integrated into the language. A designation of nonce loan is probably most appropriate here.

In all of the examples used here, non-native nouns are fully integrated into the normal NP structure, including, where appropriate, a number-marked determiner, not present in Malay. Poplack et al. (1987) note the presence of determiners, either English forms or adapted Finnish words, used with English-origin lexical items in Finnish discourse by bilinguals as a flag for code-switching by the speaker. Here the situation is switched: the matrix language makes use of determiners while the donor language does not. In this case we might expect the lack of a determiner to flag a code switch rather than a loan; in the examples used here the determiner is present where required by Wamesa syntax, and it bears a number affix *-i* 'sg.' or *-si* 'pl' where appropriate. In one instance, a speaker used the phrase *bunga-bunga pasi* 'the flowers', with plurality marked both by reduplication of the Malay word *bunga* 'flower' and on the Wamesa article *pasi*. This is an instance of a code switch, in which the speaker could not remember the Wamesa word for 'flower' (*ariou*), and applied the morphological requirements of the phrase to each word as required by the language in which it was spoken. Other instances of the phrase *bunga pasi* produced at other occasions by other speakers, minus the reduplication of the code-switched example, show the Malay word fully incorporated into the Wamesa phrase, with plurality marked only as required by Wamesa, on the determiner, and not on the noun, where it would appear in Malay.[24] These are more likely to be nonce loans, though final determination would also need to take into account the other indicators discussed here.

As mentioned above, loans often do not undergo full phonological integration into Wamesa. Given the high levels of community-wide bilingualism, this is unsurprising. LaCharité & Paradis (2005), in their study of loan word adaptation, predict that speech communities with high levels of fluent bilingualism in both the donor and recipient languages will show a relatively high rate of importation of non-native phonemes without subsequent mapping to native sounds. This is borne out in their data, for example in the relative rates of importation of English sounds between Parisian French (low bilingualism; low rates of segment importation), Quebec City French (intermediate bilingualism; intermediate importation rates), and Montreal French (high bilingualism; high importation rates). These findings are replicated in the World Loan Word database, where recent loans and those found in languages whose speakers are also very familiar with the donor language will often fail to undergo phonological nativization (Haspelmath 2009). This exactly predicts the situation we find with Malay loans into Wamesa: extremely high levels of bilingualism correspond to high rates of retention of phonemes in the donor language which do not occur in the native lexicon of the recipient language. More locally, this same scenario holds true of Malay loans into Biak, another community with high levels of bilingualism and little phonological integration of borrowed forms (van den Heuvel 2007).

Table 2.8 gives the mapping of Papuan Malay consonants into Wamesa. Unless noted otherwise in the following paragraphs, English phonemes (where attested) are adapted in the same ways. The English phoneme /v/ is the only consonant attested in a loan word which does not occur in PM; it appears in italics in the table. The upper line gives the PM

24. Number marking of NPs is optional in Malay, and is marked by reduplication.

	labial	alveolar	palatal	velar	glottal
stops	p b	t d		k g	ʔ
	p b	t d/t		k g	-
fricatives/	f v	s tʃ dʒ		h	
affricates	p p	s tʃ dʒ/di		h	
nasals	m	n ɳ		ŋ	
	m	n -		ŋ/ŋg/n/∅	
liquids		l r			
		l/r r			
glides	w	j			
	-	j			

Figure 2.8: Papuan Malay - Wamesa Consonant Mappings. The PM input appears on the upper line, and the WAD output below.

source phoneme, while the lower line gives the WAD outcome.

While many of these segments are faithfully imported into Wamesa in loan words, that is not always the case. Specific segments whose realization in Wamesa differs in some or all attested cases are discussed in the following sections. It is generally accepted in the study of loanword phonology that when a word is borrowed into a language its component segments are altered to conform to native phonology, the resulting segments will differ as little as possible from the original form, though whether similarity is computed on the basis of the surface phonetic or underlying phonological form of the source word is debated (Kang 2002, 2011; Kenstowicz 2003; Steriade 2009). Steriade (2009) proposes the concept of the P-map, a quasi-universal collection of comparisons of the relative perceptual similarity of different segments in different contexts. These will generally be universal, though some features may be language-specific. According to Steriade, the P-map stands as a solution to the too-many-solutions problem in OT,[25] whereby many

25. Though it is proposed to address this problem, the P-map is not in itself theory-specific, and its predictions can be used to constrain a variety of formalisms.

disallowed structures have only a single prototypical repair across languages, despite the predictions of Optimality Theory (Prince & Smolensky 1993/2004) that a number of (unattested) repairs should be possible given different constraint rankings. It is a truism about OT that the too-many-repairs problem cannot be solved by adding additional constraints to Con, but only by changing the set of constraints or restricting the ways in which existing constraints can be ordered. In this view, the P-map constrains the possible orderings of constraints within the grammar such that of the candidate forms which are licit in the language's phonotactics the one most perceptually similar to the input will be the winner. Since the least salient repair will in most cases be the same cross-linguistically, the unattested 'extra' repair strategies will thereby be ruled out, as the constraint rankings necessary to obtain them will be disallowed across languages.

In most cases the repairs seen in Wamesa fit neatly with the idea that the form of the loan word in the host language should be as perceptually similar as possible to that in the source language. Where segments are altered, they tend to differ minimally from the input. The major exceptions to this, the bilabial fricatives realized in WAD as [β], are analyzed in more detail below.

2.4.3. Adaptation of Consonants

[d].

The treatment of PM [d] depends on its position in the word. Word-medially, whether in intervocalic position or as part of a cluster, it is borrowed into Wamesa intact, as shown in (2.48). Word-finally, however, it is devoiced to [t], as in (2.49).

(2.48) a. [ku**d**a] ⟶ [ku**d**a] 'horse'
b. [kan**d**aŋ] ⟶ [kan**d**aŋ] 'cage'

c. [kar**d**us] ⟶ [kar**d**us] ‘box’ (< Dutch)[26]

(2.49) [spi**d**] ⟶ [spi**t**] ‘speed boat’ (< English?)

The word *spit* almost certainly came into Wamesa via Malay. It is ultimately an English form, and Stevens & Schmidgall-Tellings (2004) claims that this is the source from which it was borrowed into Indonesian, though Dutch has the same form (*speedboot*), also a loan from English.[27] Though Dutch, like German, does have final devoicing, it also shows regressive voicing assimilation in compounds (Booij 1995; Grijzenhout & Krämer 2000), which renders the /d/ voiced in this form. The truncated form found in Malay shows voicing on the final stop, consistent with either Dutch or English as the immediate source, since if the form was truncated in Malay, Dutch final devoicing would not apply. Chronology argues for this Malay form as the immediate source of the loan into Wamesa. As previously discussed, what little direct contact existed betwen Dutch and Wamesa came primarily from missionaries such as van Balen, who arived in Windesi in 1889, only three years after the invention of the motorboat. The word *speedboat* is first attested in English in 1911 (oed 2013), too late for it to have been adopted by the Dutch and brought to Papua by the missionaries. There are two arguments against the word’s having moved directly from English into Wamesa. The first is its truncated form, which is the same as that found in Malay. The second is the limited reach even now of the English language into Wamesa territory: exposure is almost exclusively through television and international media, which is not commonly encountered in villages where the only electricity comes from a generator run for a few hours only on nights when enough fuel is available.[28] Canoes, usually motorized to handle the long distances involved, are often

26. Unless otherwise noted, source languages cited here are from Stevens & Schmidgall-Tellings (2004).

27. The native Dutch word for speed is *snelheid*; *boot* is cognate with English *boat*.

28. English instruction in schools is poor and often limited to higher education, for which students must leave the villages. No one I met knew more than a few basic words of English, though in one elicitation the speaker uses the word *fish*, pronounced [pis], repeatedly, for comedic effect.

the only way to move between coastal villages, necessary for trade, and the only way to bring outside supplies into the village from port cities, as well as being useful for fishing. Given their central role in village life, motor canoes are a common topic of conversation, both in Wamesa and in PM with members of other groups. It is these conversations in trading hubs such as Wasior which led to the adoption of the Malay word *spid* (and its synonym *jonson*, after the type of engine used) into Wamesa.

The devoicing of final [d] is an instance of what Kenstowicz (2005) refers to as Retreat to the Unmarked, in reference to McCarthy & Prince's (1994) Emergence of the Unmarked. In this case, a final devoicing process applies to avoid the voiced stop in word-final position. While Wamesa has no productive synchronic final devoicing alternations, only one example of a voiced obstruent in coda position appears in my dictionary, in the word *Yob*, the name for an island in Cenderawasih Bay just south of Windesi. This suggests that Wamesa is one of the many languages of the world in which a constraint banning voiced coda obstruents is ranked above one requiring faithfulness to underlying voicing, though this is not manifested in any alternations in the native vocabulary. (See for example Blevins (2004) and Kiparsky (2006) for discussions of the markedness of final voiced segments.) In accordance with the P-map, devoicing is the canonical repair for an illegal final voiced segment, as the absence of a following vowel means the VOT value of the stop, one of the primary acoustic cues to the voicing value of a segment, is rendered less perceptible, thereby increasing the similarity of the corresponding voiced and voiceless segments in that environment (Steriade 2009). There are several examples in the literature of spontaneous final or coda devoicing by language learners, even when neither L1 or L2 has such a process; see for example Wissing & Zonneveld (1996); Broselow et al. (1998), and Broselow (2004).

[ʔ].

One word which appears repeatedly in the corpus, thanks to Mercer Mayer's Boy/Dog/Frog series of children's books, is *kodo* 'frog'. In Standard Indonesian, this word is pronounced [kodoʔ], with a final glottal stop not present in the Wamesa form. In PM, it is simply [kodo], with no glottal stop. Were this word borrowed from Indonesian, it would be unsurprising for the final glottal to delete; glottal stops are not present in the native Wamesa phoneme inventory, they are only minimally featurally specified, and consist only of a single glottal gesture, with no oral component at all. Substituting the closest licit stop, [k], would in this case be more salient than deleting the segment, making deletion the preferred repair according to the P-map.

However, it is more likely that the word was borrowed from Papuan Malay, which has no glottal stops in its inventory (Kluge 2014). Of the six speakers recorded narrating from a set of three of these stories (nine recordings in total), all exclusively use *kodo*, rather than either of the native words *vakirini* and *rosua*, to describe the frog in the story. As discussed above, each of these three words refers to a specific type of frog or toad; *kodo* is also the generic term for anurans. Given the limited foothold of Standard Indonesian in West Papua, paricularly when compared to the extensive influence of Papuan Malay, combined with the thoroughly integrated status of *kodo* in the Wamesa lexicon, PM is far more likely to be the source of the form in Wamesa. While the P-map would predict the deletion of [ʔ] word-finally, there is no data available to either confirm or disprove this claim. As there are also no examples of intervocalic or initial [ʔ] in loan words, the treatment of the segment in any position remains in question.

[f] and [v].

There are two cases attested in the data of Dutch or English [v] surfacing in Wamesa as [p].[29] From the point of view of the P-Map, which says that a non-native segment will be replaced with the most similar native segment available, this is puzzling: as a voiced fricative, surely [β] is perceptually closer to [v] than the voiceless stop [p] is. This discrepancy is accounted for by the fact that the form first passed through Papuan Malay, which lacks any labial fricatives, before entering Wamesa. The second can be accounted for by analogy with the former, perhaps first passing through a PM phonotactic filter. The two forms are given in (2.50).

(2.50) a. Dutch [**v**ork] ⟶ [**p**orok] 'fork'

b. English [hei**v**ən] ⟶ [he**p**en] '(New) Haven'

Had these words been imported directly into Wamesa from their source languages, the least salient change in realization of the [v] would be [β], which preserves the voicing and manner features of the original, and makes a minor, non-contrastive change to the place feature. This in fact is the realization of [v] in English loans into Fijian, a related language within the Eastern Malayo-Polynesian branch of Austronesian, where for example English *velvet* surfaces as [βɛlɛβɛti] and *devil* as [tɛβɔrɔ] (Kenstowicz 2007).[30] Realizing the segments as [b] might be assumed to be the runner up, as it changes the manner feature, with [p] coming in third, changing both manner and voice. Steriade (2009) cites studies suggesting that a change in manner of articulation is one of the more salient possible alterations, conclusions which are supported by the loan adaptation patterns of languages

29. One additional example, [telepon] 'telephone', is a common Malay word with [p] for English [f], and was clearly borrowed into Wamesa without modification. A fourth form, Malay [fles] 'bottle' (originally from Dutch), was also borrowed intact, with the [f] retained.

30. Fijian does not allow closed syllables or consonant clusters; the epenthetic vowels are a means of avoiding these structures.

such as Selayarese, which repairs illegal final consonants in loanwords by changing their place of articulation, not their manner (Broselow 1999).

Were these words borrowed into Papuan Malay first, however, [p] is a more plausible outcome. PM has no [v], necessitating some adjustment to the input, but [f] was introduced into the language via Arabic loans. Given this state of affairs, the least salient change for the [v]s in *vork* and *Haven* is for them to be realized as [f]. However, [f] is saliently a non-native phoneme in PM - speakers recognize that it does not appear in native vocabulary - and thus may be avoided. The perceptually closest sound to [f], if these words are to be nativized, is [p], yielding [pork], a form which is present but now archaic in PM, having lost ground to Indonesian *garpu.* Many other loans into Malay with original [v] alternate between the two possible outcomes; for example, *November*, borrowed from Dutch, can be pronounced either [nofember] or [nopember], with [nopember] as the preferred pronunciation for all but well-educated, middle-class speakers (de Vries 1988).

From PM [pork]/[hepen], the final step into Wamesa required no further modification of the consonants.[31] This path, rather than direct importation into Wamesa, may also have been reinforced by the fact that [β] is the only WAD phoneme which is not also present in PM. This may have the psychological effect of making it seem somehow special to native vocabulary, leading to avoidance of its use in all but the longest-established loan words.

The place name *New Haven* was introduced in discussions of where I'm from and what 'home' is like for me, particularly when I gave postcards of New Haven as gifts to language consultants. In these cases the name was generally spoken with English phonology but embedded in an Indonesian sentence (*Saya tinggal di kota New Haven* 'I live in New Haven'), not a Wamesa one, and most conversations on the topic were held in Indonesian. This linguistic context alone may have been enough to give the word a Malay

31. The addition of the epenthetic vowel [o] to break up the illegal consonant cluster is discussed in more detail below.

pronunciation, with [p] in place of English [v], in the speakers' lexicons. Failing that, we can say that the segment was adapted by analogy to other known loans following the same pattern. This latter explanation is, however, insufficient, as other examples of forms with the bilabial fricative on which to form the analogy are rare. Only three possible comparable forms appear in my data: *porok*, *telepon*, and the playful use of the word *fish* pronounced *pis*. In the case of the first and second, it is unclear whether speakers are aware of the original Dutch and English forms on which the already-adapted Malay forms they borrowed are based. In the case of *fish*, this was not a loan but rather a speaker's joking attempt to throw some English terms into her sentences. More likely then that *Haven* was pronounced as though it were a Malay word; any analogy would have taken place in that language's grammar, not Wamesa.

[s], [tʃ], and [dʒ].

The voiceless alveolar stridents /s/ and /tʃ/ are realized faithfully in loans to Wamesa, despite /tʃ/'s absence in the native phoneme inventory. This is true of /s/ in all positions in the word; /tʃ/ is only attested in loans word-medially in an /ntʃ/ cluster.

(2.51) a. [**s**urat] ⟶ [**s**urat] 'letter' (< Arabic[32])

b. [ro**s**leten] ⟶ [ro**s**leten] 'zipper' (< Dutch)

c. [kardu**s**] ⟶ [kardu**s**] 'box' (< Dutch)

(2.52) a. [kun**tʃ**i] ⟶ [kun**tʃ**i] 'key'

The voiced alveolar affricate /dʒ/, absent in the native phonology, may be realized faithfully or as a /di/ sequence, depending on the utterance. The /di/ is often, though not always, realized as [dj], with the vowel surfacing as its glide allophone, and this may be the source of the sequence. The original /ʒ/ shares its apical place of articulation with

32. Etymology not noted in Stevens & Schmidgall-Tellings (2004).

the glide, differing in its less-constricted manner of articulation. Fricatives require very precise tongue placement to achieve frication, and it is quite a small physical adjustment to go from [ʒ] to [j]. Since there is no underlying /j/ phoneme in Wamesa, the form of the sequence is stored as /di/, and the vowel is frequently but not obligatorily produced as a glide.

(2.53) a. [**dʒ**us] ⟶ [**di**us] ~ [**dj**us]'juice'

b. [**dʒ**aga] ⟶ [**dʒ**aga] 'to guard'

c. [keran**dʒ**aŋ] ⟶ [keran**dʒ**an] ~ [karan**di**a] 'basket'

[ŋ].

The realization of /ŋ/ in loan words is largely unpredictable. It may surface faithfully as /ŋ/, move to a coronal place of articulation as /n/, receive a following /g/ to form a homorganic cluster /ŋg/, or delete entirely.

(2.54) a. [buŋa] ⟶ [buŋa] 'flower'

b. [si**ŋ**a] ⟶ [si**ŋg**a] 'lion'

c. [kanda**ŋ**] ⟶ [kanda**n**] ~ [kanda**ŋ**] 'cage'

d. [i**ŋg**ris] ⟶ [i**ŋg**ris] 'England, English'

e. [k(e)randʒa**ŋ**] ⟶ [kerandʒa**ŋ**] ~ [kerandʒa**n**] ~ [karandia] 'basket, hamper'

No examples exist of /ŋ/ in initial position. Where the donor language has /ŋ/ directly preceding a /g/, as in (2.54d), the sequence is preserved in Wamesa, as this is a licit cluster in the language. Intervocalically between vowels, /ŋ/ may be retained unchanged, as occurred in two different speakers' pronunciation of the Malay word *bunga* [buŋa][33] 'flower' during a frog story (Mayer & Mayer 1971). In another Malay word, *singa* 'lion', a

33. As in Wamesa, <ng> denotes [ŋ] in Malay orthography.

following /g/ was added to bring the word into alignment with the Wamesa phonotactic requirement that /ŋ/ occur outside of a homorganic NC cluster only morpheme-initially.

In final position, the segment's behavior is much more variable. Unlike the other nasals /m/ and /n/, there are no native Wamesa words which end in /ŋ/. The repair of this segment can vary within a single speaker. Example (2.54e) above shows three possible realizations of the word *keranjang*, borrowed from Malay, all from the same speaker within a single telling of a frog story.[34] In the first listed form, the word is pronounced exactly as it would be in PM. In the second, the final nasal is changed to /n/. In the third, the vowel of the first syllable is changed,[35] the affricate becomes a /di/ sequence as discussed above, and the /ŋ/ deletes entirely. Though /n/ is acceptable in word-final position, as shown by words like *sien* 'master', *ron* 'ironwood', and *utin* 'twenty', the vast majority of words in the language end in open syllables. Given that steep statistical skew, it is not surprising that an illegal final segment in a loan word might be deleted rather than adapted when nativized.

[l].

In most instances, /l/ is retained in loans. One example exists where the original /l/ has been changed to /r/.

(2.55) a. [**l**i**l**in] → [**l**i**l**in] 'candle'

b. [boto**l**] → [boto**l**] 'bottle'

c. [ba**l**] → [ba**r**] 'ball'

Featurally [r] and [l] are similar in that both are apical liquids, though acoustically they are quite different. Cross-linguistically, their phonological distribution, as surveyed

34. Other speakers telling this same story chose other words for the object in question, a bucket, including *ember* (originally from the Dutch *emmer*, appearing in Malay as *ember*) and the native Wamesa form *nawa*.

35. See §2.4.5.

by Proctor (2009), supports the grouping of laterals and rhotics into a single liquid class. Common phenomena include allophony between liquids and rhotics, whether phonologically conditioned (Sranan, Japanese, Hausa) or in free variation (Sentani, Jita); historical mergers (Maori) and splits (Avok) within the category; and common behavior in metathesis, dissimilation, assimilation, harmony, and post-vocalic deletion. Despite the phonological evidence, no simple acoustic phonetic property exists which is present for all liquids and can serve as a signal of class membership, unlike other major natural classes such as nasals and stops. Gick et al. (2006), in a cross-linguistic articulatory study, found that all post-vocalic liquids in their sample had an accompanying dorsal constriction, and that intervocalic liquids were characterized by multiple simultaneous articulatory gestures. Proctor (2009) goes farther, showing that coronal liquids, such as those involved in Malay and Wamesa, prototypically entail a consonant-like tongue tip constriction coordinated with a stable, vowel-like tongue body gesture. These more-complex phonetic features must form the basis of the [l] $\longrightarrow$ [r] mapping here.

As well as being the sole example of [l] adapted to fit the native phonology rather than simply imported, *bar* 'ball' is also the only example in my data of a Dutch word which must have been borrowed directly into Wamesa without PM as an intermediary language. While *bal* does exist in Malay, the much more common word is *bola*, originally from Portuguese. This form has a final vowel not present in the Wamesa form, which would be expected to be retained if that were the source of the loan. Dutch has the form *bal*, with no final vowel, and therefore must be the donor language. This form appears as such in other languages of the area as well, for example in Moor, Yaur, and Yerisiam, as *bar* or *bare* depending on the phonotactic requirements of the language (none of these varieties allow /l/, and some require a final vowel) (David Kamholz p.c.).

2.4.4. Adaptation of Vowels

Vowel adaptation into Wamesa from Malay is for the most part very straightforward. PM has a five-vowel system identical to that found in Wamesa - /a, e, i, o, u/ - with the addition of a marginal /ə/. Schwas are uncommon in the variety of PM spoken in this area; most words with schwas in Standard Indonesian instead show [e] or ∅. There are no loans in my corpus with schwa in PM. All other PM vowels are present in the native Wamesa phoneme inventory as well, and so are brought in unchanged.

(2.56) *Words from Malay*

a. [dʒaga] ⟶ [dʒaga] ‘guard’

b. [medʒa] ⟶ [medja] (< Portuguese) ‘table’

c. [siŋa] ⟶ [siŋga] (< Sanskrit) ‘lion’

d. [gomo] ⟶ [gomo] ‘breadfruit’

e. [buŋa] ⟶ [buŋa] ‘flower’

English and Dutch have a more extensive vowel inventory than Wamesa, and loans from these languages must therefore undergo more modification when brought into the language. Because of the dearth of direct contact, however, most examples enter the language with their vowels already altered to fit the Malay inventory, making further adaptation unnecessary. The few items which might be considered to have entered the language directly from English are my name and the names of places I have lived, but as discussed above these items were most likely assimilated into Malay phonology before being used in Wamesa. One possible direct Dutch loan is *bar* ‘ball’. In all of these cases, lax vowels in English or Dutch are realized as their tense vowel counterpart in Wamesa (as well as in PM), schwas become [e], and the English diphthong [e^j] is monophthongized to [e].

(2.57) *English words via Malay*

a. E [əmɛɹɪkɑ] → M [amerika] → WAD [amerika] 'America'

b. E [æpl̩] → M [apel] → WAD [apel] 'apple'

c. E [hɛlɪkoptəɹ] → M [helikopter] → WAD [helikopter] 'helicopter'

d. E [dʒus] → M [dʒus] → WAD [dius] 'juice'

(2.58) *Possible direct loans from English and Dutch*

a. E [ɛmɪli] → [emili] 'Emily'

b. E [jeʲl] → [jel] 'Yale'

c. E [nu heivən] → [nu hepen] 'New Haven'

d. D [bɑl] → [bar] 'ball'

2.4.5. Illegal Clusters

Wamesa does not allow words to contain clusters which consist of other than a nasal plus homorganic stop; complex onsets and codas are similarly disallowed. As with illegal segments in loaned forms from Malay, these are often permitted to stand as is. When they are adapted to fit native phonotactics, it is always through epenthesis, not deletion, as predicted by the Preservation Principle (Paradis & LaCharité 1997), which states that languages prefer to preserve as much structure as possible from an input form, choosing insertion over deletion wherever possible. Many of these items come from English or Dutch via PM. Malay allows certain clusters in onset position and across a syllable boundary. As seen in the forms given in (2.59), those forms in which illegal clusters are preserved in Wamesa tend to be less phonologically nativized overall, preserving non-native segments such as [l], [f], [dʒ], and [tʃ].

(2.59) *Preserved Clusters*

a. [ro**sl**eten] → [ro**sl**eten] 'zipper'

b. [**fl**es] ⟶ [**fl**es] ‘bottle’

c. [heliko**pt**er] ⟶ [heliko**pt**er] ‘helicopter’

d. [o**rg**an] ⟶ [o**rg**an] ‘piano’

e. [ma**mbr**uk] ⟶ [ma**mbr**uke] ‘crowned pigeon (*Goura sp.*)’

f. [ku**ntʃ**i] ⟶ [ku**ntʃ**i] ‘key’

g. [kera**ndʒ**aŋ] ⟶ [kera**ndʒ**an] ‘hamper, basket’

There are a number of words in PM, both native and borrowed, which may surface with an initial cluster when the [e] vowel of the first syllable is optionally dropped in fast and casual speech. Cognate forms in Standard Indonesian show a schwa in this position, which may also undergo deletion in fast speech. These words may be borrowed into Wamesa in either their full or reduced forms. Compare the forms in (2.60), where the vowel is retained, to those in (2.61), with the optional Malay vowel absent.

(2.60) a. [k(**e**)lintʃi] ⟶ [k**e**lintʃi] ‘rabbit’ (< Dutch)

b. [k(**e**)ra] ⟶ [k**e**ra] ‘monkey’

c. [k(**e**)randʒaŋ] ⟶ [k**e**randʒan] ‘basket, hamper’[36]

(2.61) a. [g(**e**)las] ⟶ [glas] ‘glass’ (< Dutch)

b. [p(**e**)raŋko] ⟶ [praŋko] ‘stamp’(< Dutch)

A small number of words, three in the sample, make use of an epenthetic vowel to break up an illegal cluster. These forms are given in (2.62). As noted above, the form *karandia* in (2.62c) alternates with less-nativized *keranjang* and *keranjan* as the Wamesa pronunciation of Malay *keranjang*.

(2.62) a. [pork] ⟶ [por**o**k] ‘fork’

36. This is one of three attested realizations of *keranjang*.

b. [s(e)kop] ⟶ [s**i**kop] 'shovel'

c. [k(e)randʒaŋ] ⟶ [k**a**randia] 'basket, hamper'

Contrary to common claims (Lombardi 2003; Bermúdez-Otero & Börjars 2006; Rose & Demuth 2006 among many others; see Uffmann 2006 for a more complete list) that languages will choose the most unmarked, most underspecified, or least phonetically salient default vowel for epenthesis independent of context, the cases above show three different vowels used in Wamesa. More evidence is needed to firmly establish the environment is which each segment is called upon, but the pattern appears to be this: after an apical consonant, the inserted vowel is [i], as found also in Shona (Uffmann 2006); when adjacent to [r], a copy vowel is inserted, matching the vowel on the other side of the [r] in quality.[37] It is surprising that vowel quality should be copied only over [r]; as discussed by Proctor (2009), the production of liquids specifies gestural targets for both the tongue tip and body, making them less prone to coarticulation and acoustic influence by neighboring vowels. Given the paucity of data for this point however, it may prove that [r] is one of a larger natural class of consonants, such as sonorants, across which the quality of an epenthetic vowel can be determined by its neighbors; this would be a more plausible claim (see i.e. Hall 2006).

As for the predictions of the P-map in regards to epenthetic segments, Steriade (2009: 175) says that "If a language contrasts schwa and zero in some context, or if it contains non-alternating forms with schwa, and if it resolves clusters through epenthesis, then the choice of productive epenthetic vowel is limited to schwa." For a language such as Wamesa, which does not make use of a schwa, Steriade does not explicitly predict

37. Alternatively, one could claim that the round vowel [o] appears after a labial, and the low [a] after a dorsal; this however begs the question of why [o] is the chosen round vowel rather than [u], as in Japanese and Sranan (Uffmann 2006), when [u] is more common across the Wamesa lexicon than [o] is. Uffmann (2006) surveys loans into the genetically and geographically diverse languages Shona, Sranan, and Samoan and finds that [o] is significantly and consistently less common as an epenthetic vowel in illegal clusters than [u].

how vowel epenthesis should be handled. If the P-map enforces the insertion of whatever segment is least different from zero in that environment, then it would make sense in a language with no minimal vowel such as schwa that the ideal epenthetic vowel should vary based on its immediate environment. An epenthetic vowel which gets its features from neighboring segments rather than projecting its own full feature set adds similarly minimal amounts of extra information to the underlying form of the word, and will be less perceptually salient than a default vowel since it varies less from the sounds in its vicinity. On this interpretation, it would be surprising to find a single default epenthetic vowel in Wamesa, in contrast to languages such as Indonesian which have schwa in their phoneme inventory.

Here too we see evidence that loan adaptations are not always drawn from the same level of representation. In (2.62b) and (2.62c), the Wamesa pronunciation must have been adapted from the casual register surface form of the Malay word, with the [e] deleted. Were they derived from the underlying forms, we would expect the vowel of the first syllable to appear unchanged as an [e], as it is unchanged in all other examples of Malay loans into the language (c.f. the Wamesa pronunciations [amerika], [nelon] 'fishing line', etc). That [i] and [a] respectively are used instead of [e] is evidence that these vowels are epenthetic, not adapted from the Malay. Further supporting this claim is the fact that all three forms are fully in accordance with the Wamesa phoneme inventory. While [pork] and [skop] need no intervention beyond epenthesis to render them fully phonologically nativized, [karandia] shows two additional modifications, the replacement of [dʒ] with [di] and the deletion of final [ŋ]. This fully-adapted realization of the Malay source word *keranjang* contrasts with the two other attested Wamesa realizations, unmodified [kerandʒaŋ] and intermediate [kerandʒan], both of which are taken from the underlying, not surface, form of the word.

Wamesa is not the only language which makes use of more than one strategy to determine the quality of an epenthetic vowel. Uffmann (2006) gives three options, of which a language may make use of any combination: default insertion, vowel copy (or harmony), and consonantal assimilation. Wamesa appears to make use of the latter two. Within Austronesian, the same two strategies are used by Selayarese (South Sulawesi) and Cook Islands Maori (Oceanic) (Kitto & De Lacy 1999). Samoan (Oceanic) makes use of all three strategies, with consonantal place of articulation as the strongest predictor variable foe epenthetic vowel quality while default epenthesis and vowel copying play a smaller role (Uffmann 2006). In Samoan, as well as Sranan (Smith 1977) and Fula (Paradis 1996), epenthetic copy vowels get their quality from a preceding vowel; Wamesa is unusual in that here the process is bidirectional.

Chapter 3

Stress

3.1. Introduction

Wamesa is a stress accent language (as opposed to tone or pitch-accent). Every word in the language bears one primary stress; in sufficiently long words secondary stress may appear as well. This chapter begins in §3.2 by laying out the basic stress pattern of the language, which is lexically determined but always appears within a word-final three-syllable stress window. This section discusses the acoustic correlates of stress and gives statistical tendencies as to its placement. §3.3 gives a brief overview of how the Wamesa pattern compares to other neighboring languages. §3.4 gives some theoretical background on gradient versus categorical constraint evaluation, and §3.5 discusses approaches to the size and position of the stress window. §3.6 gives an account of the placement of basic stress in the language.

In certain circumstances, Wamesa primary stress undergoes a pattern of rightward shift. Some Wamesa words have antepenultimate stress; when these forms bear an enclitic, their primary stress appears instead on the final syllable of the stem, with secondary stress on the penultimate syllable. This only occurs on words of three or four syllables, and only those with antepenultimate stress in clitic-free contexts. §3.7 of this chapter gives an account of stress shift, proposing a cap on the number of secondary stresses a monomorphemic word may bear, demonstrating the insufficiency of stress faithfulness constraints to account for this pattern, and discussing native speaker intuitions regarding stress shift.

3.2. Basic Stress Pattern

3.2.1. Primary Stress

As with most eastern Austronesian languages, Wamesa has no lexical tone.[1] Wamesa stress placement is lexically determined and not predictable. The sole restriction is that in monomorphemic forms, primary stress always occurs within a three-syllable window at the end of the stem, similar to the pattern found in, for example, Macedonian (Comrie 1976; Franks 1989; Hyde 2012). In a list of 105 words spoken in isolation, a sample of convenience taken from the three highest-quality field recordings, 66 have clear penultimate stress, 23 have final stress, 13 have antepenultimate stress, and in one form stress placement varies unpredictably between tokens.

Here stress was primarily diagnosed impressionistically; key forms were checked with other linguists. Stress is primarily signaled by a lengthening of the syllable rhyme as compared to unstressed syllables, sometimes with an additional pitch peak early in the syllable followed by a fall, particularly in word-final stress. Unstressed vowels are somewhat centralized and reduced. Vowel length was measured in 64 tokens of 22 words of various lengths and with stress in various positions, all pulled from natural speech as produced by a single speaker. This included 234 total syllables, 64 of which bore primary stress. 32 of these vowels were in closed syllables; only the vowel length, not that of the

1. One notable exception is Mor, a Cenderawasih Bay language spoken to the southeast of Wamesa (Laycock 1978; Kamholz 2009).

entire rhyme, was measured. The mean raw vowel length overall was 103ms; mean length of primary stressed vowels was 143ms; mean length of non-primary-stress-bearing vowels was 89ms. Under a two-sample T-test, each of these means is significantly different from each of the others, $p < 0.001$. These values were also normalized to account for variable speech rate by computing the mean of the vowel length for each token, then computing the ratio between each individual vowel's duration and the mean for that word token. Vowels bearing primary stress were on average 1.39 times longer than the mean length for their word token; the ratio of durations of un- or secondarily-stressed vowels to token mean was 0.92. These numbers are again significantly different with a p value of under 0.001. Similar calculations were done for pitch and intensity of vowels. Pitch values for primary-stressed vs. other vowels were not significantly different; differences in intensity were significant for both raw ($p = 0.032$) and normalized ($p = 0.017$) measurements, though much less so than for length.

The spectrograms in Figures 3.2.1 - 3.2.1 demonstrate this. Figure 3.2.1 shows penultimate stress on a disyllabic word and Figure 3.2.1 on a trisyllabic word; Figure 3.2.1 gives an example of final stress. These tokens are pulled from natural speech (frog story narration); all are phrase-medial followed immediately by another word within the phrase, with no intervening pause, to avoid initial and final lengthening effects.[2] In all of these examples, the stressed vowel is visibly longer than the others in the word.

Examples of the three possible loci of stress are given in (3.1), on words of various lengths and on a range of syllable types. Only primary stress is transcribed here; secondary stress in almost all cases predictably appears to the left of primary stress, on alternating syllables.[3] High vowels in these examples are transcribed as unreduced, though if

2. Figure 3.2.1: *Kodo nei kota kiopa* 'The frog jumps too.'; Figure 3.2.1: *Eh, pibata pai vioru pa* 'Eh, the turtle died.'; Figure 3.2.1: *Tiau rawa so kambu pai* 'He falls down into the water.'

3. §3.6.6 accounts for an exceptional case, *sìvereréi*.

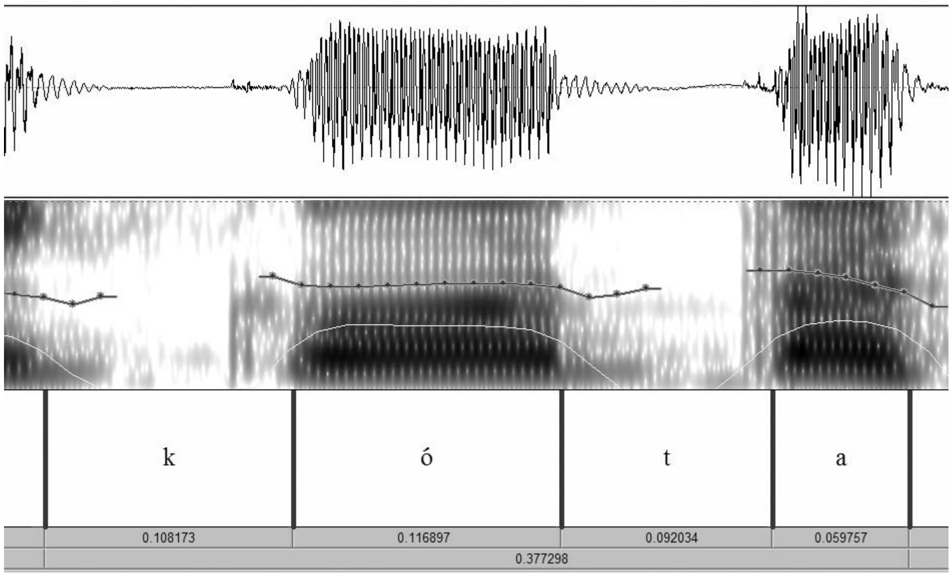


Figure 3.1: *kóta* ‘also’

unstressed this will vary in actual production. (See §2.3.6 for a discussion of the reduction of unstressed high vowels to glides.) The placement of stress in the word is unaffected by whether a high vowel is reduced or full in a particular token; reduction of a vowel does not trigger relocation of stress. This is because non-obligatory reduction - that which takes place when the resulting glide does not form a simplex syllable onset - is an optional post-lexical phonetic process, and as such takes place late in the derivation, after stress placement.

(3.1) a. *Penultimate Stress*

[ˈdi.a] ‘fish’

[ˈko.ta] ‘also’

[mu.ˈan.du] ‘two’

[a.ˈni.bar] ‘bee’

[a.muŋ.ˈge.ri] ‘sago grub’

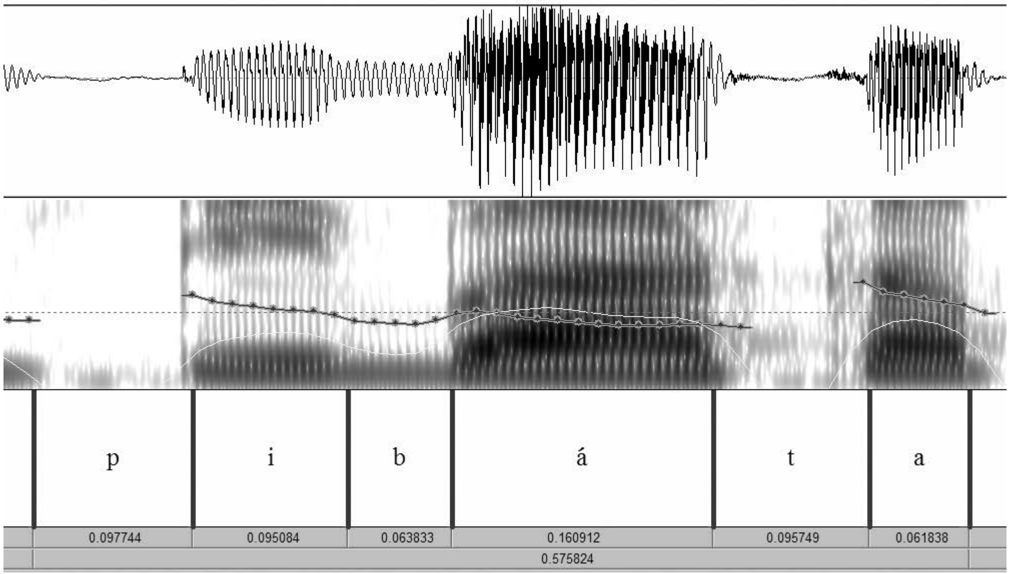


Figure 3.2: *pibáta* ‘turtle’

b. *Final Stress*

[ˈnu] ‘island’

[bu.ˈo] ‘fruit’

[vaˈra] ‘hand’

[kam.ˈbu] ‘water’

[pi.ri.ˈe] ‘shell’

[an.am.ˈbet] ‘cold sago pudding’

c. *Antipenultimate Stress*

[ˈaŋ.ga.di] ‘coconut’

[na.ˈna.ri.a] ‘slowly’

[mi.mi.ˈo.ta.ri] ‘common snakehead fish (genus *channa*)’

d. *Variable Stress*

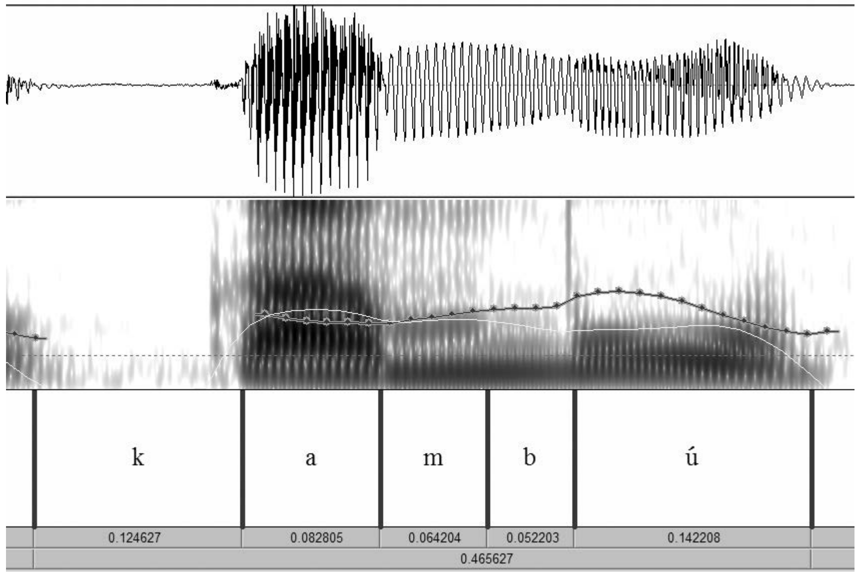


Figure 3.3: *kambú* ‘water’

[ˈni.o.i] ~ [ni.ˈo.i] ‘knife’

Stress in Wamesa is not predictable by syllable weight. As demonstrated by the examples in (3.1), closed syllables are not reliably stressed. Of the 27 total underlyingly closed syllables[4] in the 105-word sample, roughly half (56%) are unstressed. Vowels preceded by a reducible high vowel are slightly more likely to bear stress; of the 54 instances in the sample, 33 (61%) are stressed. More predictive is whether a given vowel is followed by a high vowel. Of the 17 V+V[HIGH] sequences in the sample, 14 (82%) have stress on the first (non-high) vowel. Twelve of these constitute default penultimate stress in tokens where the high vowel is unreduced. Examples include [moˈmoi] ‘breadfruit’, [niˈau] ‘cat’, and [ˌniniˈai] ’floor’. This constitutes one piece of evidence for the trochaic analysis put forth here; given trochaic footing, the comparatively-low-sonority high vowel which ends these words falls in the weak branch of the foot. There are nonetheless lexical exceptions

4. That is, those closed by a consonant and not a vowel reduced to a glide.

to this pattern; not all word-final V+V[HIGH] sequences bear stress on the first of the two vowels or even at all. Counterexamples include [anaŋˈganai] 'bait', and optionally [ˈnioi] 'knife'.

One stress-based minimal pair appears in my data: [ˈra.wa] 'side' vs. [ra.ˈwa] 'to there (distal)'. The latter *rawá* 'to there' is synchronically bimorphemic; it is a compound composed from the locational/directional particles *ra* 'to, towards' and *wa* '(distal)', but comprises a single word. *Rawa* 'side' is monomorphemic. Cowan (1955), citing van Balen, mentions two other examples of contrastive stress: [ˈsasi] 'salt; ocean' vs. [saˈsi] 'dung' and [ˈbabar] 'clapping of wings' vs. [baˈbar] 'don a loincloth'. The first pair was successfully elicited in my data but no consistent contrast in stress appeared; *sasi* meaning 'salt' had variable stress while stress on 'dung' was consistently final. The second pair could not be elicited.

3.2.2. Secondary Stress

Wamesa has audible secondary stress on eligible words of three or more syllables. This stress was again diagnosed by audition, with independent confirmation of key forms by other linguists. Measurements show that vowels bearing secondary stress are not reliably longer than unstressed vowels. They often show higher intensity, but because they occur early in the word this may, at least in part, be a function of the intensity contour of the word, which generally starts high and falls steadily. The main cue for secondary stress is the absence of vowel reduction. The spectrograms in Figures 3.4 and 3.5 show secondary stress on the words *kàmaréni* 'many-pointed fishing spear' and *àparápiri* 'gnat'. *Kàmaréni* in this case is phrase-final, and shows the long duration and low intensity found on many unstressed phrase-final vowels.

Secondary stress is almost always predictable in its placement, surfacing on alternating syllables to the left of the main stress, as in *àrióu* 'flower' and *suòmabút* 'large forest

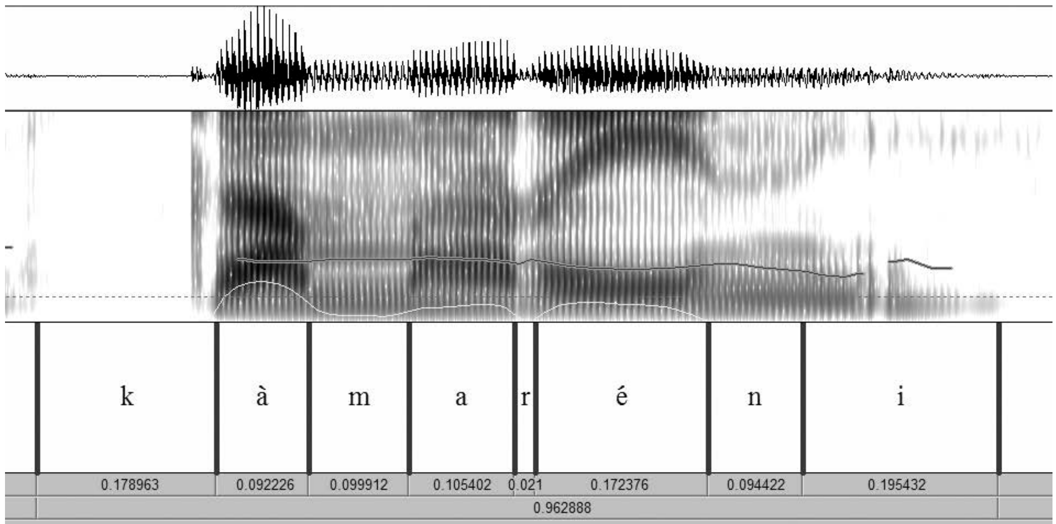


Figure 3.4: *kàmaréni* 'many-pointed fishing spear'

rat' (though see §3.6.6 for one lexical exception). Wamesa places a limit on the number of secondary stresses which may appear on any monomorphemic word; no simplex or encliticized form attested in my data bears more than one primary and one secondary stress. This generalization does not hold when the stem is morphologically complex, as with compounds and forms bearing prefixes and/or suffixes. The lack of secondary stress is not simply a matter of length, though the vast majority of Wamesa words are not sufficiently long to host more than one secondary stress. In principle, a five-syllable word could bear three stresses, two secondary and one primary on the final syllable, as (ˌσσ)(ˌσσ)(ˈσ); this is not attested in the lexicon, though this could simply be an accidental gap. Not accidental is the fact that in a five-syllable word like *àparápiri* 'gnat', the language tolerates a word-final lapse rather than introducing a third stress, even when the addition of a non-stress-bearing enclitic, such as a determiner or negation particle, extends the lapse to three or more syllables in length. On shorter forms, this induces shift of the primary stress to the final syllable; in five-syllable roots no shift occurs and no additional stress is added to fill the lapse. I argue that this results from a constraint requiring

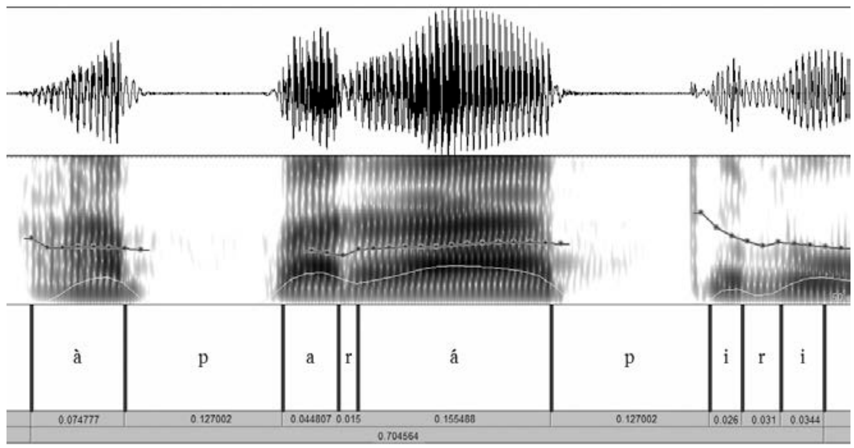


Figure 3.5: *àparápiri* 'gnat'

PWords to be binary, ruling out a third foot in the root and therefore a second primary stress. Stress shift, and its failure to apply to long roots, is discussed in §3.7.

3.3. Areal Context

The tendency within the Austronesian family is for languages to have penultimate stress. In her overview of Austronesian typology, Klamer (1998) cites a study by van Zanten et al. (2010) based on a sample of 117 languages, which finds 92 with penultimate stress. (In the remaining 25 languages, stress was either initial, final, or variable.) The picture within the Cenderawasih Bay family is somewhat cloudier. Klamer asserts that no more than 14 tonal Austronesian languages have been reported; two, Mor and Ma'ya, are Cenderawasih Bay languages. Dalrymple & Mofu (2012) say only that stress is usually penultimate on polysyllabic words in Dusner; they do not clarify the extent or nature of the variation. Biak stress, according to van den Heuvel (2006), results from the interaction of phonemic vowel length and postlexical rhythmic processes affecting larger phrasal units, rendering it quite variable. Laycock (1978) claims that Mor has lexical stress in addition to contrastive tone, and that stress is unpredictable.

Ambai, Wamesa's closest well-documented relative, has a more regular stress pattern. Here primary stress is penultimate, with secondary stress falling two syllables (one foot) prior (Silzer 1983). Silzer notes that diphthongs in Ambai attract stress, but does not provide any examples where this violates the pattern described above. Price & Donohue (2007) point out further exceptions to the default pattern of penultimate stress in Ambai, including lexically-determined stress shift triggered by the addition of an object suffix and apparent internal extrametricality, but the majority of words bear trochaic penultimate stress. This areal context, combined with the skew towards penultimate stress in Wamesa, suggests that an earlier form of the language or an ancestor had regular penultimate stress, and that the current mixed pattern is the result of historical change.

3.4. Analytical Preliminaries: Non-Gradient Constraints

Though constraints requiring the alignment of two elements or structures within a word play only a small role in the analysis presented here, they are prominent in the broader literature on stress assignment (and will feature far more prominently in the discussion of infixation in Chapter 6). After McCarthy (2003), all alignment constraints in this dissertation will be evaluated categorically, never gradiently.

Under Generalized Alignment, each affix is specified as to how a particular edge of that affix aligns with an edge of another prosodic or morphologized domain. Align constraints are specified in the format Align(Category$_1$, Edge$_1$, Category$_2$, Edge$_2$), where Cat1 and Cat2 are members of the union of the sets of prosodic and grammatical categories (PCat and GCat) of a given theoretical framework, and Edge1 and Edge2 may be either Left or Right. Universal quantification applies over Cat1, and existential quantification over Cat2. Alignment constraints are formally defined as follows:

(3.2)

Align(Cat1, Edge1, Cat2, Edge2) $=_{def}$
∀ Cat1 ∃ Cat2 such that Edge1 of Cat1 and Edge2 of Cat2 coincide.
Where
Cat1, Cat2 ∈ PCat ∪ GCat
Edge1, Edge2 ∈ {Right, Left}

(McCarthy & Prince 1993:80)

As originally proposed by McCarthy & Prince (1993) and in much of the literature since, Align constraints are evaluated gradiently - that is, a single locus of violation (a mis-aligned foot, affix, or feature) can incur multiple violation marks based on the extent of its misalignment. The structure [$_{PWd}$σσ(σσ)], for example, will incur two violations of an Align constraint requiring feet to be aligned with the left edge of a PWd because, though there is only a single mis-aligned foot (ergo a single locus of violation), that foot is misaligned by two syllables. The constraint will prefer [$_{PWd}$σ(σσ)σ] (one violation) over the previous example, and [$_{PWd}$(σσ)σσ] (no violations) over that.

While this ability to discriminate between degrees of misalignment has proven useful for many analyses, several authors, notably Eisner (1997); McCarthy (2003) and Buckley (2009) have argued that it is too powerful and makes pathological predictions. McCarthy (2003) points out that a factorial typology of gradient Align constraints predicts the existence of unattested stress systems such as right-to-left iambs. This may or may not be problematic - unattested is not the same as impossible - and McCarthy also gives an example of a more clearly troubling result. He lays out a hypothetical language where Onset dominates Align(Stem, L, Wd, L) (as, perhaps, in Western Aranda; see §3.5.1). High-ranked Exhaustivity will cause the PWd and stem to be co-extensive in V-initial words, shown below, despite the preference of the Onset constraint. In this case, violations of Align are evaluated in terms of the number of intervening segments.

(3.3) *Gradient Alignment I*

/VCVCV/			Exhaustivity	Onset	Align(Stem, L, Wd, L)
a.	☞	$[_{Wd}$VCVCV]		*	
b.		V$[_{Wd}$CVCV]	*!		*
c.		VCV$[_{Wd}$CV]	*!**		***

If in this language recursion is allowed in the PWd - that is, Non-Recursion(Wd) is low-ranked - the addition of a CVC- prefix will cause the word to be parsed into two PWds, one of which is not quite aligned with the original stem. With the given constraint set the grammar will not differentiate between candidates (a) and (b) but either outcome is equally problematic typologically.

(3.4) *Gradient Alignment II (adapted from (McCarthy 2003: 88)*

/CVC-VCVCV/			Exhaust	Onset	Align(Stem, L, Wd, L)	Non-Rec(Wd)
a.	☞	$[_{Wd}$CV$[_{Wd}$C-VCVCV]]			*	*
b.	☞	$[_{Wd}$CVC-V$[_{Wd}$CVCV]]			*	*
c.		$[_{Wd}$CVC-VCVCV]			**!*	
d.		$[_{Wd}$CVC-$[_{Wd}$VCVCV]]		*!		*

Exhaustivity, which enforced the inclusion of the initial V into the PWd on the V-initial forms, is here satisfied in all of the given candidates. Of the three forms which obey Onset, gradient evaluation of the Align constraint rules out that with the single PWd, leaving the two candidates in which the internal PWd is mis-aligned with the stem by a single segment. In either case, McCarthy points out, we have a problem: either the final C of the prefix or the first (though non-initial) C of the root is PWd-initial, and would be expected to behave as a word-initial segment. This, according to McCarthy, is unattested and implausible.[5] If the ALIGN constraint is instead evaluated categorically, with one violation mark per locus of violation, each of candidates (a), (b), and (c) will receive only one violation mark, leaving NON-RECURSION to choose candidate (c), with a single PWd. This, McCarthy claims, is the desirable result, as it does not predict word-initial behavior of root- or affix-internal segments.

Hyde (2012) gives another problem with gradient ALIGN constraints, modeled on an observation by Eisner (1997), which he terms the *Midpoint Pathology*. Take the constraint ALIGN(σ, L, Ft, L, σ), which is well-formed within the Generalized Alignment framework proposed by McCarthy & Prince (1993). The constraint is defined by Hyde (2012: 793) as follows:

(3.5) ALIGN(σ, L, Ft, L, σ): The left edge of every syllable coincides with the left edge of some foot. Assess a violation mark for each syllable intervening between misaligned edges.

When ranked above any countervailing constraints, this constraint will align the left edge of the foot with the left edge of the medial syllable of the word, as demonstrated in (3.6). This is a clearly undesirable outcome.

(3.6) *Gradient Alignment III*

5. It is possible that Western Aranda presents a counterexample, as Goedemans (1996) argues that the PWd does indeed mis-align with the stem, leaving initial vowels stressable; in this case the second syllable of the word acts as though it were initial at least for the purposes of stress assignment.

/σσσσσσσ/			Align(σ, L, Ft, L, σ)
a.		(σσ)σσσσσ	21!
b.		σ(σσ)σσσσ	16!
c.		σσ(σσ)σσσ	13!
d.	☞	σσσ(σσ)σσ	12
e.		σσσσ(σσ)σ	13!
f.		σσσσσ(σσ)	16!
g.		σσσσσσ(σ)	21!

Hyde uses this to argue for a family of constraints specifying the direction of misalignment, but simply categorically evaluating the Align constraint also removes the pathological prediction. In that case, each syllable which is not leftmost within the foot incurs precisely one violation regardless of its distance from the desired spot, yielding exactly six violations per candidate, and leaving it up to other (hopefully more plausible) constraints to determine the winner.

In response to these sorts of problems, McCarthy (2003) asserts that all OT constraints must be categorical, and proposes a family of categorical Anchor constraints to replace gradient Align constraints. Like Align constraints, Anchor constraints must specify two categories and the edges of each which are to align, as well as the intervening category by which violations will be assessed. The difference is that where Align constraints assign a violation mark for each segment/syllable/foot/etc which intervenes between two misaligned edges (multiple violations per locus), Anchor constraints assign one violation mark if the edges are misaligned by one or more instances of the intervening category (one violation per locus, regardless of severity). This approach captures the attested patterns without the undesirable side effects of gradient ALIGN. In this dissertation, any constraints used in the analysis in which the intervening category is null (such that any misalignment incurs violations) are referred to as ALIGN, but they are always evaluated categorically.

3.5. The Three-Syllable Stress Window

3.5.1. WINDOW Constraints

Hyde (2012) cites a number of languages which, like Wamesa, confine primary stress to the final three syllables of the word. These include Latin, Macedonian, Maithili, and Pirahã. Hyde proposes a constraint FINAL-WINDOW (alongside a corresponding INITIAL-WINDOW for languages requiring stress to fall within a certain distance of the left edge of the word), which, as he defines it, restricts stress to the final foot of the word or the syllable adjacent to it. The formal definition of FINAL-WINDOW, paraphrased from Hyde (2012) and formulated in terms of his *Relation-Specific Alignment* theory (RSA), is given in (3.7).

(3.7) FINAL-WINDOW: $*\langle x_\omega, F, \sigma \rangle / \; x_\omega ... \sigma ... F$

'Assess a violation mark for every syllable which intervenes between a primary stress and the left edge of a following foot.'

Hyde argues, *contra* McCarthy (2003), that distance-sensitive constraints, assigning more violation marks the further mis-aligned an edge is, are necessary, and this constraint is thus formulated as such. Though the constraint does not overtly specify that primary stress must fall on or be adjacent to the rightmost foot, it has that effect - given the configurational nature of the RSA, no violations are incurred when any number of feet precede primary stress, but if any more than a single foot follows the stress then one or more syllables will necessarily intervene in the prohibited configuration, each one constituting a violation.

This approach has the advantage that it does not require ternary feet to define the three-syllable domain; formally, FINAL-WINDOW simply assigns one violation mark for each syllable intervening between the stressed syllable and final foot of the word. This requires the presence of non-stress-bearing (covert) feet. Evidence for such feet has been presented for a number of languages, for example by Hayes (1995); Buckley (1992, 1997, 2009) for Kashaya; Crowhurst (1996) for Cairene Arabic; and González (2005, 2007) for Huariapano, Panoan languages, and a number of others.

There are, however, several drawbacks to the FINAL-WINDOW approach. The first is that unless we rule out strictly right-edge-adjacent, disyllabic (non-recursive) feet, it easily produces pre-antepenultimate stress patterns, unattested in natural language (Buckley 2013). This rules out instances of final extrametricality, an analytical construct Hyde argues in favor of elsewhere (Hyde 2011), and to which he refers here as well. The simplest counterexample, as in (3.8) below, is a language which shows regular left-aligned feet (ALIGNFOOTLEFT), with FINAL-WINDOW outranking MAIN-STRESS-LEFT, as Hyde proposes for Macedonian. Assuming that FT-BIN outranks PARSE, the final syllable of an odd-parity word will remain unfooted. The rightmost foot here encompasses the penultimate and antipenultimate syllables, so a stressed pre-antepenultimate syllable, with nothing intervening between it and the rightmost foot, will fail to incur any violations. Only stress which falls five or more syllables to the left of the Pword edge will run afoul of FINAL-WINDOW and be ruled out; any odd-parity word of five or more syllables will surface with pre-antepenultimate stress, while any even-parity word of sufficient length will have antepenultimate or pre-antepenultimate stress, depending on the language's preference for iambs or trochees. (The purely iambic pattern is shown below; a strong enough preference for trochees will cause these to surface except in long, odd-parity forms.)

(3.8) Left-aligned feet:
$(\sigma\text{ˈ}\sigma)$
$(\sigma\text{ˈ}\sigma)\sigma$

$(\sigma{}^{\prime}\sigma)(\sigma\sigma)$
$(\sigma{}^{\prime}\sigma)(\sigma\sigma)\sigma$
$(\sigma{}_{\scriptscriptstyle |}\sigma)(\sigma{}^{\prime}\sigma)(\sigma\sigma)$
$(\sigma{}_{\scriptscriptstyle |}\sigma)(\sigma{}^{\prime}\sigma)(\sigma\sigma)\sigma\ldots$

/σσσσσ/		Ft-Bin	Parse	Fin-Win	AlignFtLeft	Main-Str-Left
a.	('σσ)(σσ)σ		*	*!	**	
b.	('σσ)σσσ		**!*			
c. ☞	(σ'σ)(σσ)σ		*		**	*
d.	(σσ)('σσ)σ		*		**	**!
e.	(σσ)(σ'σ)(σ)	*!			******	***

While it is an advantage of the Final-Window approach that it does not require ternary feet to account for antepenultimate stress, it also rules out better-motivated trisyllabic structures, in particular recursive feet as discussed, for example, by Kager (2012); Martínez-Paricio & Kager (2013); Bennett (2012) and Martínez-Paricio (2012, 2013). Recursive feet, in which one branch of a binary foot is itself a foot, account elegantly for patterns such as English aspiration (see Martínez-Paricio 2013 §5.4 and the references therein) and are argued to be necessary to account for stress in Wargamay, Chugach Alutiiq (Martínez-Paricio 2012),[6] and Kashaya (Buckley 2013). A maximal foot with one level of recursion, while binary-branching at each level, ultimately encompasses three syllables; any language with such recursive feet at the right edge (Parse » NonRecursivity), as well as an active Final-Window constraint, would be predicted to have regular pre-

6. Martínez-Paricio also uses Yidiny to argue for recursive feet, but as Dixon's (1977, 1990) description of the facts of Yidiny stress, on which most subsequent work is based, has since been shown by Bowern, Alpher, & Round (2013, in progress) to be problematic, I ignore that aspect of her analysis here.

antepenultimate stress. This is shown in (3.9). Exhaustivity (Selkirk 1995), which bans level-skipping (i.e. a syllable directly dominated by a Pword without an intervening foot level), here replaces Parse-Syll.

(3.9) Final Recursive Feet:
(σˈσ)
((σˈσ)σ)
(σˈσ)(σσ)
(σˈσ)((σσ)σ)...

/σσσσσ/		Ft-Bin	Exhaust	F-Win	AlignFtL	Non-Recurs	Str-L
a.	(ˈσσ)((σσ)σ)			*!	**	*	
b. ☞	(σˈσ)((σσ)σ)				**	*	*
c.	(σˈσ)(σσ)σ		*!		**		*
d.	(σ(σˈσ))(σσ)				***!*	*	**
e.	(σσ)((ˈσσ)σ)				**	*	**!
f.	(σσ)(σˈσ)(σ)	*!			******		***

A similar problem arises unique to word-initial stress windows. As Hyde (2011) points out, the initial syllables of vowel-initial words in Western Aranda are unstressable. Goedemans (1996) attributes this to a requirement in W. Aranda that the left edge of a foot align with a consonant, leaving the word-initial vowel unfooted. (This is here formalized as the constraint Align(Ft, L, C, L).) A language W. Arandaˈ with this same requirement plus an active Initial-Window constraint (outranking Main-Stress-Right) would have regular stress on the fourth syllable of sufficiently long vowel-initial words, and on the post-peninitial syllable of consonant-initial forms.

(3.10) Unfootable Initial Vowels (W. Arandaˈ):

($C\sigma\sigma$)(σ'σ)

V($\sigma\sigma$)(σ'σ)

/VCVCVCVCV/		Aln(Ft, L, C, L)	Parse	Init-Win	AlnFtL	Str-R
a.	('V.CV)(CV.CV)(CV)	*!			******	****
b.	V('CV.CV)(CV.CV)		*		****	****!
c.	V(CV.'CV)(CV.CV)		*		****	***!
d.	V(CV.'CV)CV.CV		**!*		*	***
e. ☞	V(CV.CV)('CV.CV)		*		****	*
f.	V(CV.CV)(CV.'CV)		*	*!	****	

One could further imagine a language W. Aranda'' where strings of unfootable initial vowels force stress even farther rightwards in the word, though at a certain point the learnability, and thus the sustainability, of such a pattern comes into question, and this may well be a problem with Align(Ft, L, C, L) more than the Window constraint. Examples of this are given in (3.10) and (3.11).

(3.11) Strings of Unfootable Initial Vowels (W. Aranda''):

($C\sigma\sigma$)(σ'σ)

V.($\sigma\sigma$)(σ'σ)

V.V.($\sigma\sigma$)(σ'σ)...

/VVCVCVCVCV/		Aln(Ft, L, C, L)	Parse	Init-Win	AlnFtL	Str-R
a.	('V.V)(CV.CV)(CV.CV)	*!			******	*****
b.	V('V.CV)(CV.CV)(CV)	*!	*		*********	****
c.	V.V('CV.CV)(CV.CV)		**		******	***!
d.	V.V('CV.CV)CV.CV		***!*		**	**
e.	V.V(CV.'CV)(CV.CV)		**		******	**!
f. ☞	V.V(CV.CV)('CV.CV)		**		******	*
g.	V.V(CV.CV)(CV.'CV)		**	*!	******	

While syllable extrametricality is not as widely accepted word-initially as it is word-finally, Martínez-Paricio (2012) and Buckley (2013) argue that initial recursive feet are required to account for Wargamay, Chugash Alutiiq, and Kashaya, leading to the same prediction of the existence of four-syllable stress windows as encountered word-finally, as in (3.12). This is the precise mirror image of the problem encountered with final recursive feet.

(3.12) Initial Recursive Feet:
$(\sigma\text{ˈ}\sigma)$
$(\sigma(\sigma\text{ˈ}\sigma))$
$(\sigma\sigma)(\sigma\text{ˈ}\sigma)$
$(\sigma(\sigma\sigma))(\sigma\text{ˈ}\sigma)$...

/σσσσσ/		Ft-Bin	Exhaust	Init-Win	AlnFtR	Non-Recurs	Str-R
a.	(σ(σσ))(σˈσ)			*!	**	*	
b. ☞	(σ(σσ))(ˈσσ)				**	*	*
c.	σ(σσ)(ˈσσ)		*!		**		*
d.	(σσ)((σˈσ)σ)				***!*	*	**
e.	(σ(σˈσ))(σσ)				**	*	**!
f.	(σ)(ˈσσ)(σσ)	*!			******		***

Even in a language where binary, right-aligned feet combined with Final-Window make for a well-behaved three-syllable stress window, there is a further undesirable, though non-fatal, result. In languages like Wamesa and Macedonian, where stress may be lexically assigned to any one of the final three syllables (rather than consistently and by default surfacing on the antepenult), the language will be inconsistent as to the headedness of its feet, as shown below in (3.13). (A superscript *h* on the forms marks a syllable underlyingly marked as the head of a syllable which will surface with stress; this assumption will be discussed in more depth in §3.6.1.) Given right-aligned feet, penultimately-stressed words will always be trochaic, while words with antepenultimate or final stress require iambs. For some languages[7] inconsistent directionality of foot heads may be a desirable result. However for others, such as those with a clear contrast in intensity or duration between head and dependent syllables, this may be an undesirable result, as the vast majority languages are consistent (Prince 1990), and iambs and trochees are known to have different (though not absolute) preferences in regards to these features, as laid out

7. Such as, reportedly, Huariapano (Bennett 2012), Nuuchahnulth (Lee 2008), Akkadian (Kager 1994), and Guahibo (Kondo 2001); see Bennett (2012: 62) for more examples and references.

in Hayes' (1985, 1995) Iambic-Trochaic Law.

(3.13) *Inconsistent Footing under* FINAL-WINDOW

a. Penultimate Lexical Stress ⟶ Trochees:

$(_{ˌ}\sigma\sigma)(_{ˌ}\sigma\sigma)(^{ˈ}\sigma\sigma)$

b. Antepenultimate or Final Lexical Stress ⟶ Iambs:

$(\sigma_{ˌ}\sigma)(\sigma_{ˌ}\sigma)(\sigma^{ˈ}\sigma)$

$(\sigma_{ˌ}\sigma)(\sigma^{ˈ}\sigma)(\sigma\sigma)$

Based on these pathologies, it seems prudent to exclude FINAL-WINDOW from CON. Other solutions have been posited to account for final stress windows, including NON-FINALITY (Prince & Smolensky 1993/2004), *EXTENDED-LAPSE-RIGHT (Gordon 2002; Kager 2005), and WEAK-LOCAL-PARSING (Kager 1994; Green 1995; Green & Kenstowicz 1995). Hyde states that these approaches are in fact inadequate for languages such as Macedonian and Maithili. A full analysis of window effects in these languages is beyond the scope of this dissertation, and the answer may well turn out to be different for each language, depending on its particular idiosyncrasies of stress placement (i.e. lexical exceptions in Macedonian, weight/stress interactions in Maithili, etc.). In Wamesa, the stress window is the result of the interaction of several constraints, most importantly LAPSE-AT-END and END-RULE-R.

3.5.2. The Stress Window in Wamesa

The three-syllable stress window found in Wamesa does not require a WINDOW constraint to produce the attested output; its presence can be explained by the confluence of two facts: 1) Wamesa prefers any lapses to occur adjacent to the right word-edge, and 2) secondary stress in Wamesa always appears to the left of the primary stress. These generalizations can be seen throughout the lexicon; words with antepenultimate stress always have a final lapse, but any potential lapses preceding primary stress are filled by a secondary stress. That this is an active constraint in the language and not only a statistical or historical accident is shown by words like *àparápiri*, *mìmiótari* 'snakehead fish', and *kòmamásare* 'funny', where clitic-induced stress shift is blocked because it would either create a word-initial lapse (**aparàpirí=wa*) or require a second secondary stress to fill the lapse (**àparàpirí=wa*), which is prohibited.

This preference for word-final lapses is the result of a Lapse-at-End constraint, defined below. This explains the size of the window: a two-syllable lapse at the right word edge, as occurs when the stress is three syllables in, incurs no Lapse-at-End violations, while moving the stress any farther leftwards creates a violation-inducing three-syllable lapse.

(3.14) Lapse-at-End: Assign one violation for every sequence of two adjacent unstressed syllables which is not adjacent to the right word edge (i.e. σ̆σ̆σ) (Kager 2001, 2005).

The fact that secondary stress in Wamesa always temporally precedes primary is enforced by an End-Rule-R constraint (Prince 1983; McCarthy 2003; Kager 2004), banning any other feet from appearing to the right of the head foot. This ensures that it is the primary stress which consistently surfaces within the final window. Placement of secondary stress within the final three syllables of the word would avoid a Lapse-at-End-violating trisyllabic lapse just as effectively as would a primary stress, but would entail placing another foot to the right of the head foot, violating End-Rule-R. The ordering of secondary stress to the left of primary stress means that it will never be in position to appear within a word-final stress window.

(3.15) End-Rule-R: Assign one violation mark for every foot which follows the head foot within the Pword.

These two constraints combine to give us the final trisyllabic stress window found in Wamesa and elsewhere. Their interaction is demonstrated in tableau (3.16). Any of candidates (a) - (c), with primary stress within the final trisyllabic stress window, can win under these constraints; which one ultimately prevails will depend on the given language's preferences for foot type, foot alignment, and exhaustiveness of parsing. Candidates (d) and (e), with primary stress outside the window, are ruled out by virtue of having a non-final lapse or secondary stress following primary stress.

(3.16)

		/σσσσ/	End-Rule-R	Lapse-at-End
a.	☞	(σˌσ)(σˈσ)		
b.	☞	(ˌσσ)(ˈσσ)		
c.	☞	(σˈσ)σσ		
d.		(ˈσσ)(ˌσσ)	*!	
e.		(ˈσσ)σσ		*!

The above pattern also explains why there should be no constraint Extended-Lapse-at-End, penalizing trisyllabic-or-longer lapses not adjacent to the right edge of the word, admitted into Con. Constraints functionally equivalent to *Extended-Lapse, violated by strings of three or more unstressed syllables, have been proposed under various names by a number of authors (Elenbaas & Kager 1999; Gordon 2002; Houghton 2006; Kager 2007), generally to account for ternary stress patterns. Houghton (2006) proposes Extended-Lapse-at-End on analogy with Lapse-at-End as part of her analysis of Tripura Bangla. However, a language with the same constraint rankings as Wamesa, but with Extended-Lapse-at-End in place of Wamesa's Lapse-at-End, would have a four-syllable stress window word-finally, for the same reasons that LAPSE-AT-END creates a three-syllable window in Wamesa.

Under this analysis, stress which is lexically assigned to any of the rightmost three syllables of a lexical word should surface precisely there. Stress lexically assigned earlier in the word will surface on either the penultimate or antepenultimate syllable, depending on the parity of its original location.[8] Here I assume faithfulness to headedness rather than to underlying stress, after McCarthy & Pruitt (2013), discussed in more detail below. This has several advantages over the WINDOW approach: it avoids predicting the existence of languages with four-syllable stress windows; it allows Wamesa to be a strictly trochaic language; and it relies on constraints which are necessary elsewhere in the language to account for the stress shift pattern found on words with antepenultimate stress, as discussed in §3.7.

3.6. Placing Wamesa Stress

I assume that Wamesa has consistent trochaic footing, regardless of the location of stress in a given word. The primary evidence for this lies in the fact that penultimate stress is by far the most common pattern in the language, accounting for approximately two thirds of the lexicon, while the remaining third is divided between final and penultimate main stress. While it is certainly possible to account for this data using iambs, a trochaic analysis incorporating degenerate feet is more parsimonious. Given the set of constraints and ranking laid out here, penultimate stress falls out as the default for words not otherwise specified in the lexicon, accounting for the skew found in the data. This is demonstrated in §3.6.6. That degenerate feet are allowed in Wamesa is demonstrated by monosyllabic words, such as *nú* 'island', *rón* 'ironwood', and *át* 'four', which do bear

8. Though underlying stress earlier in the word is unsustainable in the language, as the learner will never see it surface anywhere other than in the final three syllables and have no evidence of its origin elsewhere, it must be accounted for under Richness of the Base.

primary stress. A degenerate foot will be built on a multisyllabic word in case of final stress, otherwise disyllabic feet will be built leftwards from the primary stress. Feet to the right of the head foot again are banned by End-Rule-R. Examples of the footing assumed here are given below.

(3.17)
a. *(nú)* 'island'
b. *(kó.ta)* 'also'
c. *kam(bú)* 'water'
d. *(áng.ga)di* 'coconut'
e. *pi(mú.na)* 'pig'
f. *(à.ka)(nák)* 'breadfruit'
g. *na(ná.ri)a* 'slowly'
h. *(kò.rom)(bó.wi)* 'cowrie shell'
i. *va(wà.ta)(tár)* 'naughty'
j. *(kò.ma)(má.sa)re* 'funny'
k. *ma(rì.si)(á.ni)* 'spicy'

Locating primary stress in monomorphemic Wamesa words of five or fewer syllables requires six active constraints: Parse-Syllable, Foot-Binarity, *Lapse, End-Rule-R, Trochee, and h⟶Head (McCarthy & Pruitt 2013). Parse-Syllable (Liberman & Prince 1977; Prince 1980, etc.), as defined below in (3.18), derives from the Strict Layer Hypothesis of Selkirk (1984) and Nespor & Vogel (1986), in effect prohibiting syllables from being directly dominated by a Pword with no intervening foot layer. The second of these, Foot-Binarity, is standard in most OT theories of foot-based stress and traceable back to Prince (1980), Kager (1989), and others; it requires that all feet be binary-branching. Since Wamesa is not quantity-sensitive - there is no contrastive vowel length or gemination, and closed syllables and diphthongs do not deterministically attract stress - this must be calculated in terms of syllables, not moras. *Lapse (Prince 1983; Selkirk 1984; Kager 2001, *inter alia*), another by-now-standard constraint, bans adjacent unstressed syllables. In the case of words of the length considered so far, its primary function is to ensure that all feet are stress-bearing, since higher-ranked Parse will disallow strings of unfooted syllables in pre-stress position.

(3.18) Parse-Syllable: Assign one violation-mark for each syllable not parsed into a foot.

(3.19) Foot-Binarity: Assign one violation-mark for each foot which is not binary-branching in structure (McCarthy & Prince 1993).

(3.20) *Lapse: Assign one violation-mark for each pair of adjacent unstressed syllables (McCarthy & Prince 1993).

End-Rule-R (McCarthy 2003, adapted from Prince 1983) requires that primary stress fall on the final foot of the word. Depending on its ranking with regards to Parse-Syll, this can result either in the primary stress being attracted to the right word edge, or in a failure of syllables following the head foot to themselves be parsed into a foot. In Wamesa, its main effect is to ensure that secondary stress, where it occurs, is always to the left of primary stress.

Trochee will become relevant later in the analysis, in §3.6.3. It avoids the creation of iambic feet, enforcing consistency in the directionality of foot heading throughout the Wamesa lexicon. The typological evidence for trochees vs. iambs in Wamesa is mixed: its insensitivity to weight suggests a trochaic analysis (Hayes 1985), while the realization of stress primarily as length hints at an iambic one (Prince 1990). That the majority of the lexicon is composed of disyllabic words with stress on the first syllable, however, tips the scales in favor of trochees, which allow a more parsimonious analysis.

3.6.1. Motivating Faithfulness to Heads: McCarthy & Pruitt (2013)

The fourth constraint, h→Head, requires somewhat more explanation. McCarthy & Pruitt (2013) propose this constraint, which enforces faithfulness to a syllable's underlyingly indexed status as foot head, as a solution to the problem of lexical exceptions to predictable stress within the framework of Harmonic Serialism (HS). Though their motivation for proposing the *h* and *d* diacritics is particular to the assumptions of HS, particularly the non-revisability of feet once built, this machinery will prove useful in the Wamesa analysis as an alternative to constraints requiring faithfulness to underlying stress.

Faith(Stress) constraints are often used in the literature but rarely formally defined; many of those who do give definitions, for example McCarthy (1995), assume that faithfulness enforces not just the location of a stress but also its status as primary or secondary (though McCarthy for one does not address secondary stress at all). This does not necessarily need to be the case; Revithiadou (1999) uses a constraint Max(LA) which, as defined, simply says that every lexical accent present in the input must have a correspondent in the output, and more specific constrains are proposed targeting head vs, non-head accents. Use of this primacy-non-specific approach would work equally as well in Wamesa as the *h* diacritic does. In fact, in combination with a high-ranked Express-Head-as-Stress constraint, McCarthy & Pruitt's *h* diacritic has exactly the same result as such a Faith(Stress) constraint. A Faith(Stress) constraint which enforces degree of stress as well as location creates a ranking paradox; this is discussed in detail in §3.7.6, after the full constraint hierarchy has been laid out.

One additional complication is that if stress is taken to be a structural property of the foot rather than an autosegmental feature associated with the vowel, as it often is (Liberman & Prince 1977, among others), then underlying stress entails the presence of underlying feet. Richness of the Base supposes that, if underlying feet are allowed at all, they may be either trochaic or iambic, regardless of the surface preferences of the language. Though I assume here, for uniformity's sake (Prince 1990), that Wamesa allows only trochees to surface, the possible presence of underlying iambs is not a problem in parallel OT; a highly-ranked TROCHEE constraint will reconfigure underlying iambs int trochees as needed. While an analysis based on FAITH(Stress) is certainly possible, I find the diacritic approach to be more parsimonious, as it entails less underlying metrical structure, and I will follow that avenue here.

McCarthy & Pruitt's argumentation on behalf of headedness diacritics runs as follows: Many languages of the world have stress which is predictable for most lexical items, but maintain a few exceptions to the wider pattern. A standard assumption in the metrical literature is that, unlike those forms with predictable stress, these exceptional words have enough metrical structure included in their lexical entries to account for their unpredictable stress patterns. McCarthy & Pruitt take the view, after i.e. Liberman & Prince (1977), that stress is defined over the foot, and thus stress cannot exist on a syllable in the absence of foot structure; therefore underlying stress requires underlying footing. If underlying footing can exist in some languages, then under the assumptions of Richness of the Base the grammar of a language with entirely predictable stress and no lexical exceptions must be able to derive the attested stress pattern from an underlying form with foot structure anywhere in the word.

Under the assumptions of Harmonic Serialism (McCarthy 2000, 2002, 2007), this is problematic. Unlike in classic, parallel OT as originally proposed by Prince & Smolensky (1993/2004), in HS GEN may only make one change at a time to the input form, for example adding or deleting a foot. The winner of each round of evaluation becomes the input for the following round, until the fully faithful candidate is chosen as optimal and the derivation converges. In this framework, the winning candidate at each successive step of the derivation is more harmonic than the last. The authors give the example of

Pintupi, a Pama-Nyungan language of Australia, which has exceptionless initial stress, with secondary stress on alternating syllables thereafter. Final syllables in Pintupi are never stressed. Assuming that stress never needs to be part of the lexical entry for a Pintupi form, the attested patterns can be derived by the interaction of three constraints: Ft-Bin and Parse-Syllable (as defined above), and Align-Left(foot, word) (McCarthy & Prince 1993), ranked in just that order. Though they leave this out of their tableaux, an undominated constraint enforcing trochaic foot structure is also necessary.

(3.21) Align-Left(foot, word): For each foot in a word assign one violation mark for every syllable separating it from the left edge of the word (McCarthy & Pruitt 2013: 113).

In HS, the process of stress assignment to an underlying form such as /puɭiŋkalat^{j}u/ 'we (sat) on a hill' has three steps. First, a foot is built on the the first and second syllables of the word, removing two Parse-Syll violations from the fully faithful candidate while failing to incur any others. Next a second foot is built around the third and fourth syllables, removing two additional Parse-Syll violations at the expense of one violation of lower-ranked Align-Left(ft). In the third step the derivation converges on the faithful candidate, with undominated Ft-Bin blocking footing of the final syllable. The tableaux illustrating this derivation are given in (3.22).

(3.22) *HS Stress Assignment in Pintupi* (McCarthy & Pruitt 2013)

a. *Step 1: Footing*

/pu.ɭiŋ.ka.la.tʲu/		Ft-Bin	Parse	Align-Left(ft)
a. ☞	(ˈpu.ɭiŋ)ka.la.tʲu		***	
b.	pu.ɭiŋ.ka.la.tʲu		****!* W	
c.	pu.(ˈɭiŋ.ka)la.tʲu		***	*! W
d.	(ˈpu)ɭiŋ.ka.la.tʲu	*! W	**** W	

b. *Step 2: Footing Continued*

(ˈpu.ɭiŋ)ka.la.tʲu		Ft-Bin	Parse	Align-Left(ft)
a. ☞	(ˈpu.ɭiŋ)(ˈka.la)tʲu		*	**
b.	(ˈpu.ɭiŋ)ka.la.tʲu		**!* W	L
c.	(ˈpu.ɭiŋ)ka(ˈla.tʲu)		*	***! W
d.	(ˈpu.ɭiŋ)(ˈka)la.tʲu	*! W	** W	**
e.	pu.ɭiŋ.ka.la.tʲu		**!*** W	L

c. *Step 3: Convergence*

(ˈpu.ɭiŋ)(ˈka.la)tʲu		Ft-Bin	Parse	Align-Left(ft)
a. ☞	(ˈpu.ɭiŋ)(ˈka.la)tʲu		*	**
b.	(ˈpu.ɭiŋ)(ˈka.la)(ˈtʲu)	*! W	L	****** W
c.	(ˈpu.ɭiŋ)ka.la.tʲu		**!* W	L
d.	pu.ɭiŋ(ˈka.la)tʲu		**!* W	**

The ranking of these three constraints is made clear in Tableau (3.22c): Ft-Bin must outrank Parse to rule out candidate (b), and Parse in turn must dominate Align-Left(ft) to rule out candidate (c) (as well as candidates (b) and (e) in step two). As seen with candidate (e) in Tableau (3.22b) and candidates (c) and (d) in Tableau (3.22c), it will never be harmonically improving within a grammar like that of Pintupi to delete a binary foot which is present in the input. A lexically-assigned degenerate foot will be removed, because Ft-Bin outranks Parse, but the only constraint which might oppose a binary foot, Align-Left(ft), will be overruled by the higher-ranked Parse, which will always prefer the form with the highest number of footed syllables, regardless of the location or form of those feet.

Therein lies the problem. If our theory allows stress (and therefore foot structure) to be present underlyingly in any language, under Richness of the Base, we must admit the possibility of underlying forms in any language with underlying footing incompatible with that attested in the surface forms. The grammar of Pintupi, then, must be able to take as input a form with underlying footing of the second and third syllables, and, as there are no attested lexical exceptions to Pintupi stress, return a surface form with regular initial stress. But with the given constraint ranking, which is necessary to produce the regular pattern, this cannot take place, as demonstrated in (3.23). Because Parse outranks Align, it will not be harmonically improving to remove the mis-aligned underlying foot, and with that foot in place Ft-Bin will block footing of the initial syllable, which should bear primary stress. Simply moving the offending foot leftward into initial position is not a licit move, as it involves deleting and then adding back the foot - two derivational steps. The output then is one which is totally unattested in the language, with primary stress on the second syllable of the word, even in the absence of any active constraint enforcing faithfulness to underlying stress.

(3.23) *Footing with Irregular Underlying Stress in Pintupi*

a. *Step 1: Footing*

	/pu('ɭiŋ.ka)la.tʲu/	Ft-Bin	Parse	Align-Left(ft)
a.	☞ pu('ɭiŋ.ka)('la.tʲu)		*	****
b.	pu('ɭiŋ.ka)la.tʲu		**!*	*
c.	pu.ɭiŋ.ka.la.tʲu		**!***	
d.	('pu)('ɭiŋ.ka)la.tʲu	*!	**!	*

b. *Step 2: Convergence*

	pu('ɭiŋ.ka)('la.tʲu)	Ft-Bin	Parse	Align-Left(ft)
a.	☞ pu('ɭiŋ.ka)('la.tʲu)		*	****
b.	pu('ɭiŋ.ka)la.tʲu		**!*	*
c.	pu.ɭiŋ.ka('la.tʲu)		**!*	***
d.	('pu)('ɭiŋ.ka)('la.tʲu)	*!		****

To fix this problem within HS, either Richness of the Base or underlying stress must be discarded, and McCarthy & Pruitt choose the latter. Some mechanism is still necessary, however, to produce lexical exceptions to regular stress patterns in languages which allow them, and with faithfulness to Faith(stress) no longer an option they instead propose h⟶Head and its opposite, d⟶Dependent. Rather than underlying foot structure determining exceptional stress, here they build on the use of diacritic marking in accentual systems (Goldsmith 1976 *et seq*; Haraguchi 1977; Hyman 1981, 1982; *inter alia*). Under this analysis, unpredictably stressed syllables (or the vowels which comprise their heads) are underlyingly marked with a phonetically uninterpretable diacritic *h* or *d*. The constraint h⟶Head prefers that *h*-marked vowels are parsed into the head syllable of a foot; d⟶Dependent prefers that *d*-marked vowels are parsed into the dependent syllable.

(3.24) *Headedness Faithfulness Constraints (McCarthy & Pruitt 2013: 128)*

a. h⟶HEAD: Assign one violation mark for every *h*-bearing segment that is not in the head syllable of a foot.

b. d⟶DEPENDENT: Assign one violation mark for every *d*-bearing segment that is not in the dependent syllable of a foot.

These are violable constraints like any other, so a language like Pintupi which admits no lexical exceptions to stress will rank them low, and they may be violated when the feet they prefer would offend a higher-ranked constraint, as in (3.25), which proceeds exactly as did the derivation in (3.22) above, despite the distribution of diacritics equivalent to the underlying footing in (3.23). Because there are no feet to erase and rebuild - the diacritics are present in the surface form, but phonetically null - the problems which come up with underlying stress assignment do not arise here.

(3.25) *Stress Diacritics in Pintupi: No Lexical Exceptions*

a. *Step 1: Footing*

/pu.ɭiʰŋ.kaᵈ.la.tʲu/	FT-BN	PARSE	ALN-LEFT(ft)	h⟶HD	d⟶DEP
a. ☞ (ˈpu.ɭiʰŋ)kaᵈ.la.tʲu		***		*	*
b. pu.ɭiʰŋ.kaᵈ.la.tʲu		****!* W		*	*
c. pu.(ˈɭiʰŋ.kaᵈ)la.tʲu		***	*! W	L	L
d. (ˈpu)ɭiʰŋ.kaᵈ.la.tʲu	*! W	**** W		*	*

b. *Step 2: Footing Continued*

(ˈpu.ɭiʰŋ)kaᵈ.la.tʲu	Ft-Bn	Parse	Aln-Left(ft)	h⟶Hd	d⟶Dep
a. ☞ (ˈpu.ɭiʰŋ)(ˈkaᵈ.la)tʲu		*	**	*	*
b. (ˈpu.ɭiʰŋ)kaᵈ.la.tʲu		**!* W	L	*	*
c. (ˈpu.ɭiʰŋ)kaᵈ(ˈla.tʲu)		*	***! W	*	*
d. (ˈpu.ɭiʰŋ)(ˈkaᵈ)la.tʲu	*! W	** W	**	*	*
e. pu.ɭiʰŋ.kaᵈ.la.tʲu		**!*** W	L	*	*

c. *Step 3: Convergence*

(ˈpu.ɭiʰŋ)(ˈkaᵈ.la)tʲu	Ft-Bn	Parse	Aln-Left(ft)	h⟶Hd	d⟶Dep
a. ☞ (ˈpu.ɭiʰŋ)(ˈkaᵈ.la)tʲu		*	**	*	*
b. (ˈpu.ɭiʰŋ)(ˈkaᵈ.la)(ˈtʲu)	*! W	L	****** W	*	*
c. (ˈpu.ɭiʰŋ)kaᵈ.la.tʲu		**!* W	L	*	*
d. pu.ɭiʰŋ(ˈkaᵈ.la)tʲu		**!* W	**	*	*

A language like Pintupi which allows exceptional forms will rank them higher, allowing their preferred footings to surface, as in (3.26), which shows a language just like Pintupi but with lexical exceptions to stress. This derivation proceeds along the same lines as did that in (3.23), but with an additional first step of building the foot which there was underlying.[9]

(3.26) *Stress Diacritics in Pintupi′: Lexical Exceptions Allowed*

a. *Step 1: Footing*

9. Either diacritic individually would be sufficient here to produce the mis-aligned foot; both are included for purposes of illustration.

/pu.ɭiʰŋ.kaᵈ.la.tʲu/			h→Hd	d→Dep	Ft-Bn	Parse	Aln-Left(ft)
a.	☞	pu(ˈɭiʰŋ.kaᵈ)la.tʲu				***	*
b.		pu.ɭiʰŋ.kaᵈ.la.tʲu	*! W	* W		***** W	L
c.		(ˈpu.ɭiʰŋ)kaᵈ.la.tʲu	*! W	* W		***	L

b. *Step 2: Footing Continued*

pu(ˈɭiʰŋ.kaᵈ)la.tʲu			h→Hd	d→Dep	Ft-Bn	Parse	Aln-Left(ft)
a.	☞	pu(ˈɭiʰŋ.kaᵈ)(ˈla.tʲu)				*	****
b.		pu(ˈɭiʰŋ.kaᵈ)la.tʲu				**!* W	* L
c.		pu.ɭiʰŋ.kaᵈ.la.tʲu	*! W	* W		*****W	L
d.		(ˈpu)(ˈɭiʰŋ.kaᵈ)la.tʲu			*! W	** W	* L

c. *Step 3: Convergence*

pu(ˈɭiʰŋ.kaᵈ)(ˈla.tʲu)			h→Hd	d→Dep	Ft-Bn	Parse	Aln-Left(ft)
a.	☞	pu(ˈɭiʰŋ.kaᵈ)(ˈla.tʲu)				*	****
b.		pu(ˈɭiʰŋ.kaᵈ)la.tʲu				**!* W	* L
c.		pu.ɭiʰŋ.kaᵈ(ˈla.tʲu)	*! W	* W		*** W	*** L
d.		(ˈpu)(ˈɭiʰŋ.kaᵈ)(ˈla.tʲu)			*! W		****

3.6.2. Placing Stress on Shorter Words

The inability of the system to produce both lexically-determined and exceptionlessly predictable stress patterns under Richness of the Base does not apply to classic parallel OT, as used in this analysis. Since Gen in this framework may make more than one change at a time, shifting an offending foot in an exceptionless language does not pose a problem. The proposed constraints, however, are still of use. Some faithfulness-like constraint is necessary in order for lexical stress preferences to have any effect at all on surface forms. As will become clear later in the analysis, making use of a Faith(stress) constraint in Wamesa actually leads to a ranking paradox in cases of stress shift encoded by an enclitic. This will be illustrated in detail in §3.7.6, when the stress shift facts and the constraints needed to account for them have been more fully elaborated. For now, I will assert that h⟶Head is a better choice to fill that role; the derivations below will demonstrate that it produces the desired output.

Wamesa ranks h⟶Head high in its grammar; the constraint is undominated with respect to the others considered here. Its effect can be seen in odd-parity antepenultimately-stressed words, where adherence to diacritically-marked headedness leads to violations of *Lapse, which must therefore be dominated by h⟶Head. These forms also give a ranking argument for Ft-Bin. Wamesa could retain underlying headedness and avoid lapse by fully parsing the word with one disyllabic and one degenerate foot, both bearing stress; that it fails to do so shows that Ft-Bin crucially outranks Parse and *Lapse.

(3.27) a. *Antepenultimate Stress on Trisyllabic Roots*

/siʰ.ni.tu/			h→Head	Ft-Bin	Parse	End-Rule-R	*Lapse
a.		siʰ.ni.tu	*! W		*** W		** W
b.	☞	(ˈsiʰ.ni)tu			*		*
c.		(ˌsiʰ)(ˈni.tu)		*! W	L		L
d.		(ˌsiʰ.ni)(ˈtu)		*! W	L		L
e.		siʰ(ˈni.tu)	*! W		*		L
f.		(ˈsiʰ.ni)(tu)		*! W	L	* W	*
g.		(ˈsiʰ)(ni.ˌtu)		*! W	L	* W	*

The same ranking of constraints also produces the attested output on disyllabic words, as in (3.28a), and monosyllabic words, as in (3.28b). For reasons of space, the constraint End-Rule-R will be omitted from the remaining tableaux in this section, as its only function so far is to decide the relative order of primary and secondary stress, not their location or the structure of the feet on which they fall. (Preventing the footing of post-stress syllables is accomplished independently of End-Rule-R by Ft-Bin.) Penultimately-stressed disyllabic roots like *kóta* 'also' in (3.28a) are the most common in the language, and presumably most of these have stress placed there by default rather than by diacritic. However, under Richness of the Base we must be able to account for cases where the diacritic is indeed placed in this position, though presumably this would be a redundant and historically ephemeral lexical representation.

(3.28) a. *Penultimate Stress on Disyllabic Roots*

/koʰ.ta/	h→Head	Ft-Bin	Parse	*Lapse
a. koʰ.ta	*! W		** W	* W
b. ☞ (ˈkoʰ.ta)				
c. (ˈkoʰ)ta		*! W	* W	

b. *Stress on Monosyllables*

/nu/	h→Head	Ft-Bin	Parse	*Lapse
a. nu		*	*! W	
b. ☞ (ˈnu)		*		

3.6.3. Trochee

To account for instances of antepenultimate stress on four-syllable (or longer even-parity) words, as well as penultimate and final stress on trisyllabic words, we must introduce one additional constraint: Align(Foot, L, Headσ, L), aka Trochee.

(3.29) Align(Foot, L, Headσ, L) (Trochee): Assign a violation mark for every foot whose leftmost sylable is not the head syllable. McCarthy & Prince (1993: 12)

One important feature of this formulation, as with any Align-style version of Trochee, is that it does not ban degenerate feet; (ˈσ), even if that syllable is light,[10] satisfies its definition as well as (ˈσσ) does, as in both cases the left edge of the head syllable is aligned with the left edge of the foot. It is only (σˈσ) and headles (σσ) which violate it.

10. Keep in mind that stress in Wamesa is not weight-sensitive; heavy/closed syllables are statistically more likely to bear stress but not deterministically so, as demonstrated by counterexamples such as *kam.bú* ‘water’.

(3.30) *Antepenultimate Stress on Longer Roots*

a. *Even Parity*

		/ma.raʰ.re.a/	h→HEAD	TROCHEE	FT-BIN	PARSE	*LAPSE
a.		ma.raʰ.re.a	*! W			**** W	*** W
b.	☞	ma(ˈraʰ.re)a				**	*
c.		(ma.ˈraʰ)re.a		*! W		**	*
d.		(ma.ˌraʰ)(re.ˈa)		*!* W		L	L
e.		(ˌma.raʰ)(ˈre.a)	*! W			L	L

b. *Odd Parity*

		/a.pa.raʰ.pi.ri/	h→HEAD	TROCHEE	FT-BIN	PARSE	*LAPSE
a.		a.pa.raʰ.pi.ri	*! W			***** W	**** W
b.	☞	(ˌa.pa)(ˈraʰ.pi)ri				*	*
c.		a.(pa.ˌraʰ)(pi.ˈri)		*! W		*	*
d.		a(ˌpa.raʰ)(ˈpi.ri)	*! W			*	L
e.		(ˌa.pa)(ˌraʰ.pi)(ˈri)			*! W	L	L

Without TROCHEE, there is no way prevent the unattested candidate (d) [(ma.rà)(re.á)] in Tableau (3.30a) from winning over the attested form [ma(rá.re)a], here candidate (b). The attested candidate is in fact harmonically bounded by unattested (d) - as well as phonetically identical but structurally different (c) - in the absence of TROCHEE. This candidate pair also gives us the first ranking argument for TROCHEE, which must dominate PARSE-SYLL and *LAPSE. Using TROCHEE also allows attested [(ˌa.pa)(ˈra.pi)ri], candidate (b), to win over unattested [a.(pa.ˌraʰ)(pi.ˈri)], candidate (c), in Tableau (3.30b), and permits us to distinguish between candidates (b) and (c) in Tableau (3.30a), which have different footing but the same stress pattern, making them phonetically identical in Wamesa. TROCHEE similarly allows us to choose between phonetically identical but structurally different surface forms for trisyllabic words with stress on either of their rightmost two syllables, as in candidates (b) and (c) in each of Tableaux (3.31) and (3.32).

Candidate (e) in Tableau (3.30a) also demonstrates that h⟶HEAD must dominate PARSE-SYLL as well as *LAPSE.

(3.31) *Final Stress on Trisyllabic Roots*

		/a.ri.riʰ/	h⟶HEAD	TROCHEE	FT-BIN	PARSE	*LAPSE
a.		a.ri.riʰ	*! W			*** W	* W
b.	☞	(ˌa.ri)(ˈriʰ)			*		
c.		(ˌa)(ri.ˈriʰ)		*! W	*		
d.		(ˈa.ri)riʰ	*! W		L	* W	* W
e.		a(ˈri.riʰ)	*! W		L	* W	

The above tableau also gives us reason to rank h⟶HEAD above FT-BIN, in order to rule out the two unattested candidates (d) and (e), with main stress on the antepenultimate and penultimate syllables, respectively.

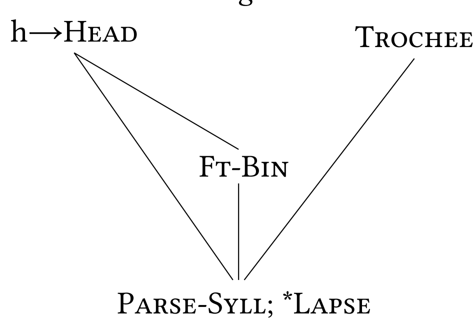
(3.32) *Penultimate Stress on Trisyllabic Roots*

	/pi.mu^h.na/	h→HEAD	TROCHEE	FT-BIN	PARSE	*LAPSE
a.	pi.mu^h.na	*! W			*** W	** W
b. ☞	pi(ˈmu^h.na)				*	
c.	(pi.ˈmu^h)na		*! W		*	
d.	(ˌpi.mu^h)(ˈna)	*! W		* W	L	
e.	(ˌpi)(mu^h.ˈna)	*! W	* W	* W	L	

3.6.4. Intermediate Summary and Constraint Ranking

So far we have accounted for the placement of stress in forms where one of the final three syllables is underlyingly marked for its status as head of a syllable. In terms of rankings, we so far have three tiers of constraints. A Hasse diagram laying out these dominance relationships is given below in (3.33). END-RULE-R is omitted here, as it is never surface-violated and therefore unrankable with respect to any of the constraints used here.

(3.33) *Constraint Rankings*



3.6.5. Longer Words

So far we have only addressed words of three syllables or less, in which stress will necessarily fall within the three-syllable stress window, and those of four and five syllables where headedness is specified for a syllable within that window, as in the tableaux in (3.30) above. We must also, however, consider underlying forms in which headedness may be specified outside of the stress window. In these words, whether primary stress falls on the penultimate or antepenultimate syllable will depend on the parity of syllable bearing the headedness diacritic - that syllable will be footed as the head of a trochee, with further disyllabic trochees built adjacent to that one towards the right word edge.

Here I will reintroduce End-Rule-R (abbreviated ER-R) into the following tableaux, as it illustrates why main stress rather than secondary stress surfaces within the final three syllables, though it is so far still unrankable, as it never uniquely conflicts with other constraints. For longer words with headedness specified on the pre-antepenultimate syllable - to the left of the stress window - the constraints already discussed are sufficient to produce the desired output. Tableau (3.34) below demonstrates this for four-syllable words; Trochee, in combination with h⟶Head, will ensure the same pair of right-aligned feet in a five-syllable word, with a single unfooted syllable at the left edge of the word.

(3.34) *Longer Words: Penultimate Stress*

/σʰσσσ/			h→Head	Trochee	Ft-Bin	ER-R	Parse	*Lapse
a.		σʰσσσ	*! W				*** W	**** W
b.	☞	(ˌσʰσ)(ˈσσ)						
c.		(ˈσʰσ)(ˌσσ)				*! W		
d.		(ˈσʰσ)σσ					*!* W	** W
e.		σʰσ(ˈσσ)	*! W				** W	* W

When a five-syllable word bears the *h* diacritic on the initial syllable, however, the current set of constraints is insufficient to decide between the desired form with regular alternating stress on the first and third syllable and a final unfooted syllable, represented by candidate (b) in Tableau (3.36) below, and a candidate with the unfooted syllable intervening between the two feet and apparent ternary stress on the first and fourth syllables, represented by candidate (d).[11] A categorical Align or Anchor constraint tying each foot to the left edge of the word won't fix the problem, as both candidates have one foot misaligned a foot or more; without gradiency these will each only incur a single violation and thus fail to be differentiated (though this constraint will prove useful later in the analysis). Instead, we return to the Lapse-at-End constraint discussed in §3.5.2. Its definition is reproduced in (3.35).

(3.35) Lapse-at-End: Lapse must be adjacent to the right edge (Kager 2001, 2005).

Because these two candidates differ in the location of the lapse which they contain, Lapse-at-End will prefer the one in which it is word-final, allowing the candidate with

11. Candidate (c) in (3.36) is phonotactically acceptable in Wamesa, but would not be expected as the output of this underlying form.

regular alternating stress to win out. It is so far unrankable with respect to the other constraints.

(3.36) *Antepenultimate Stress*

	/σʰσσσσ/	h→Hd	Troch	Ft-Bin	ER-R	LaE	Parse	*Lapse
a.	σʰσσσσ	*! W				*** W	**** W	***** W
b. ☞	(ˌσʰσ)(ˈσσ)σ						*	*
c.	σʰ(ˌσσ)(ˈσσ)	*! W					*	L
d.	(ˌσʰσ)σ(ˈσσ)					*! W	*	*
e.	(ˌσʰσ)σσσ					*!* W	*** W	*** W

There are no monomorphemic words of more than five syllables attested in my data, and few multi-morphemic words of six or more syllables. Because there is no way to verify from the existing data how the language would deal with the longer monomorphemic forms (particularly given the limit of one secondary stress per word), they will not be discussed here.

3.6.6. Multiple/No *h*-Specified Syllables

So far this chapter has only discussed words with exactly one syllable bearing the *h* diacritic. How then does the grammar handle words with none, or two, or more? At least in words of up to five syllables, this analysis predicts that default stress with no diacritics to guide footing will always be penultimate. Monosyllabic words are the exception, as they have no penultimate syllable, and will bear their stress in the only available location. Otherwise, once iambs and degenerate feet have been ruled out by the appropriate constraints, the winning candidate will be that with the fewest number of lapses. Disyllabic words are footed with a single trochaic foot covering the entire word, as [(ˈσσ)]. Trisyllables avoid lapse by leaving their initial syllable, rather than the final one, unfooted, as in [σ(ˈσσ)]. Like disyllables, four-syllable words can be exhaustively trochically footed, with one secondary and one primary stress: [(ˌσσ)(ˈσσ)]. In a five-syllable word, it is once again the initial syllable which goes unfooted, as demonstrated in Tableau (3.37) below. This is the desired result, as penultimately-stressed words make up the majority of the Wamesa lexicon; an analysis which placed default stress elsewhere in the word would be suspect.

(3.37) *Default Stress*

/σσσσσ/			h→Hd	Troch	Ft-Bn	ER-R	LaE	Parse	*Lapse
a.		σσσσσ					*!** W	***** W	**** W
b.	☞	σ(ˌσσ)(ˈσσ)						*	
c.		σ(ˈσσ)(ˌσσ)				*! W		*	
d.		(ˌσσ)(ˈσσ)σ						*	* W
e.		(ˌσσ)σ(ˈσσ)					*! W	*	* W

There is some evidence that words with more than one syllable bearing the *h* diacritic exist in Wamesa. There are two acceptable pronunciations of the word *sivererei* 'star'. Both have primary stress on the penultimate [e],[12] but secondary stress may fall on either the first or second syllable. In the former case, we must posit that both the first and last syllables are marked with the *h* diacritic, as in tableau (3.38). Both diacritics are necessary;

12. Actually there is more variation than just stress placement: depending on speech rate, the final underlying /ei/ sequence may be pronounced as two syllabified vowels [e.i] or more commonly with the final /i/ vowel reduced to a glide, as the single syllable [ej]. This is a phonetic process, applying after stress assignment.

if only the first were present we would expect antepenultimate primary stress, as in (3.36), rather than penultimate. The latter possible realization, with alternating stress, suggests a historical change-in-progress of regularization providing an alternate lexical entry with no *h*-marking on the initial syllable.

(3.38) *Multiple* h-*Marking*

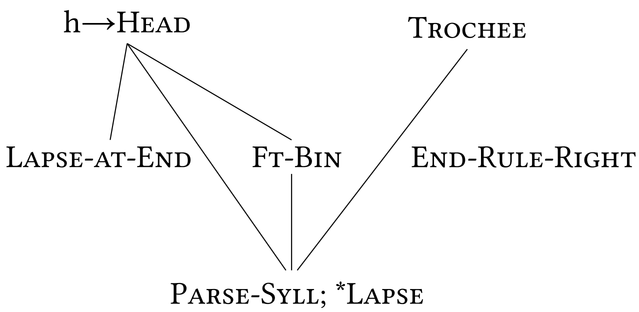
/si^h.βe.re.re^h.i/	h⟶Hd	Troch	Ft-Bi	ER-R	LaE	Parse	*Lapse
a. si^h.βe.re.re^h.i	*!* W				*** W	***** W	**** W
b. ☞ (ˌsi^h.βe)re(ˈre^h.i)					*	*	*
c. ☞ (ˌsi^h.βe)(re.ˈre^h)i		*			*	*	L
d. si^h(ˌβe.re)(ˈre^h.i)	*! W				L	*	L
e. (ˌsi^h.βe)(ˈre.re^h)i	*! W				L	*	*

Two things must be said about Tableau (3.38). First, it gives us our first ranking argument for Lapse-at-End, which must be dominated by h⟶Head in order to block candidate (d), the regularized variant, and (e), an unattested candidate, from winning out. Second, given our current knowledge of the relative rankings of Trochee and Ft-Bin, there is no way to differentiate between the two phonetically-identical candidates (b) and (c). Either, however, produces the attested stress pattern.

3.6.7. Intermediate Summary and Constraint Ranking

The analysis thus far accounts for the regular stress pattern of Wamesa monomorphemic words. Lapse-at-End can be added to the previous constraint ranking laid out in §3.6.4. An updated Hasse diagram of the constraints discussed so far is given below in (3.39). End-Rule-R remains unrankable with regards to the other active constraints discussed here.

(3.39) *Constraint Rankings*



3.7. Stress Shift With Clitics

As noted above, a small proportion of Wamesa stems have antepenultimate stress when pronounced in isolation. When these words are followed by an enclitic, primary stress shifts to the stem-final syllable, as shown in (3.40a). Words with penultimate stress, such as *maneta* in (3.40b), do not undergo this shift. The examples used here will make use of the definite determiner clitics *ne* (proximal), *=pa* (middle distance/default), and *=wa* (distal), but shift is also attested preceding *=va* 'neg.', the focus particle *=ma*, and *=ya* 'again'. The pattern is also not limited to the nouns used here; the adjective *pímasa* 'big' and the adverb *nanária* 'slowly' both appear in my data with shift in the relevant environments. There are so far nine 3- and 4-syllable shifting roots identified in my corpus and three 5-syllable non-shifting roots. Roots of three or more syllables with antepenultimate stress, like *manéta* 'friend' below, are very common.

(3.40) a. [(ˈsi.ni)tu] 'person' ⟶ [(ˌsi.ni)(ˈtu=pa)] 'the person'

b. [ma(ˈne.ta)] 'friend' ⟶ [ma(ˈne.ta)=pa] 'the friend'

This pattern is interesting in part because, at least on the level of the syllable, it does not appear to be local - the stress appears not one but two syllables to the right of its default location in forms undergoing shift. It is only on the foot level that we see the locality of the process: rather than moving to an adjacent syllable, it moves to the head of the adjacent foot. I propose that shift is triggered by a combination of lapse avoidence and the effects of END-RULE-RIGHT.

3.7.1. Domain of Lapse

This shift occurs only between a lexical word and a clitic; lapses that cross word boundaries do not trigger shift. When two shift-eligible forms occur in a row followed by a clitic, it is only the rightmost that undergoes shift; as a shifted primary stress is replaced in its original location by a secondary stress, no lapse is created and the effect is not transitive. An example of this is in the phrase *rírio pìmasá=pa-i* 'the big leech', in which the adjective *pímasa* 'big' has undergone stress shift before the determiner, while preceding *rírio* 'leech' retains its antepenultimate stress. Also at the Pword level, compound words, too, show differences in stress from their components spoken in isolation, though those processes include clash avoidance and other lexically idiosyncratic changes as well as the lapse avoidance whose effects are described here.

Based on the sorts of processes which do and don't occur in various environments, we can posit a level of attachment for the clitics. Within the Pword, lexical processes such as cluster reduction, v/r/k splitting, and vowel reduction in onsets take place. These processes are seen at the affix-stem boundary but not between a clitic and its host.

(3.41) a. *sur-* '3DU' + *pota* 'sick, hurt' ⟶ [su**p**ota] 'they two are sick'

b. *anibar* 'bee' + *=pa-i* 'DET-SG' ⟶ [aniba**rp**ai] 'the bee'

(3.42) a. *sur-* '3DU' + *vavu* 'go home' ⟶ [su**mb**avu] 'they two go home'

b. *anibar* 'bee' + *=va* 'NEG' ⟶ [aniba**rβ**a] 'not a bee'

This suggests that clitics attach higher in the structure (or later in the derivation) than compounding or affixation. There is, however, a closer relationship between a clitic and its host than between two adjacent words in the same phrase, as pointed out above. Though lapse does not induce stress shift between adjacent lexical words, clash does.

(3.43) a. Lapse-induced shift with clitics: /marárea=pa-i/ ⟶ [maràreá=pa-i]

b. No shift despite lapse between lexical words: /marárea katú/ ⟶ [marárea katú]

c. Clash-induced shift between lexical words: /marárea katú yána/ ⟶ [marárea kátu yána]

If a compound consists of two stems in a single Pword, and adjacent lexical words combine into a Phonological Phrase, then the clitic must attach between these two levels. The inability of clitics to bear stress, no matter how many syllables they add to the word, as well as the failure of word-internal phonological processes to apply, suggests that they attach at a minimal PPhrase rather than at a maximal PWd . Either interpretation requires recursivity in the prosdic structure, violating the Strict Layer Hypothesis (Selkirk 1981, 1984; Nespor & Vogel 1986). The proposed structure, along with the relevant domains, is given in Figure 3.6 for the phrase *vedianggariria pa muandu* 'having the properties of two crocodiles; the two things which have the properties of crocodiles'. The foot and syllable levels are here omitted.

3.7.2. Accounting for Shift: The Basic Pattern

This basic pattern is straightforwardly accounted for by our current constraint set and ranking. The tableaux in (3.44) show how shift occurs in three- and four-syllable words with antepenultimate stress, while (3.45) shows how the high-ranked h⟶HEAD constraint blocks shift from occurring on forms with antepenultimate stress.

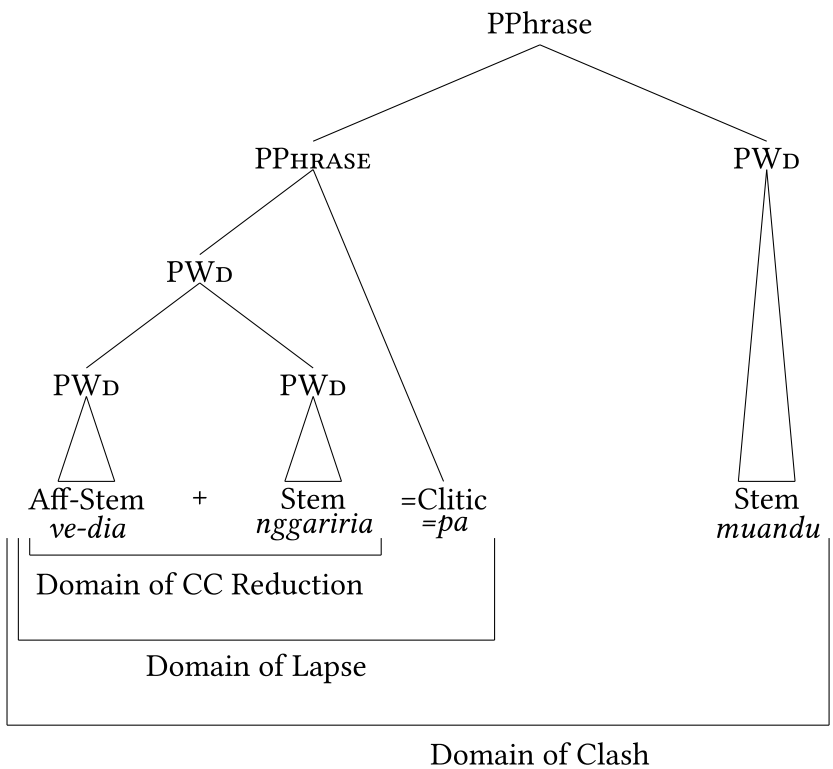


Figure 3.6: Prosodic Structure

(3.44) a. *Stress Shift: Three Syllables*

/si^{h}.ni.tu=pa/	h→Hd	Troch	Ft-Bn	ER-R	LaE	Parse	*Lapse
a. si^{h}.ni.tu.pa	*!* W				** W	**** W	*** W
b. ☞ (ˌsi^{h}.ni)(ˈtu.pa)							
c. (ˈsi^{h}.ni)(ˌtu.pa)				*! W			
d. (ˈsi^{h}.ni)tu.pa					*! W	** W	** W

b. *Stress Shift: Four Syllables*

/ma.raʰ.ri.a=pa/		h→Hᴅ	Tʀᴏᴄʜ	Fᴛ-Bɴ	ER-R	LᴀE	Pʀs	*Lᴘs
a.	ma.raʰ.ri.a.pa	*!* W				*** W	***** W	**** W
b. ☞	ma(ˌraʰ.ri)(ˈa.pa)						*	
c.	ma(ˈraʰ.ri)(ˌa.pa)				*! W		*	
d.	ma(ˈraʰ.ri)a.pa					*! W	*** W	** W

(3.45) *Penultimate Stress: No Shift*

/ma.neʰ.ta=pa/		h→Hᴅ	Tʀᴏᴄʜ	Fᴛ-Bɴ	ER-R	LᴀE	Pᴀʀsᴇ	*Lᴀᴘsᴇ
a.	ma.neʰ.ta.pa	*!* W				** W	**** W	*** W
b. ☞	ma(ˈneʰ.ta).pa						**	*
c.	(ˌma.neʰ)(ˈta.pa)	*! W					L	L
d.	(ma.ˌneʰ)(ˈta.pa)		*! W				L	L
e.	ma(ˌneʰ)(ˈta.pa)			*! W			* L	L

Wamesa avoids lapses of three or more syllables within the appropriate domain by shifting stress to the final syllable from the antepenult. (Though if we assume a non-stratal approach where the grammar gets the stem+clitic complex fully formed, this only constitutes 'shift' in comparison to the clitic-free form of the root, not in the sense of actually moving any underlying stress.) The h→Hᴇᴀᴅ, Tʀᴏᴄʜᴇᴇ, and Fᴛ-Bɪɴ constraints bar shift from occurring on normally penultimately-stressed forms. There are three possible ways for the grammar to place stress on the final syllable: candidate (c), with two binary trochees, violates h→Hᴇᴀᴅ; candidate (d), with secondary stress on the second syllable to satisfy h→Hᴇᴀᴅ nonetheless violates Tʀᴏᴄʜᴇᴇ; and candidate (e) satisfies both of these but in doing so utilizes a degenerate foot, violating Ft-Bin. Note that no Clash constraint is necessary to block candidates such as (d) and (e); the clash candidates are effectively ruled out by the other constraints.

3.7.3. Longer Enclitics

Stress never shifts rightwards beyond the edge of the stem and onto the enclitics, regardless of how many adjacent syllables this leaves unstressed at the end of the word. In the previous examples there was no Lapse-at-End-violating three-syllable word-final lapse to prompt shift in penultimately-stressed forms like *manéta* 'friend', even if final stressing of that word weren't blocked by h⟶Head, Trochee, and Ft-Bin; in these cases there is, but shift is still impossible. I assume that the clitics are marked lexically for stresslessness - this is, after all, part of what qualifies them as clitics - and that inability to bear stress will be instantiated here as an undominated *StressClitic constraint. The clitics, plus their agreement affixes, may add up to three additional syllables, but still stress remains on the stem in all cases, and fails to shift at all in roots with penultimate or final stress, as in (3.46b) and (3.46c). As can be seen from these examples, there is no prohibition against parsing the clitics into feet; they are only barred from overlapping with the head syllable of the foot.

(3.46) a. [(ˈsi.ni)tu] 'person' ⟶ [(ˌsi.ni)(ˈtu=pa)-ta.ta] 'we people'

b. [ma(ˈne.ta)] 'friend' ⟶ [ma(ˈne.ta)=pa-si.a] 'the friends'

c. [a.wa(ˈdu)] 'thigh' ⟶ [ˌa.wa(ˈdu=pa)-si] 'the/his thighs'

To account for these forms, we need to introduce a constraint barring stress from falling on the clitics, reflecting their structural position outside of the domain of stress with no dominating PWd. This constraint is unviolated in Wamesa, except for occasional cases of higher-level intonational prominence such as contrastive or emphatic stress, usually on =*va*, the negator.

(3.47) *StressClitic: Assigns one violation to every stress, primary or secondary, which falls on an enclitic.

This may well be a structural restriction rather than a true prohibition stressing clitics. As proposed in §3.7.1, clitics attach at the PPhrase level, rather than the PWd. If only the PWd and not the PPhrase can assign lexical stress, then the effect is the same as that of an undominated *StressClitic constraint. With that in mind, I will continue to use *StressClitic here in the tableaux that follow.

With this constraint in place, we can now account for the failure of stress to shift beyond the edge of the stem, even in cases of extensive word-final lapse, as we saw in the examples in (3.46).

(3.48) *Limiting Stress Shift*

a.

/si^h.ni.tu=pa-ta.ta/	*StCl	h→Hd	Troch	Ft-Bn	ER-R	LAE	Parse	*Lapse
a. si^h.ni.tu.pa.ta.ta		*! W				**** W	****** W	***** W
b. ☞ (ˌsi^h.ni)(ˈtu.pa)ta.ta						*	**	**
c. (ˌsi^h.ni)(ˌtu.pa)(ˈta.ta)	*! W					L	L	L

b.

/ma.neʰ.ta=pa-ta.ta/		*StCl	h⟶Hd	Troch	Ft-Bn	ER-R	LAE	Parse	*Lapse
a.	ma.neʰ.ta.pa.ta.ta		*! W				**** W	****** W	***** W
b. ☞	ma(ˈneʰ.ta)pa.ta.ta						**	****	***
c.	ma(ˌneʰ.ta)(ˈpa.ta)ta	*! W					L	** L	*L
d.	(ma.ˌneʰ)(ˌta.pa)(ˈta.ta)	*! W		* W			L	** L	L
e.	(ˌma.neʰ)(ˈta.pa)ta.ta		*! W				* L	**L	**L

Again the combination of h⟶Head, Trochee, and Ft-Bin prevents the stress from shifting a single syllable rightward in penultimately-stressed forms such as *manéta* while the ban on stressing clitics prevents it from surfacing any farther right than that; taken together, these constraints explain why penultimately-stressed forms remain penultimately-stressed even in the face of such extensive clitic-based lapse. These tableaux also give us the first ranking argument for *StressClitic: in order to block the unattested candidate (c) in both tableaux in (3.48), it must dominate Lapse-at-End, Parse, and Lapse.

3.7.4. Five-Syllable Words

One final case bears investigation here, and accounting for it will require two additional constraints. When an enclitic is added to a five-syllable antepenultimately-stressed word, shift fails to occur. Three examples occur in my data, one is given in (3.49).

(3.49) [(ˌa.pa)(ˈra.pi)ri] ‘gnat’ ⟶ [(ˌa.pa)(ˈra.pi)ri=wa-i] ‘the gnat’,
*[(ˌa.pa)(ˌra.pi)(ˈri=wa)-i]

Why stress would fail to shift is at first puzzling. In the unattested form, the headedness of the syllables remains the same, suggesting that h⟶Head remains satisfied, and the alternating stresses suggest a series of binary trochaic feet. *Lapse and Lapse-at-End are fully satisfied, unlike in the attested form. Why then does *àparàpirí=wa* fail to surface as such? A look at the broader lexicon provides an answer: in Wamesa, monomorphemic words never have more than one secondary stress. Affixed or compound words may have additional stresses, but clitics do not license this expansion. A form like *àparàpirí=wa*, while otherwise more harmonic than its unshifted counterpart, would break this pattern. It is therefore ruled out as an option.

This limitation is best viewed not as a cap on the number of secondary stresses in a word but rather as a cap on the number of feet. Hyman (2006: 231) describes primary stress as being defined by two properties: under Obligatoriness, each word has *at least* one syllable bearing primary stress; under Culminativity, each word has *at most* one syllable bearing primary stress. McCarthy (2003) considers these to be axiomatic properties of Gen rather than violable constraints, and Wamesa provides no counterevidence to that claim. Words with only one foot, therefore, will have primary stress; a second foot will bear a single secondary stress. If no third foot is possible, no second secondary stress can appear. Binary-branching structure is common to many aspects of the grammar; many theories of syntax require binary branching structures, and constraints like Ft-Bin enforce them in metrical structure. Here a highly-ranked constraint MaxBin(Wd) (Itô & Mester 2007: 9) will effectively limit the number of feet in the word to two.[13] This formulation allows for recursive structures, where a PWd is composed of two other PWds or a PWd and a foot, and for level-skipping structures, where a foot and an unfooted syllable may constitute a PWd, so long as there are no more than two components involved.

(3.50) MaxBin(Wd): Assign one violation mark for every PWd which is more than binary branching.

13. A stricter Wd-Bin constraint, also proposed by them as part of the Binarity family of constraints, requires that the PWd contain exactly two constituents This would also produce the attested output, if ranked below Ft-Bin.

Maximal binarity of PWds is borne out by the Wamesa lexicon, and it will be crucial to our analysis here in blocking otherwise ideal unattested stress-shifted candidates with two secondary stresses from surfacing.

It will not, however, rule out a candidate with shift and only one secondary stress, represented by candidates (d) and (e) in tableau (3.52) below. Lapse-at-End by itself is insufficient; as attested *àparápiri=wa* has the same number of non-final lapses as an unattested candidate like *aparàpirí=wa*, and more total lapses by one. In order to block the shifted candidate from winning out, we need one additional constraint. The one which will serve our purposes here requires the left edge of each foot to align with the left edge of the Pword, and is defined below.

(3.51) Align(PWd, L, Ft, L): Assigns one violation for each foot whose left edge does not coincide with the left edge of the Pword.

Note that, after McCarthy (2003), this constraint is evaluated categorically: each misaligned foot incurs only a single violation, regardless of how far rightwards of the Pword edge it falls. Feet are not exclusively left-aligned in Wamesa, as seen in a number of winning candidates above; AlignFtL must be dominated by h⟶Head and Trochee in order not to disrupt the ability of the *h* diacritic to locate the left edges of feet, but must dominate *Lapse to prevent the shifted candidate (d) from beating out the attested non-shifted *àparápiri=wa* in tableau (3.52) below. In combination with the constraints already discussed, this gives us the desired output.

(3.52) *Five-Syllable Non-shift*

/a.pa.ra^h.pi.ri=wa/		MaxBin(Wd)	*StCl	h→Hd	Troch	Ft-Bn	ER-R	LAE	All	Parse	*Lapse
a.	a.pa.ra^h.pi.ri.wa			*! W				**** W		****** W	***** W
b. ☞	(ˌa.pa)(ˈra^h.pi)ri.wa							*	*	**	**
c.	(ˌa.pa)(ˌra^h.pi)(ˈri.wa)	*! W						L	** W	L	L
d.	a.pa(ˌra^h.pi)(ˈri.wa)							*	**! W	**	* L
e.	(ˌa.pa)ra^h.pi(ˈri.wa)			*! W				** W	*	**	**

Based on the preceding tableau, it would seem that MaxBin(Wd) is redundant - ranking AlignFtL above Lapse-at-End would successfully rule out candidate (c), the only case here in which MaxBin(Wd) has an effect. Other forms, however, show that omitting MaxBin(Wd) leads to a ranking paradox. This configuration - no MaxBin(Wd), AlignFtL over Lapse-at-End - is shown in the tableaux in (3.53) for the forms *àparápiri=wa* 'the gnat', using candidates (b) and (c) from tableau (3.52), and *maràreá=pa* 'the child'. In (3.53a), AlignFtL must dominate all of Lapse-at-End, Parse, and *Lapse; in (3.53b) it must be dominated by at least one of them. The unviolated constraints *StressClitic, h⟶head, Trochee, Ft-Bin, and End-Rule-R are here omitted for clarity.

(3.53) *Removing Word Binarity: Ranking Paradox*

a.

/a.pa.ra^h.pi.ri=wa/		AlignFtL	LaE	Parse	*Lapse
a. ☞	(ˌa.pa)(ˈra^h.pi)ri.wa	*	*	**	**
b.	(ˌa.pa)(ˌra^h.pi)(ˈri.wa)	** W	L	L	L

b.

/ma.ra^h.re.a=pa/	AlignFtL	LaE	Parse	*Lapse
a. ma(ˌra^h.re)('a.pa)	**!		*	
b. 💣 ma('ra^h.re)a.pa	* L	* W	*** W	** W

It has already been demonstrated in (3.52) that AlignFtL must dominate *Lapse; the above tableaux show that it must be dominated by either Lapse-at-End or Parse (or both), though it is impossible to determine which. The reintroduction of MaxBin(Wd) dominating the higher-ranked of these two, and by transitivity also AlignFtL, makes this possible.

3.7.5. Summary and Rankings

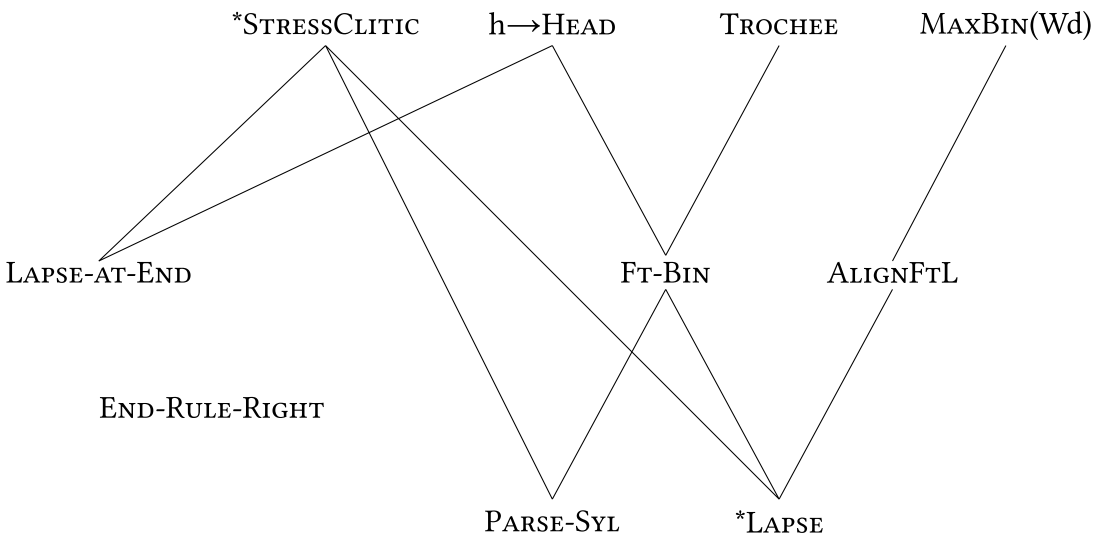
This set of constraints will successfully account for default stress placement in Wamesa, as well as for the stress shift pattern found before enclitics. For many pairs of constraints a direct ranking is not possible. Here I would like to address one pair in particular. So far there is no ranking argument available for Trochee vs. Ft-Bin. One case in which these constraints will clash is in disyllabic roots with final stress, such as *kambú* 'water'. There are two possible footings for this form: *kam('bu)*, with a degenerate foot, or *(kam.'bu)*, with an iambic foot. There is no way in Wamesa to differentiate between these footings based on surface forms, so a positive determination of the correct form is impossible; however, given that there is no evidence of iambs surfacing anywhere else in the language, I argue that, for the sake of consistency (see Prince 1990: 7), *kam('bu)* is the better analysis. For this form to surface, Trochee must outrank Ft-Bin, putting it in the top stratum of constraints. This analysis is given in (3.54).

(3.54) *Trochee » Ft-Bin*

/kam.bu^{h}/		MaxBin(Wd)	*StCl	h→Hd	Troch	Ft-Bn	ER-R	LAE	AlL	Parse	*Lapse
a.	☞ kam(ˈbu^{h})					*			*	*	
b.	(kam.ˈbu^{h})				*! W	L			L	L	

This candidate pair give us one more ranking argument for Trochee as well: in addition to Ft-Bin, it must also dominate AlignFtL. The full Hasse diagram of all relative constraint rankings is laid out in (3.55).

(3.55) *Final Constraint Rankings*



3.7.6. FaithStress

Let us return briefly to the h⟶Head constraint, and its superiority in this case over a constraint requiring (positional and prominence-level) faithfulness to lexically-assigned stress rather than headedness. In the analysis presented here, the lexically-placed diacritic places a single foot, and the Lapse constraints place the others, while End-Rule-R chooses the head foot. This approach allows the main stress to appear in whatever the optimal position is, so long as the *h*-bearing syllable is footed as a head and falls an even number of syllables away. Using a stress faithfulness constraint instead narrows down the admissible locations for primary stress to only one. While h⟶HEAD (in concert with TROCHEE and FT-BIN) differentiates between a footing pattern with primary stress one syllable to the right of the diacritic (violating) and one with primary stress two syllables over (allowed), FAITHSTRESS is equally violated by both. This difference is demonstrated in (3.56), where the superscript *h* stands in for either the headedness diacritic or underlying stress, as appropriate to the constraint.

(3.56)

/σ^{h}σσσ/	h⟶HEAD	FAITHSTRESS
(ˈσ^{h}σ)σσ	✓	✓
σ^h(ˈσσ)σ	*	*
(ˌσ^{h}σ)(ˈσσ)	✓	*

This lack of flexibility leads to problems in the analysis: in order to block shift in five-syllable words and those with penultimate stress, as in (3.57a), it must dominate all four of LAPSE-AT-END, ALIGNFTL, PARSE, and *LAPSE, but in order for FAITHSTRESS to be violated in cases of shift, as in (3.57b), it must be ranked below at least one of the four. This conflict is evident from the tableaux in (3.57), where FAITHSTRESS replaces h⟶HEAD. The unviolated constraints MAXBIN(Wd), *STRESSCLITIC, TROCHEE, FT-BIN, and END-RULE-R are here omitted for clarity.

(3.57) *FAITHSTRESS: Ranking Paradox*

a.

/ma(ˈne.ta)=pa-si/	FAITHSTR	LAE	ALL	PARSE	*LAPSE
a. ☞ ma(ˈne.ta).pa.si		*	*	***	**
b. (ˌma.ne)(ˈta.pa)si	*! W	L	*	* L	* L

b.

/(ˈsi.ni)tu=pa-si/	FAITHSTR	LAE	ALL	PARSE	*LAPSE
a. (ˌsi.ni)(ˈtu.pa)si	*!		*	*	*
b. 💣 (ˈsi.ni)tu.pa.si	L	*!* W	L	*** W	*** W

This ranking paradox does not arise when h⟶HEAD and its associated diacritics are used in place of underlying stress with FAITHSTRESS. The former is therefore the preferable solution, despite the fact that the pathological predictions made by underlying stress in Harmonic Serialism, which constitute the the original motivation for this bit of theoretical machinery, do not apply in parallel OT.

3.8. Multi-Morphemic Words

The focus of this chapter is on stress in monomorphemic words, but multimorphemic-words bear mentioning as well, if only briefly. In compound words, the rightmost constituent retains its primary stress, while the leftmost constituent, the head, sees its primary stress downgraded to secondary. This is shown in (3.58a). Stress in compounds may move to avoid clash, as in (3.58b), or undergo ideosyncratic changes, as in (3.58c). As illustrated in Figure 3.6, each component of a compound constitutes its own Pwd, expanding the number of possible stresses above two.

(3.58) a. *día* 'fish' + *sínia* 'mother' ⟶ *dìasínia* 'crocodile'

b. *sasú* 'clothing' + *sáma* 'buttocks' ⟶ *sàsusáma* 'pants'

c. *vará* ‘hand’ + *kiái* ‘digit’ + *dír* ‘nail’ ⟶ *vàrakiàidír* ‘fingernail’

Affixed forms too may bear more than two stresses when sufficiently long. I suggest that this is because, unlike (phonological) clitics, which attach at the PPhrase level, affixes project to a PWd. A longer stem such as *rariate* ‘dirty’, when combined with a subject agreement prefix and the esive prefix *ve-*, will surface with one primary and two secondary stresses, as will other forms like it. The root is parsed by the minimal PWd and the affixes by the next highest one. A negation clitic *=va* is included show the contrast between its level of attachment. Only the PWds, and not the PPhrase, can assign lexical stress, leading to the generalization that clitics always surface unstressed. The proposed structure is given in Figure 3.7.

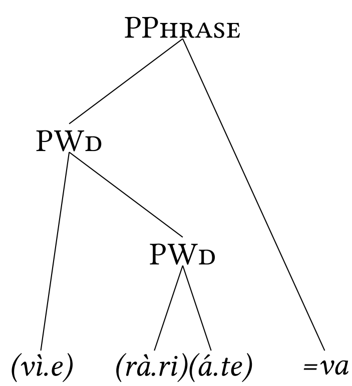


Figure 3.7: Prosodic Structure: [v<í>e-rariate=va] ‘3sg isn’t dirty’

3.9. Native-Speaker Intuitions

The forms referenced here all occurred naturally during elicitation with native speakers; none were elicited by asking speakers directly about stress shift. To test whether speakers are aware of this phenomenon, two Wamesa speakers, IMK and TLB, were each played a set of twelve audio clips. These clips were extracted from recordings of a frog story and elicited sentences as spoken by IMK and recorded with a head-mounted microphone to limit background noise. Each clip consisted of a target word which was either shift-eligible (*mararea* 'child' and *ririo* 'leech') or ineligible (*aya* 'bird' and *aparapiri* 'gnat'), and which appeared either in isolation, followed by a non-shift-inducing adjective, or by a shift-inducing determiner clitic. Half of these forms had stress in the appropriate place; the other half either underwent shift inappropriately or failed to shift when called for. The forms which do not naturally occur were constructed in one of two ways. Shift was created on forms which never naturally undergo it (*aya*, *aparapiri*) by manipulating the pitch, intensity, and particularly length in the affected vowels using the Audacity computer program. For forms which do naturally undergo shift, recordings of shifted forms were spliced into non-shift environments, and vice versa. The resulting clips were played for each of the two speakers individually, using external speakers plugged into a laptop, and in random order. The subjects were informed that I had manipulated some of the recordings but not all (they were not told how), and told for each recording to say whether it sounded good, bad, or strange ('bagus, tidak bagus, aneh').

Both speakers accepted eleven recordings and rejected one as 'bad' or 'strange', but they disagreed on which recording that was. In both cases the rejected form was a shifted instance of a shift-ineligible word. This may reflect some knowledge that the stress pattern was impossible, or could have been the result of manipulation of the audio signal creating some other problem. The speakers both unhesitatingly accepted all instances of shift occurring when it shouldn't have or failing to occur when it should have. This suggests that stress is not salient to Wamesa speakers. This test allowed them to point out if something sounded 'off' or 'not quite right' even if they were unable to pinpoint what the problem was; for the most part that did not happen. Speakers instead did not notice when stress fell on the wrong syllable. Nonetheless, they are highly consistent in their production of stress, suggesting that it is a defined part of the grammar, even if speakers do not consciously notice its placement. These results recall work by Peperkamp, Dupoux, and others (Dupoux et al. 1997, 2001; Peperkamp & Dupoux 2002) demonstrating similar 'stress deafness' in speakers of French, Finnish, and Hungarian. These results are not directly comparable to theirs - their subjects were faced with an ABX task and asked to identify identical repetitions of a word, while the speakers here simply gave grammaticality or naturalness judgements - but it bears noting that the typology discussed in Peperkamp & Dupoux (2002) and Dupoux & Peperkamp (2002) predicts that Wamesa speakers should not be stress deaf, as though stress is rarely contrastive in Wamesa, it is also not predictable. This issue bears further investigation.

Chapter 4

Dependent Morphemes: Affixes and Clitics

4.1. Clitics, Affixes, and Stems

This chapter explores the distinction between clitics and affixes and the behaviors of these two morphological categories in Wamesa. The stems to which clitics and affixes attach will be discussed in Chapter 5.

4.1.1. Wordhood

Before discussing the dependent morphemes found in Wamesa, a short discussion is in order on what constitutes a word in the language. I follow Dixon & Aikhenvald (2002) in differentiating between a form which qualifies as a word phonologically, one which qualifies grammatically, and one which fits both sets of criteria. Unlike in some highly morphologically complex languages, of which Cup'ik (Woodbury 2002) is a widely-cited example, in Wamesa these categories are highly, though not entirely, overlapping. Four classes may be distinguished in Wamesa with regards to wordhood: affixes are words in neither sense, simple clitics (in the sense of Anderson 2005) are syntactic but not phonological words, and lexical roots, with any accompanying affixes, inhabit both categories. This chapter discusses those morphemes which, in isolation, are either phonological or grammatical words or neither, but not both.

Phonological words, in Dixon & Aikhenvald's sense, may be distinguished by phonotactic, phonological, and prosodic restrictions and processes which apply only within a word and not across word boundaries. A Wamesa word may have maximally one syllable bearing primary stress: unbound lexical words have exactly one, while simple clitics, which are words grammatically but not phonologically, have none (§3). Wamesa also disallows any consonant clusters other than those made up of a homorganic nasal plus voiced stop word-internally;[1] this restriction does not apply across word boundaries or between a clitic and its host. The processes of cluster reduction and v/r/k splitting, described in §2.3.2 and §2.3.3, also occur only within a word, and not across a word boundary or between clitic and host. The word boundary further creates one of the environments under which obligatory high vowel reduction takes place (§2.3.6).

Dixon & Aikhenvald (2002: 19) define a grammatical word as a group of one or more grammatical elements which:

(a) "always occur together, rather than scattered throughout the clause (the criterion of cohesiveness);

(b) occur in a fixed order;

(c) have a conventionalised coherence and meaning."

These criteria describe well a Wamesa root and any attached affixal morphology, but (a) is violated by a stem plus clitic.[2] A root and its affixes always appear together, regardless

1. Though see §2.3.6 for discussion of surface consonant/glide clusters.

2. See §4.2 for a discussion of why.

of any movement or other syntactic processes which may have been applied. They are strictly ordered with regards to one another; object agreement always follows the verb root, subject agreement always precedes it, and the applicative prefix when present always appears between the root and its subject agreement marker. When repeating an utterance word-by-word, Wamesa speakers do not pause between a stem and affix, as they do at word boundaries and often between a clitic and its host. Each of these units has a conventionalized meaning.

Dixon & Aikhenvald (2002: 24) discuss a further criterion, based on suggestions by Sweet (1875/6) and Bloomfield (1933b), that in at least some languages an item which is both a grammatical and a phonological word may in itself constitute a complete utterance. All lexical stems other than verbs in Wamesa may occur in isolation in at least some environments, suggesting that verbs require some affixation - minimally subject agreement markers or the essive prefix - to constitute a grammatical word. This criterion has some problems, however. Unlike the definite determiners, the Wamesa indefinite determiner *pe*, for example, is not a clitic, and fulfills the criteria to be considered both a phonological and grammatical word. It is hard, however, to imagine a scenario in which it could independently constitute an utterance. It may well be that this particular criterion is useful for describing content words in Wamesa but not function words like *pe*.

4.1.2. The Dependent Morphemes

Wamesa affixal morphology includes the verbal subject agreement markers, essive *ve-*, applicative *it-*, and causative *on-*, as well as the number suffixes *-i* and *-si*, which may be found on both verbs (to fill an argument position in case of a non-overt or raised direct object) and determiners (to mark agreement with the head noun). A full list of Wamesa affixes is given in §4.4.

Examples of clitics in Wamesa include the definite determiners, directional particles, the phrase-final negator *va*, and the plural pronouns. Other than the pronouns, whose direction of attachment varies, these are all enclitics and must follow their host. Examples in the following discussion will focus primarily on the determiners for evidence, but the same patterns hold of the other types. A full list of the Wamesa clitics is given in §4.3.

§4.2 below will describe the differences between clitics and affixes and tests to distinguish between the two categories, as applied to Wamesa data. §4.4 will describe the meaning and behavior of the Wamesa affixes; §4.3 describes that of the clitics.

4.2. Distinguishing Clitics from Affixes

Both clitics and affixes share the property of being dependent on a lexical host; neither can appear independently.[3] Affixes attach to the root as part of word formation. Wamesa has only one derivational affix, the essive *ve-*, which creates adjectives; the remainder are inflectional. Changing the class of a word is most often carried out through zero-derivation.

Anderson (2005) defines two types of clitics: *phonological* or *simple clitics* and *syntactic clitics*. Unlike unbound lexical items, phonological clitics are phonologically deficient; they lack sufficient prosodic structure to constitute words on their own, though they can potentially be assigned stress as part of a larger domain (Anderson 2005: 23). In Wamesa, clitics may receive contrastive stress, but they never receive normal word-level stress, as discussed in §3. Morphosyntactic or special clitics are those "whose position with respect to the other elements of the phrase or clause follows a distinct set of principles, separate from those of the independently motivated syntax of free elements in the language" (Anderson 2005: 31). Most Wamesa clitics are phonological clitics, which, as we saw in §3.7,

3. While a distinction is drawn here between dependent (affix and clitic) and independent (lexical root) morphemes, this is somewhat reductive, as not all lexical roots may appear independently either. Wamesa verb roots, for example, require subject agreement or category-changing morphology, and never appear bare. (See §5.3.)

merge with the phonological word of the preceding lexeme. Many may be syntactic clitics as well, as they always appear as the final element of some phrase, though given a theory of syntax which defines a phrase-final structural position this distribution would not be evidence for syntactic clitichood.

Neither clitics nor affixes may appear independently, though some are homophonous with independent words; instead, they must surface on a host lexeme. The examples in (4.1) demonstrate the dependence of the plural marking suffix *-si* on its host, a behavioral pattern which may be generalized to all other affixes. In (4.1a) and (4.1b), *-si* attaches to a verb, marking a non-overt plural direct object, and to a determiner clitic, marking plural number on the direct object DP, respectively.[4] In (4.1c), however, the affix appears unattached to any host; this is ungrammatical.

(4.1) a. *Yau i-rora-**si**.*
I 1SG-hit-PL
'I hit (something plural).'

b. *Yau i-rora wona=pa-**si**.*
I 1SG-hit dog=DET-PL
'I hit the dogs.'

c. * *Yau i-rora **si**.*
I 1SG-hit PL
'I hit them.'

Despite this similarity, a distinction must still be made between affixes and clitics. Following Anderson's (2005) hypothesis that clitics combine with their host post-lexically, while affixes are incorporated earlier in the derivation, we expect to see a number of differences in the behavior of the two morpheme types which fall out from the differences in attachment time. As laid out by Zwicky & Pullum (1983), these include: a) a lower degree

4. The number marker may not appear both on the verb and on a directly following DP; see §4.4.6 for discussion.

of host selectivity in clitics than in affixes; b) a lower probability of arbitrary paradigmatic gaps for clitics; c) a lower instance of morphophonological and semantic idiosyncrasies for host + clitic groups than affixed words; d) the ability of syntactic rules to affect clitics and their hosts independently; and e) the ability of clitics, but not affixes, to attach to material already containing clitics.[5] Anderson et al. (2006) discuss specifically those morphemes which always appear at the edge of a phrase, as most Wamesa clitics do, and agree that the behavior described here points towards an analysis of these items as clitics rather than edge features. The following sections apply these tests to Wamesa morphemes, allowing us to differentiate the clitics, such as the definite determiners, from the affixes, such as the verbal agreement markers and number suffixes.

4.2.1. Test 1: Host Selectivity

According to Anderson, affixation occurs within the lexical phonology, giving affixes access to the phonological and semantic content of their stems, and thereby opening up the possibility of idiosyncratic interactions. Clitics, however, attach post-lexically, and are not therefore subject to the same interactions as the affixes.

As predicted, affixes in Wamesa are particular as to their hosts, exemplified here by the verbal agreement affixes. While these affixes appear to combine with stems from range of categories, primarily verbs and predicative adjectives, but also prepositions, locational nouns, and the adverb *kota* 'also', these stems have undergone zero-derivation to the category of 'verb', which then requires the presence of subject agreement marking. In every case, the affixed stem functions as a predicate in the clause, with the same restrictions and requirements as intrinsic verbs.[6] This zero-derivation to verbhood is not possible

5. This criterion is apparently violated by many of the examples already given, most recently (4.1b), in which a number agreement affix follows a determiner clitic. Crucially, the agreement affix is attaching specifically to the determiner clitic as head of the DP, not to the clitic's host stem. See §4.3.1 for further discussion.

6. See §5.3 for an account of these.

for every word category; nouns, for example, cannot fulfil the predicate role without the addition of the essive prefix *ve-*. The examples in (4.2) show the ability of the subject agreement affixes to signal a change of category in an adverb but not a noun.

(4.2) a. *Kodo=ne-i kota d<**i**>as.*
frog=DET-SG also 3SG-swim.
'The frog swims too.' [S29 IMK]

b. *Kodo=ne-i k<**i**>ota.*
frog=DET-SG 3SG-also.
'The frog joins in.' [S29 IMK]

c. * *Andi k<**i**>odo.*
he/she 3SG-frog.
'He is a frog; he frogs.'

Determiner clitics, on the other hand, may attach to any host which precedes them, regardless of category, without any category-shifting or other effects. The determiners, discussed in detail in §4.3.1, are phonological but not special clitics, and appear finally in the DP but for numerals and quantifiers. The examples in (4.3) show the determiners in combination with a range of hosts: a verb (4.3a), a noun and a directional clitic (4.3b), and an adjective (4.3c). See the above examples in (4.18a) and (4.1b) for examples of the most common hosts, nouns.

(4.3) a. *Kue y-a=**pa**-i=ma di-te to diadi=va.*
cake 1SG-eat=DET-SG=TOP 3SG-good until may=NEG
'The cake I ate was really delicious.' [S41 IMK]

b. *I-panande sinitu=**pa**-i papopa i-se=ra=**wa**-i.*
1SG-forget person=DET-SG earlier 1SG-see=to.there=DET-SG
'I forget the person I saw (him) before.' [S42 IMK]

c. *mararea katu=**pa**-i*
child small=DET-SG
'the little child'

Flexibility as to host word does not preclude selectivity as to the types of phrases in which a given clitic may appear; a clitic may attach to a host word of any category that linearly precedes it (or follows it, as appropriate), but many only surface in a specific category of phrase. Therefore homophonous clitics are often distinguishable by what type of phrase they appear in. The determiners, for example, must be associated with an NP. There is, however, a clitic *=pa* which marks aspect and appears VP-finally and is homophonous with the definite determiner *=pa*. Unlike the determiner, the tense-marking *=pa* never bears number agreement and almost always directly follows a verb in an independent clause. When the determiner *=pa* appears in a direct object DP, the two can appear consecutively. Example (4.4a) shows the determiner clitic *=pa* on a nominal host; (4.4b) shows the aspectual clitic *=pa*[7] on a VP host; (4.4c) shows both together.

(4.4) a. *Aya pimasa=**pa**-i*
bird big=DET-SG
'the big bird'

b. *Rebuki=pa-i t<i>au=wa=**pa**.*
stone=DET-SG 3SG-fall=down=pa
'The stone hasn't fallen yet' [S46 IMK]

c. *Yau i-serei aya=**pa**-i=**pa**.*
I 1SG-see bird=DET-SG-PA
'I haven't seen the bird yet.'

The presumed phrase structure for VP in the sentence in (4.4c) is [$_{\text{VP}}$ *i-serei* [$_{\text{DP}}$ *aya=pa-i*] *=pa*]. The determiner here appears in its syntactically determined position after the DO noun; the aspectual clitic, as required by the language, ends the VP.

4.2.2. Test 2: Paradigmatic Gaps

Anderson (2005) predicts that because clitics combine with their host post-lexically, they

7. The uses of aspectual *=pa* are covered in §4.3.5.

will have no access to semantic identity or other information about their hosts, while affixes, incorporated earlier in the derivation, do; therefore affixes are predicted to have more paradigmatic gaps and semantically idiosyncratic interaction with their host lexeme. This is borne out in the Wamesa data. The clitics show no paradigmatic gaps - they will attach to any word which the syntax allows to linearly precede them - and no unpredictable phonological shapes, suppletive forms, or semantic idiosyncrasies appear. This is not true of the inflectional affixes. The verbal agreement markers, for example, fail to surface on several of the sensory verbs[8] in the second- and third-person plural, shown in (4.5) and (4.6). On most Wamesa verbs, all person/number combinations receive overt agreement morphology, laid out later on in §4.4.1.

(4.5) *Subject Agreement on Sensory Verbs*

	pronoun	prefix	*sanevesie* 'to like'	*sanekariria* 'to be sad'
1sg	*yau*	/i-/	*i-sanevesie*	*i-sanekariria*
2sg	*au*	/bu-/	*∅-sanevesie*	*∅-sanekariria*
3sg	*andi*	/di-/	*∅-sanevesie*	*∅-sanekariria*
1pl incl	*tata*	/tat-/	*ta-sanevesie*	*ta-sanekariria*
...				

(4.6) a. *Yau i-sane-vesie y-unu karu-mas kopi.*
I 1SG-stomach-well 1SG-drink water-hot coffee
'I like to drink coffee.' [S41 IMK]

b. *Andi kota sane-vesie di-unu karu-mas kopi.*
he/she also stomach-well 3SG-drink water-hot coffee
'He likes to drink coffee too.' [S41 IMK]

8. As in many languages of the region, many sensory verbs in Wamesa are formed by compounding the word *sane* 'stomach' with a modifier. *Sanevesie* 'happy, like' translates literally as 'good stomach'; *sanekariria* 'sad' means 'bad/evil stomach'.

The applicative marker *it-* likewise undergoes complex semantic interactions with the stem to which it is attached, taking on a variety of unpredictable, if related, aspectual readings, shown in (4.7).[9] The first sentence in this set lacks the applicative marker and gives a baseline with which to compare the others. The morpheme surfaces in these forms as [i] rather than [it-] due to deletion of the /t/ before a root-initial C (cluster reduction).

(4.7) a. *Yau i-nai na Bintuni.*
I 1SG-be.at LOC Bintuni
'I'm in Bintuni.' (unmarked tense/aspect) [S68 IMK]

b. *Yau y-**i**-nai na Bintuni.*
I 1SG-APPL-be.at LOC Bintuni
'I live in Bintuni.' (durative interpretation) [S68 IMK]

c. *Yau y-**i**-nda so Bintuni.*
I 1SG-APPL-go to Bintuni
'I'm about to go to Bintuni.' (prospective interpretation) [S68 IMK]

d. *Yau y-**i**-mboru.*
I 1SG-APPL-die
'I already died.' (completive interpretation) [S68 IMK]

4.2.3. Test 3: Movement

While syntactic rules such as movement and deletion affect a lexical base and its affixes together as a unit, clitics merge with their hosts after the application of syntactic processes to an utterance (Anderson 2005: 34). The clitic and its host do not necessarily form an exhaustive constituent by themselves and are not, at that point in the derivation, a unitary word, meaning that syntactic rules apply to each independently.

The argument for movement of the NP independent of the DP is rather more theory-internal than those made for the preceding tests. Take for example the sentences in (4.8):

9. The applicative is more often used to add an instrumental argument to the verb; this is discussed in §4.4.2.

(4.8) a. *Yau i-rina Sutri di-o r<i>ora* ***sinitu=pa-sia****.*
I 1SG-know Sutri 3SG-want 3SG-hit person=DET-3PL.HUM
'I know Sutri wants to hit the people.' [S41 IMK]

b. *Yau i-rina* ***sinitu*** *Sutri di-o r<i>ora=****pa-sia****.*
I 1SG-know person Sutri 3SG-want 3SG-hit=DET-3PL.HUM
'I know the people Sutri wants to hit.' [S41 IMK]

There are (at least) two possible analyses of (4.8b). In the first, as put forward by, for example, Carnie (2013: 372-73), *sinitu* 'person' is generated in its surface position as the object of *irina* '1sg knows'. This analysis base-generates the relative clause '(who) Sutri wants to hit' in the same position, between noun and determiner, where other relatives with an overt relativizer *ve-* appear. Examples of these are given in §4.4.4. The only movement under this analysis is of a null operator, not the head noun.

Kayne (1994) argues for a second analysis, based on that of Vergnaud (1974) and later refined by Bianchi (2000), in which *sinitu pasia* 'the people' is generated as the complement of *riora* '3sg hits', and *sinitu* subsequently raises past the verbs to its surface position in the specifier of CP. He bases his argument on binding facts from English and the behavior of relative pronouns in Romance. Under Kayne's raising analysis, (4.8b) has the structure given in (4.9). The underlying word order is identical to that which surfaces in (4.8a) without the relative clause; *sinitu* moves through successive CP specs to its surface position.

(4.9) Yau irina [$_{\text{CP}}$ sinitu$_i$ Sutri dio [$_{\text{CP}}$ t_i riora t_i =pasia]]

Examples such as that in (4.3b) are potentially problematic for this analysis, though without a full analysis of relative clauses and resumptive pronouns, beyond the scope of this work, they do not necessarily rule it out.

If we take Kayne's analysis to be true, this constitutes an example of the NP, which hosts the clitic determiner in (4.8a), to move independently of the rest of the DP, leaving behind the determiner. Were *=pa* an affix rather than a clitic, this would not be possible. Should Carnie's analysis instead be preferred, then independent syntactic movement must be set aside as evidence for the clitichood of Wamesa definite determiners until additional examples are found. Enough other evidence exists, however, that this should not be troubling.

When the entire DP is fronted for topicalization, the noun and determiner clitic remain together. Example (4.10a) gives the default order for a post-verbal direct object DP; in (4.10b), the entire DP has been fronted, keeping the noun and the determiner together.

(4.10) a. *N<i>unu* ***dia=pa-i*** *tomanau.*
3SG-grill fish-DET-SG already
'He already grilled the fish.'

b. ***Dia=pa-i*** *n<i>unu tomanau.*
fish=DET-SG 3SG-grill already
'The fish, he already grilled it; the fish was already grilled.' [S10 IMK]

This sentence pair provides evidence that the DP singular number marker *-i*, and by extension its plural counterpart *-si* are indeed affixes. Note that, unlike the clitic and its host, which are syntactically independent of one another, in all of the above cases the number-marking suffixes *-i* and *-si* remain with the determiner, regardless of what movement has occurred. Affixes belong to the same grammatical word as their stem and are treated as a single word by the syntax. The sentence in (4.10b), modified such that the determiner *=pa* was raised while the singular suffix *-i* remained in situ,would be impossible with that reading.[10]

10. The sentence *Dia=**pa** n<i>unum-**i** tomanau* has the same superficial form as (4.10b) with the *-i* singular suffix left behind after the verb, taking into account the idiosyncratic addition of the *m* to the end of the verb root *nunu* 'grill' when *-i* is added. Structurally, however, the sentence is quite different: *-i* here is not the DP number-marking suffix but a homophonous one which fills the direct object argument slot of a verb with no overt object. The entire sentence has a topic-comment form, along the lines of 'The fish, he grilled it'. The lack of number marking on the DP (*dia=pa* rather than *dia=pa-i* or *=pa-si*) gives a generic reading to the subject, making the sentence semantically odd at best.

4.2.4. Clitic-Affix Interspersion

It is this ordering of derivational processes - first add affixes,[11] then perform syntactic processes, then join clitics to hosts - which precludes the interspersing of clitics and affixes on a base: by the time the clitics combine with their hosts, all affixes are already in place, and the phonology no longer has access to the internal structure of the word. While constraints may apply to order multiple clitics in relation to each other at the edge of the word, none of those clitics may see inside the word to order themselves between the existing affixes.[12]

This prediction is clearly borne out in the case of the VP clitics. When a verb stem is host to both a suffix filling the direct object argument slot and a clitic such as *=va* 'NEG', the affix will always precede the clitic; no clitic can intervene between the verb itself and the affix. Other words can, however, precede the clitic. The sentences in (4.11a) and (4.11b) show the mandatory ordering of affix and clitic after the verb root *rora* 'hit'; (4.11c) inserts an additional time adverb before the clitic. As in English, this last example is ambiguous as to whether it is the verb 'hit' or the time of hitting, 'yesterday', which is being negated.

(4.11) a. *Yau i-rora-**i**=**va**.*
I 1SG-hit-SG=NEG
'I didn't hit (it)'.

b. * *Yau i-rora=**va-i**.*
I 1SG-hit=NEG-SG
'I didn't hit (it)'.

11. Number affixes will be added to the determiner clitics at the same point that other affixes are added to non-clitic hosts; because the determiner clitics are phonological clitics only, they obey the normal rules of syntax and, unlike special/syntactic clitics, are not introduced into the structure of the sentence post-lexically (Anderson 2005: 34).

12. Specific ordering is possible between clitics however, as in Italian *da=me=lo* (**da=lo=me*) 'give it to me' (see i.e. Wanner 1977 for a fuller description of Italian clitic ordering) or Wamesa *wona=pa-i=ma* (**wona=ma=pa-i*) 'the dog (focus)'.

c. *Yau i-rora-**i** ravinie wani=**va**.*
I 1SG-hit-SG evening that=NEG
'I didn't hit (it) yesterday'.

This final criterion, that affixes cannot attach to material already containing clitics, is somewhat complicated by the fact that clitics themselves can be affixed, as the determiners often are in Wamesa. What is crucial is that these affixes are attaching to and modifying the clitic itself as head of the DP, not its host lexeme. (See §4.4.5 for discussion of number marking on Wamesa clitics.) Wamesa definite determiners are simple or phonological clitics; what distinguishes them from unbound forms is only their phonological deficiency, not any special syntactic properties. There is no reason they shouldn't undergo the same affixation as the indefinite determiners, which are not clitics. Wamesa is not the only language in which clitics may be affixed; Fehri (1988) claims that resumptive pronouns are affixed clitics in some varieties of Arabic, for example, and Heggie & Ordóñez (2005: 3-4) point out instances in Portuguese and Caribbean Spanish of affixes attaching to clitics, though they take these to be a historical change-in-progress of reduction from clitic to affix. Claire Bowern (p.c.) suggests that some direct object markers in Bardi (Nyulnyulan; northwest Australia) may constitute an instance of a prefix *jarr-* attaching to an enclitic, though they are described slightly differently in her (2012) grammar.

That the affix is attaching to the clitic, not its lexical host, is again demonstrated by cases of syntactic movement: the affixes remain with their clitic host, not with the head noun, when one but not the other is fronted in the sentence. Note again the examples in (4.8) above, in which the singular suffix *-i* and plural *-si* always appear with the determiner, and the verbal agreement affixes always appear on the verb, regardless of their position in the sentence. As already discussed, these elements cannot stand alone apart from their hosts, nor may they attach to an inappropriate host.

4.2.5. Intra-Word Phonological Processes

One final piece of evidence also shows that the Wamesa clitics bear a different relationship to their hosts than that of an affix to its stem. The cluster reduction processes that apply within a word do not take place when a C-initial determiner clitic *=ne* or *=pa* attaches to a C-final word; instead, both segments surface unchanged, as in (4.12), reproduced from (3.41). Example (4.12a) shows an /rp/ cluster reducing over an affix-stem boundary, while in (4.12b) the same cluster surfaces intact over a stem-clitic boundary. V/r/k splitting similarly fails to apply between a stem and its clitic, while it always applies within a word. This is shown in (4.13a), where splitting of /β/ takes place between two prefixes and between a prefix and stem, but not between a stem and a following clitic. This is because cluster reduction and splitting only take place within a word, whereas clitics attach later in the derivation.[13]

(4.12) a. *sur-* '3DU' + *pota* 'sick, hurt' ⟶ [su**p**ota] 'they two are sick'

b. *anibar* 'bee' + *=pa-i* 'DET-SG' ⟶ [aniba**rp**ai] 'the bee'

(4.13) a. *sur-* '3DU' + *vavu* 'go home' ⟶ [su**mb**avou] 'they two go home'

b. *sur-* '3DU' + *ve-* 'ESS' + *rawana* 'sea' ⟶ [su**mb**erawana] 'they two are blue'

c. *anibar* 'bee' + *=va* 'NEG' ⟶ [aniba**rβ**a] 'not a bee'

13. In Cowan's (1955) data, these clusters remain intact in both scenarios, though he has a final /n/ rather than /r/ on the dual subject agreement prefixes. van Balen (1915a) shows the same, and was one of Cowan's source texts, as does Anceaux (1961), who draws on both Cowan and van Balen. Though van Balen spent an extended amount of time in Windesi, his written materials differ substantially from modern spoken Windesi, and it is impossible to say whether those differences stem from language change, data from other dialects, or error. Therefore no no firm conclusions can be drawn based on this data about the status of clitics versus affixes in the Wamesa of 100 years ago. The data in Saggers (1979) and Henning et al. (1991) from the Wandamen dialect agrees with mine.

4.3. Wamesa Clitics

Wamesa has a wide range of clitics. All but two are phonological clitics, prosodically deficient and unable to bear accent themselves. This section discusses each, in varying levels of detail. (4.14) gives a full list of the Wamesa clitics.

(4.14) *Wamesa clitics*

DP Clitics

a)	*=ne*	definite determiner, proximal	§4.3.1
b)	*=pa*	definite determiner, default	§4.3.1
c)	*=wa*	definite determiner, distal	§4.3.1
d)	*=ma*	topic particle	§4.3.3
e)	*=ya*	topic particle	§4.3.3

Verbal and VP Clitics

f)	*=ra*	movement away from speaker	§4.3.4, §5.6.4
g)	*=ma*	movement towards speaker	§4.3.4, §5.6.4
h)	*=wa*	movement down or into	§4.3.4, §5.6.4
i)	*=pa*	aspectual	§4.3.5
j)	*=ya*	'again'	§4.3.6
k)	*=wo*	intensifier	§4.3.7
l)	*=e*	intensifier	§4.3.7

CP Clitics

m)	*=va*	negation marker	§4.3.8
n)	*sa=*	negative commands	§4.3.8

Other Clitics

o)	*=re*	durative aspect	§4.3.9
p)	*=e*	'and'	§4.3.10

Pronominal Clitics

q)	*tata=*	1pl inclusive pronoun	§4.3.11
r)	*ama=*	1pl exclusive pronoun	§4.3.11
s)	*mia=*	2pl pronoun	§4.3.11
t)	*sia=*	3pl human pronoun	§4.3.11

This section will describe the distribution and behavior of each of the clitics listed above.

4.3.1. Definite Determiners

Wamesa has three definite determiners: *=ne*, *=pa*, and *=wa*. The difference between these three is the amount of distance, literal or metaphorical, which they encode between the speaker and the modified noun.[14] When presented with three pens at various locations on a table, speaker IMK designated the pen set directly in front of her as *bolpen* ***nei***, the one half way across the table as *bolpen* ***pai***, and that at the far end of the table as *bolpen* ***wai***. She then clarified that in this case the scale was compressed; *bolpen wai* could also be outside in the yard or in another city. When used to contrast the relative positions of the three items, however, the compressed scale was acceptable. This was confirmed with other objects in the room and local landmarks and geographical features, and is consistent with the usage of these forms in stories and other recorded speech by IMK and other Wamesa speakers.

This same three-way distance distinction also appears in the Wamesa spatial deictic and demonstrative systems, with the forms *nini*/*nina* 'this/here', *yani*/*yana* 'that/there (middle distance)', and *wani*/*wana* 'that/there (far distance)', for more information on which see §5.7.2. Three-way distance distinctions are fairly common cross-lingustically; of the 234 languages for which the feature is recorded in WALS, 88 show a three-way distance distinction in their demonstratives, outnumbered only by the 127 languages with a two-way contrast (Diessel 2013). (The next most common system, with a four-way

14. How the relative position of the listener figures into this was never made explicit, as when actual physical objects rather than figures in a story or on paper were referred to, I (the listener) was seated in close proximity to the speaker, so any distinction was hard to observe. It was my impression however that the determining factor was only the location of the speaker. This is supported by the use of the determiners in inalienable constructions, where for example *vara=ne-i* 'hand=DET-SG' can only mean 'my (the speaker's) hand', and never 'your (the listener's) hand.' It would be interesting to explore the use of these determiners over long-distance communication such as Skype or telephone.

distance contrast, appeared in eight languages in the sample.) Three-way distance contrasts appear across families, in languages including K'ichee' (Mayan; López Ixcoy 1997), Alyawarra (Pama-Nyungan; Yallop 1977), Ewondo (Niger-Congo; Redden 1979), Basque (isolate; Saltarelli et al. 1988), Irish (Indo-European; Bammesberger 1982), Hmong Njua (Hmong-Mien; Harriehausen 1990), and Georgian (Kartvelian; Hewitt 1995). Indonesian has only a two-way contrast in its demonstratives (*ini* 'this' and *itu* 'that'), but a three-way contrast in its spatial deictics (*sini* 'here', *situ* 'there (middle distance)', and *sana* 'there (far distance)'). A three-way distance contrast in the demonstratives is reconstructed to Proto-Oceanic (Ross 2007), and is also present in other Austronesian languages, such as Tukang Besi (Donohue 1999) and Tagalog (Schachter & Otanes 1972). Closely related Biak (van den Heuvel 2006), Dusner (Dalrymple & Mofu 2012), and Ambai (Silzer 1983) have similar three way distance contrasts in their demonstratives, determiners, or both. It is very likely reconstructable to Proto-Eastern-Malayo-Polynesian, if not higher up within the Austronesian tree.

=Ne denotes closeness and is used to mark an object as very nearby physically to the speaker, very familiar, or highly salient to the discourse; it is often the determiner used when talking about the protagonist of a story, for example. *=Ne* also marks inalienable possession in the first person singular, denoting a closeness of association rather than spatial proximity. Inalienable possession, often referred to as 'direct' possession in the Papuanist literature, is used in Wamesa for kin terms and body parts. See §4.4.7 for a discussion of inalienable possession with possessors in other than the first and third person singular, and van den Berg (2009) for a discussion of possession in a number of SHWNG languages.

(4.15) *vara=**ne**-i*
hand=DET-SG
'the hand/my hand'

a. *ai=**ne**-i*
tree=DET-SG
'the tree (that's right here)'

=Pa indicates middle distance. It is the default choice when distance is not salient. *=Pa* is also used to indicate third person singular inalienable possession, as in (4.16b).

(4.16) a. *nando=**pa**-i*
banana=DET-SG
'the banana'

b. *ai=**pa**-i*
leg=DET-SG
'his leg'

=Wa denotes great distance. It is the least commonly used of the determiners; where *=ne* and *=pa* overlap significantly in the items they can denote and are often interchangeable, particularly with regards to highly salient objects, *=wa* emphasizes the distant nature of the noun it modifies. It can also indicate third person singular inalienable possession when distance is salient.

(4.17) a. *wise=**wa**-i*
mountain=DET-SG
'the mountain (way over there)'

b. *ai=**wa**-i*
leg=DET-SG
'his leg (at a distance)'

Clitics require the presence of an adjacent host word to which they may adjoin. As such, the determiners, most often *=ne* but occasionally also *=pa* or *=wa*, may be used as third person pronouns with a null preceding N head, but only when there is some overt preceding material to which they may attach (and when they are discourse-appropriate). In examples (4.18a) (reproduced from (4.3b) above), and (4.18b), the determiner appears in this capacity as the object of a verb or preposition, and can attach to the preceding element within its VP or PP. Examples such as these frequently occur in natural speech.

(4.18) a. *I-panande sinitu=pa-i papopa i-se=ra=**wa**-i.*
1SG-forget person=DET-SG earlier 1SG-see=to.there=DET-SG
'I forget the person I saw (him) before.' [S42 IMK]

b. *R<i>a na=**ne**-i.*
3SG-go LOC=DET-SG
'He goes to it/there.' [S50 BAK]

Cases with a clitic determiner in place of a subject pronoun are marginal in the language. They never appear in natural speech in my data, but the sentence in (4.19) and others like it were judged to be grammatical when presented in elicitation. Sentences such as that in (4.20), with no other lexical item preceding the clitic, are impossible.

(4.19) *Kausapa=**wa**-sia se-panande sinitu=pa-i.*
tomorrow=DET-3PL.HUM 3PL.HUM-forget person=DET-SG
'Tomorrow they (will) forget the person.' [S81 IMK]

(4.20) **=**Wa**-sia se-panande sinitu=pa-i.*
DET-3PL.HUM 3PL.HUM-forget person=DET-SG
'They (will) forget the person.' [S81 IMK]

In most cases, the DP must be marked for number. The details of DP number marking will be discussed in more detail in §4.4.5 on the number marking suffixes *-i* and *-si*; the relevant fact here is that when these suffixes do appear, they will always attach to the determiner if one is present.

The definite determiners are phonological clitics but not morphosyntactic/special ones in the sense of Zwicky & Pullum (1983) and Anderson (2005). They appear in the same position in the DP as the indefinite determiner *pe*, which is not a clitic (see §5.7.2); that is, following nouns, adjectives, and relative clauses but preceding numerals and quantifiers, as in (**??**).

(4.21) *anggadi pimasa=pa-i toru*
coconut big=DET-SG three
'the three big coconuts'

The definite determiners are *not* EDGE features, in the sense of Lapointe (1990, 1992) and Anderson et al. (2006), though in the vast majority of DPs, which lack a numeral or quantifier, they do appear phrase-finally; rather, their placement is governed by the normal rules of syntax also applying to free elements of the language. The following section, §4.3.2, gives an account of the typolgically unusual word order of the Wamesa DP, and should make it clear that no extra-syntactic processes are at work in the placement of the determiner clitics.

4.3.2. Wamesa DP Structure

Default DP word order in Wamesa is Noun (Adjective) (Relative Clause) Determiner (Quantifier/Numeral), which I will refer to here as NADQ. Greenberg (1966), in his Universal 20,[15] claims that this word order does not exist, at least according to what Cinque (2005) calls "the most sensible interpretation" of the universal. However, Cinque notes that this word order is in fact attested, and sets out an underlying universal ordering of elements within the DP which, when combined with the two types of movement he proposes as allowable, produces the Wamesa word order while also explaining its rarity.

The most common DP word orders, according to Cinque (2005: 319-320), are Det Num Adj Noun, as in English, and Noun Adj Num Det as in Javanese, several Tai-Kadai languages, Igbo, Kugu Nganhcara, and others. Unattested word orders include but are not limited to Num Det Adj Noun, Adj Det Num Noun, and Num Adj Det Noun.[16] While not non-existent, NADQ word order certainly does seem to be cross-linguistically unusual.

15. "When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite'' (p. 87).

16. See Cinque (2005: 319-320) for a list of all logically possible DP word orders and their prevalence.

Cinque's survey of the typological literature turns up ten examples: four Tibeto-Burman languages (Lalo, Lisu, Akha, and Qiang), three Bantoid languages (Aghem, Lingala, and Babungo), one Trans-New-Guinea language (Koiari), and one clear case (Port Sandwich) and one tentative one (Woleaian) within Oceanic.[17] This word order is not universal within Cenderawasih Bay, but neither is it limited to Wamesa; in Biak and Ambai, numerals and quantifiers precede the determiner (a very common word order according to Cinque), but Dusner patterns with Wamesa in placing determiners before quantifiers.

Cinque suggests that word order within the DP is universally base-generated as Det (his Dem) Num Adj N. He gives the basic underlying syntactic tree reproduced here in Figure 4.1.

The attested word orders are arrived at by raising elements upwards within the tree, either by moving the NP alone from Specifier to Specifier through the AgrPs, placing the NP at different points between the other three elements, which retain their ordering relative to each other; by pied-piping the entire category directly dominating the Spec hosting the NP to 'roll up' the tree and reverse the order of elements; or by some combination of the two. Under this analysis, only the NP or elements containing the NP may move. This restriction rules out those word orders which are unattested, as producing them from the base-generated Det Num Adj Noun order requires movement of constituents not containing the NP.

Following Cinque's analysis, then, the Wamesa word order results from three steps. First, the NP is raised into the Spec of Agr$_y$P. Next the entire Agr$_y$P passes first through Agr$_x$P, then finally into the Spec of Agr$_w$P. The second of these steps involves pied-piping; the first and the third do not. Ordering movement without pied-piping after movement with it is marked, according to Cinque; once a language begins to pied-pipe it generally

17. See Cinque (2005) for the relevant references.

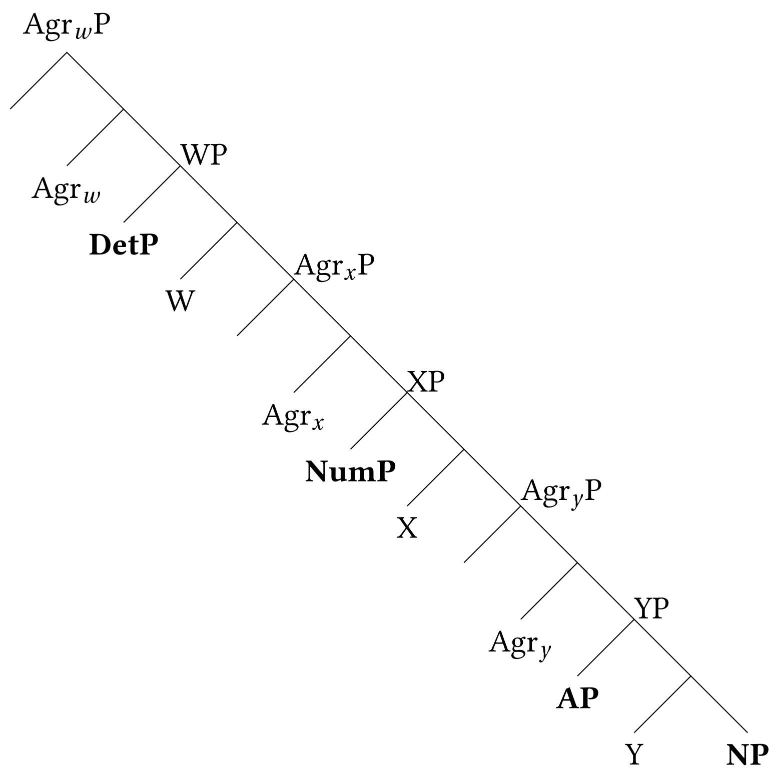


Figure 4.1: Base-Generated DP Structure (Cinque 2005: 317)

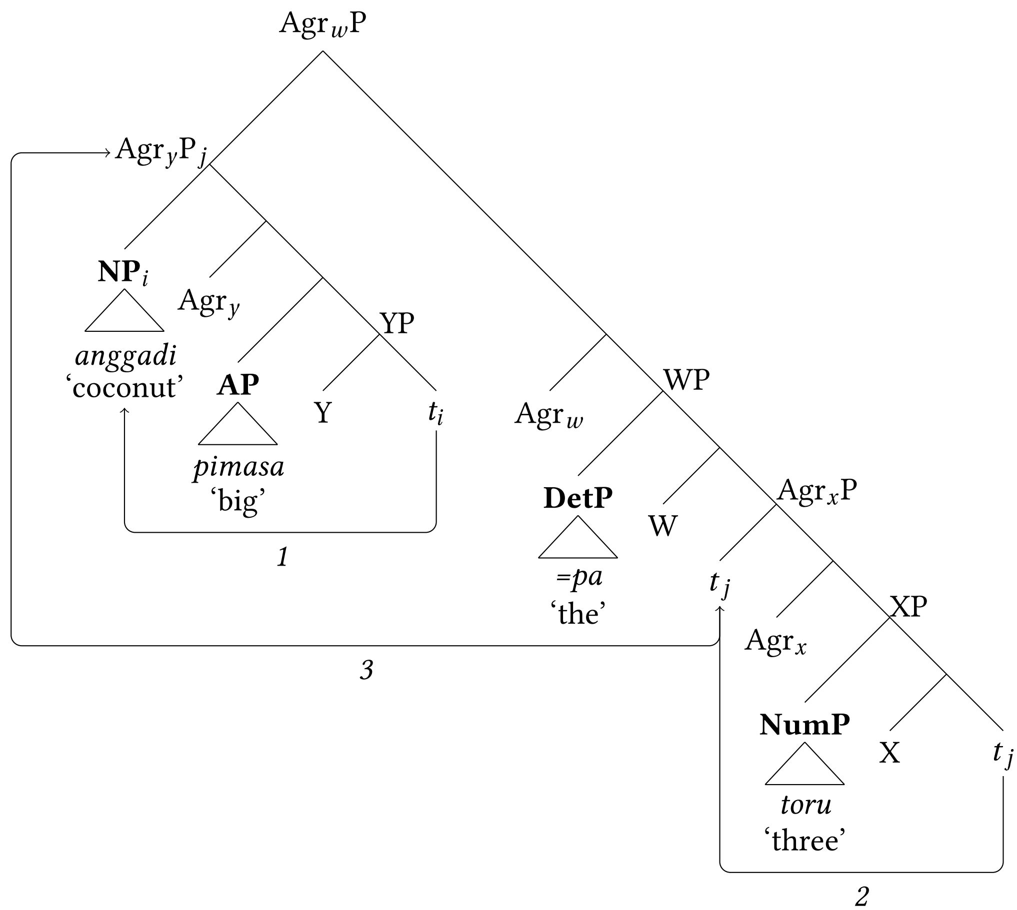


Figure 4.2: Wamesa Surface DP Structure: 'The three big coconuts'

continues to do so. This marked ordering of movement types leads to the typologically unusual status of this word order (Cinque 2005: 323). Cinque notes that other orders which, like Wamesa, include both types of movement interspersed are similarly rare cross-linguistically.

That the number-marking suffixes, numerals, and non-universal quantifiers (when used attributively) are all in complementary distribution in Wamesa suggests that all three appear in the same position, namely as head of the NumP. The universal quantifier *vura* 'all', however, shows a different behavior, co-occurring with the number-marking suffixes.[18] Cinque (2000, 2005) suggests, based on word orders such as that of phrases like 'all those four new books' in English and other Indo-European languages and the mirror image word order in some Semitic languages, that universal quantifiers are generated in a separate position, higher than D. If this is indeed the case, it would imply that in Wamesa the entire Agr$_w$P above would need to raise above the universal quantifier in one final step to produce the attested order in phrases like *wona wasi vura* 'all the dogs' (lit. 'dog=DET-PL all').

(4.22) $[[[[[[[[\mathrm{NP}]_i\ \mathrm{Adj}]\ t_i]_j\ \mathrm{Det}]\ t_j]\ \mathrm{Num}]\ t_j]_k\ \mathrm{Q_{univ}}]\ t_k]$

4.3.3. Topicalization

There are two topicalization clitics in Wamesa, *=ma* and *=ya*. These are homophonous with VP clitics meaning 'to here' and 'again', respectively (see §4.3.4 and §4.3.6), and their meanings are clearly related. They always appear in final position within the DP.

In terms of meaning, *=ma* and *=ya* are largely, but not entirely, interchangeable; the environments in which *=ya* is appropriate are a proper subset of those in which *=ma* can be used. Frascarelli (2007) and Frascarelli & Hinterhölzl (2007) claim the existence of three different types of topics: *aboutness topic*, "what the sentence is about", particularly if newly introduced or returned to (after Strawson 1964; Reinhart 1981; Givón 1983; Lambrecht 1994); *contrastive topic*, creating oppositional pairs with other topics (after Kuno 1976; Büring 1999); and *familiar topic*, a given, discourse-linked item used for topic continuity (Pesetsky 1987; Givón 1983). Wamesa *=ma* can be used to indicate any of these three types, with contextual cues used to differentiate. *=Ma* most often appears in my data marking the subject when a similar sentence is being repeated multiple times with

18. Instances of *vura* 'all' alongside a numeral do not appear in my corpus; this does not necessarily mean that they are ungrammatical. This analysis predicts that a phrase like *anggadi pimasa pa toru vura* 'all the three big coconuts' (lit. 'coconut big=DET-SGthree all') should in fact be grammatical.

slight adjustments, as with verbal paradigms or eliciting different tenses and aspects on the same sentence. Only *=ma*, not *=ya*, may appear with pronouns and demonstratives.

(4.23) *Aboutness topic:*

S<i>obatai! T<i>enam. Pibata=ne-i t<i>enam, t<i>enam di-niare!
3SG-wake.up-3SG! 3SG-live. turtle=DET-SG 3SG-live, 3SG-live 3SG-crawl!

*T<i>enam. T<i>enam di-niare. Mararea=ne-i=**ma** s<i>e=ra varami.*
3SG-live. 3SG-live 3SG-crawl. child=DET-SG-TOP 3SG-see=to.there not.yet.

Wona=ne-i tuti kodo=ne-i su-se tomanau.
dog=DET-SG with frog=DET-SG 3DU-see already

'He wakes up! He's alive. The turtle is alive, he's alive and he's crawling! He's alive. He's alive and he's crawling. The child hasn't looked over there yet. The dog and the frog already see.' [S29 IMK]

(4.24) *Contrastive topic:*

*M<i>eta. V<i>urar katu. Nini=pa-i=**ma** k<i>ke. Nini=pa-i=**ma***
3SG-black. 3SG-red small. this-DET-SG-TOP 3SG-green. this-DET-SG-TOP

*v<i>usa. Nini=pa-i=**ma** di-urar. Nini=pa-i=**ma** kumuar. M<i>eta.*
3SG-white. this-DET-SG-TOP 3SG-red. this-DET-SG-TOP dark. 3SG-black.

'It's black. It's pink. This one is green. This one is white. This one is red. This one is dark. It's black.' (after several minutes of naming colors) [S4 IMK]

(4.25) *Familiar topic:*

*Wona=ne-si vura si-ru-mi=pa-si si=pota. Wona=ne-si=**ma***
dog=DET-PL all 3PL.NH-head-POSS=DET-PL 3PL.NH-hurt. dog=DET-PL=TOP

si-ru-mi=pa-si si=pota.
3PL.NH-head-POSS=DET-PL 3PL.NH-hurt.

All the dogs' heads hurt. The dogs' heads hurt.' [S8 IMK]

It is ungrammatical for two DPs in the sentence to both be marked with *=ma*, though it may co-occur with the homophonous VP clitic meaning 'to here', as in example §4.27. Those instances of double contrastive topics in my corpus, where both the subject and object are contrastive, are divided into two sentences, with a clear intonational break between them. The subject of each is marked with =*ma*, while the predicate noun is unmarked, as shown in (4.28).

(4.26) a. *Yau=**ma** i-pote dia.*
I=TOP 1SG-go.fishing fish
'I catch fish.'

b. *Yau i-pote dia=**ma**.*
I 1SG-go.fishing fish=TOP
'It's fish that I catch.'

c. * *Yau=**ma** i-pote dia=**ma**.*
I=TOP 1SG-go.fishing fish=TOP
'I catch fish.' [S42 IMK]

(4.27) *Yau=**ma** i-kopa=**ma**.*
I=TOP 1SG-jump=to.here
'I jump over here.'

(4.28) *Nini=**ma** vavi Amerika nina. Wani=**ma** vavi Inggris.*
this=TOP woman America here. that=TOP woman England
'This is an American woman. That is an English woman.' [S29 IMK]

=*Ya* is used for familiar topics only. In Example (4.29), the turtle has already been the topic of narration for several pages of the story being described (Mayer & Mayer 1971), and continues to be for a while longer. Here the boy is holding his dog and his fishing pole, while the turtle hangs from the dog's paw, which he is biting. =*Ya* is unlikely to indicate a switch topic in this case, as the preceding verb *setapai* is marked for plural subject agreement, meaning three or more actors, which in this context must include the turtle. If it referred only to the boy and the dog, the dual form *surapai* would be used.

(4.29) *Pibata di-vakire. R<i>ute n<i>e sarera=ne-i, s<i>rio wona=ne-i.*
turtle 3SG-hang, 3SG-hold 3SG-have fishing.pole=DET-SG, 3SG-carry dog=DET-SG.

*Set-apai vera kambu. Ah, pibata=ne-i=**ya** t<i>risu ai=ne-i.*
3PL.HUM-run toward water. Ah turtle=DET-SG=TOP 3SG-let.go leg=DET-SG.

‘The turtle is hanging. He (the boy) holds the fishing pole, he carries the dog. They run to the water. Ah, the turtle let go of his (the dog’s) leg!’ [S29 IMK]

There are few instances of question/answer pairs in my data, but those which do appear include neither *=ma* nor *=ya* in the answer, though it whether they are ungrammatical in this position has yet to be determined.

(4.30) a. *Yau i-nai na toine? Au n<u>ai na aniose Manokwari.*
I 1SG-be.at LOC where? you 2SG-be.at LOC village Manokwari
‘Where am I?’ ‘You’re in Manokwari.’ [S41 IMK]

Au bu-o r<u>a topina? Yau y-o i-ra nina.
you 2SG-want 2SG-GO to.where? I 1SG-want 1SG-go here
‘Where are you going?’ ‘I’m coming here.’ [S13 IMK]

4.3.4. Directional Clitics

The directional clitics *=ra*, *=ma*, and *=wa* appear directly after the verb. *=Ra* ‘to there (away from speaker)’ is homophonous with the verb root *ra* ‘go’; *=ma* ‘to here (towards the speaker)’ is homophonous with the DP topic marker (§4.3.3); *=wa* ‘down, in’ is homophonous with the distal definite determiner (§4.3.1). The meanings of each these homophonous pairs are all related, and each likely developed from the same source. These markers are also discussed in §5.6.4.

=Ra and *=ma* are by far the most common of the three directional markers. They are attested in my corpus on a wide range of verbs of motion. The distribution of *=wa* is far more limited; the vast majority of attestations are with the verb *tau* ‘fall’. These are only grammatical with verbs of motion; use with other items was rejected by speakers, though a few naturally-produced (rather than elicited) cases do appear in my data, as with the sentence *set-i-kavio=**ra*** 3PL.HUM-APPL-speak=to.there ‘they use it to speak to over there’, used to describe a telephone. *=Ra* also commonly occurs with the verb *se* ‘see’ to indicate

the direction of looking. According to speakers, use of *=ra* and *=ma* implies, but does not entail, that the endpoint of the movement has been reached and the action is completed.

(4.31) a. *Yau y-apai.*
I 1SG-run
'I run.'

b. *Yau y-apai=**ra*** *(so Bintuni).*
I 1SG-run=to.there (to Bintuni)
'I run there (to Bintuni).'

c. *Yau y-apai=**ma*** *(so Manokwari).*
I 1SG-run=to.here (to Manokwari)
'I run here (to Manokwari).'

d. *Rebuki=pa-i t<i>au=**wa***.
stone=DET-SG 3SG-fall=down
'The stone falls down.'

4.3.5. Aspectual *=pa*

The aspectual clitic *=pa* appears finally within the VP. Its precise meaning varies depending on interactions with the other aspectual elements in the sentence. Alone, speakers translated sentences with *=pa* with either 'not yet' or 'already'. When combined with a verb bearing the applicative prefix, however, *=pa* was consistently translated as 'already', 'just now' or otherwise indicated to mean past tense. The applicative, discussed in §4.4.2, can have both an instrumental and an aspectual (usually prospective) function. *=Pa* follows any directional clitics, as in (4.32a). Both (4.32a) and (4.32b) show *=pa* without any other aspectual elements giving the interpretation 'not yet'; (4.32c) has *=pa* in isolation with the 'already' interpretation; the remaining examples show *=pa* in combination with the applicative giving the range of possible non-future TAM readngs.

(4.32) a. *Rebuki=pa-i t<i>au=wa=**pa**.*
stone=DET-SG 3SG-fall=down=PA

‘The stone hasn’t fallen down yet.’ [S46 IMK]

b. *Wona=pa-i k<i>ripe yau=**pa**.*
dog=DET-SG 3SG-bite I=PA.
‘The dog didn’t bite me yet.’ [S47 IMK]

c. *Yau i-vawou=**pa**.*
I 1SG-go.home=PA
‘I went home.’ [S68 IMK]

d. *Yau y-i-mbavou=**pa**.*
I 1SG-APPL-go.home=PA
‘I used it to go home.’ [S68 IMK]

e. *Yau y-i-mbui=**pa**.*
I 1SG-APPL-write=PA
‘I’m already just now writing.’ [S68 IMK]

4.3.6. =*Ya* ‘again’

The enclitic =*ya* appears finally within the VP, with the meaning ‘again’. This =*ya* is homophonous with the DP clitic =*ya* marking familiar topic (§4.3.3) and their meanings are clearly related; it seems likely that DP =*ya* developed by semantic shift from VP =*ya*, especially as a VP final clitic will in some cases surface directly following an object DP, which could lead to a structural re-interpretation. The examples in (4.33) show =*ya* ‘again’ in VPs with and without a direct object, as well as with an adverbial modifier.

(4.33) a. *Vavi=pa-i di-osa=**ya**.*
woman=DET-SG 3SG-stand=again
‘The woman stands up again.’ [S74 IMK]

b. *Di-unu kambu=**ya**.*
3SG-drink water=again
‘She drinks water again.’ [S74 IMK]

c. *Aya wani=ma t<i>au rawave=**ya**.*
bird that=TOP 3SG-fall down.below=again.

'That bird falls down again.'

4.3.7. The Intensifier Clitics

Wamesa has two clitics, *=e* and *=wo*, which act as intensifiers on verbs and predicative adjectives. The former, *=e*, is transparently related to the conjunction *=e*. These enclitics most often appear on predicative adjectives; when modifying true verbs they can convey a sense of surprise or of doing the action intensely.

(4.34) a. *Pibata=ne-i t<i>enam=**e**!*
turtle=DET-SG 3SG-live=INTENS
'The turtle is (really, surprisingly) alive!' [S29 IMK]

b. *Di-urar=**e**.*
3SG-red=INTENS
'It's very red.' [S35 IMK]

c. *B<i>ba=**wo**.*
3SG-big=INTENS
'It's very large.' [S35 IMK]

The second of these clitics, *=wo*, also functions as a discourse particle, whose meaning is so far still unclear. In that use it was often translated by speakers into Malay as *jadi*, as in (4.35). *Jadi* in Standard Indonesian means 'happen', 'become', and 'therefore', but in colloquial Papuan Malay it also functions as a discourse marker with a wide range of difficult-to-pin-down functions, making it difficult to translate.[19]

(4.35) a. *Mararea=wa-sia set-ane kue=pa toru=**wo**.*
child=DET-3PL.HUM 3PL.HUM-eat cake=DET three=WO
'The children eat three cakes.'
'*Anak-anak makan kue tiga jadi.*' [S77 IMK]

19. Kluge (2014) defines clause-final *jadi* in the variety of Papuan Malay spoken on the Sarmi coast to the east of Cenderawasih Bay as a conjunction marking a causal relationship with a preceding unmarked clause, where the result is expected. This covers a subset of its uses in the PM spoken in West Papua, but not, to my knowledge, all uses.

b. *Andi r<i>ama=**wo**.*
he/she 3SG-come=WO
'He comes here'
'*Dia ada datang jadi.*'

4.3.8. The Negation Clitics

The clitic *=va* is used for negation, sometimes in combination with *sa=*, discussed later in this section. *=Va* appears finally within the CP or directly after the verb, though most commonly in the former position. Exactly which element of the clause is being negated is syntactically ambiguous; in example (4.36a), it could be someone else who sees the child, the child could be heard rather than seen, I could see the dog rather than the child, or I could be looking for the child but not see him. Of all of these possible interpretations, the correct one is picked out by contextual cues, intonation, and real-world knowledge.

(4.36) a. *Yau i-sayore mararea=wa-i=**va**.*
I 1SG-see child=DET-SG=NEG
'I don't see the child.'

b. *Yau kota i-ri=**va** sasu sama ve-meta=wa-i*
I also 1SG-know=NEG clothing buttocks ESS-black=DET-SG
v<i>e-mahal.
3SG-ESS-expensive
'I don't know if the black pants are expensive.' [S29 IMK]

c. *Yau=ma i-ra=**va** so pasar.*
I=TOP 1SG-go=NEG to market
'I don't go to the market.'

Negative commands are also formed using *=va*, in conjunction with the proclitic *sa=*. *Sa=* appears clause-initially; *=va* is subject to the same positional limitations as in non-imperative negation, appearing either immediately post-verbally or clause-finally. A regular imperative is formed simply by marking the verb with second person agreement;

negative imperatives apply *sa= … =va* to the same construction.

(4.37) a. *R<u>a (so) pasar=wa-i!*
2SG-go to market=DET-SG
'Go to the market!' [S29 IMK]

b. ***Sa****=r<u>a=****va*** *so pasar=wa-i!*
NEG=2SG-go=NEG to market=DET-SG
'Don't go to the market!' [S29 IMK]

c. ***Sa****=r<u>a so sasi dire=****va!***
NEG=2SG-go to salt edge=NEG
'Don't go to the beach!' [S13 IMK]

Sa=…=va is less commonly used to negate non-imperative clauses. In this case, there must exist some expectation on the part of either speaker or hearer that the event described by the clause did place, often as something planned, or the logical consequence of an earlier statement. Use of *sa=…=va* cancels that expectation, similar to the usage of *tidak jadi* 'it didn't happen (contrary to expectation)' in Indonesian. In (4.38), the fact that the dog was chasing the pig sets up a reasonable expectation that the dog might catch the pig, but that is contradicted by *sa=tipuri=va* 'he doesn't grasp him.'

(4.38) *Wona ve-usar pimuna=pa-i* ***sa****=t<i>pur-i=va.*
Dog ESS-chase pig=DET-SG NEG=3SG-grasp-3SG=NEG
'The dog that is chasing the pig isn't catching him.' [S77 IMK]

4.3.9. Durative *=re*

The clitic *=re* is unusual in that it encodes what I here refer to as durative aspect on whatever element hosts it. It is attested in my data not only on verbs but also on prepositions and quantifiers. It indicates that the state of affairs described by the host lexeme is continuing, long-lasting, or habitual. Example (4.39a) shows *=re* on a verbal host; (4.39c) on a preposition; and (4.39d) on a quantifier. When *=re* co-occurs on verbs with the directional enclitics =*ra*, =*ma*, and =*wa*, as in (4.39b), it must precede the directional clitic.

(4.39) a. *Yau y-apai=**re** so wana.*
I 1SG-run=DUR to there
'I'm running over there.'

b. *Yau y-apai=**re**=ma.*
I 1SG-run=DUR=to.here
'I'm running here.'

c. *Mararea=ne-i di-ena tuti=**re** wona.*
child=DET-SG 3SG-sleep with=DUR dog
'The child is sleeping with his dog.' [S7 IMK]

d. *Ririo=pa pau=**re** si-karipe mararia=pa-i ai=pa-i.*
leech=DET many=DUR 3PL.NH-bite child=DET-SG leg=DET-SG
'Many leeches are biting the boy's leg.' [S75 IMK]

4.3.10. Conjunction

Enclitic =*e* 'and' is used to conjoin phrases of the same type. It appears in my corpus conjoining two or more DPs, NPs, VPs, and AdjPs. The conjunction of DPs or APs can be mono- or polysyndetonic, with =*e* appearing optionally after the final conjunct, as in (4.40c). With other types of conjuncts, =*e* does not follow the final conjunct. While my data includes examples of more than two nominals being conjoined with =*e*, as in *kodo=ne-i=e, wona=ne-i=e, mararea=ne-i=e, sarera=ne-i* 'the frog and the dog and the boy and the fishing pole', it does not include examples of greater than two conjuncts of other types.

(4.40) a. *Wona=wa-i=**e** kodo=wa=i(=**e**) sur-ase.*
dog=DET-SG=and frog=DET-SG(=and) 3DU-swim
'The dog and the frog swim.' [S77 IMK]

b. *wona ve-meta=**e** ve-vusa pau-re*
dog ESS-black=e ESS-white many-DUR
'many black-and-white dogs' [S77 IMK]

c. *Yau i-ra i-serei suomuse pimasa=**e** katu=**e**.*
I 1SG-go 1SG-see shark big=e small=e
'I go see big and small sharks.' [S34 IMK]

d. *Suo=ne-i di-urar=**e** s<i>orap.*
nose=DET-SG 3SG-red=and 3SG=sting
'My nose is red and stinging (from sunburn).' [S78 IMK]

e. *Vavi=pa-i r<i>ute stoples=pa-i=**e** s<i>or-i.*
woman=DET-SG 3SG-grasp jar=DET-SG-and 3SG-close-SG
'The woman holds the jar and she closes it.' [S74 IMK]

An equally common strategy for conjoining two DPs is to use the preposition *tuti* 'with'. This is shown in example (4.41), with a meaning equivalent to that of (4.40a) above.

(4.41) *Wona=wa-i **tuti** kodo=wa=i sur-ase.*
dog=DET-SG with frog=DET-SG 3DU-swim
'The dog and the frog swim.' [S77 IMK]

This contrasts with the non-conjunctional use of *tuti*, as in (4.42). Note that in (4.41) the verb gets dual subject agreement, while in (4.42) the agreement marker is singular.

(4.42) *Mararea=ne=i r<i>a **tuti** wona=ne-i.*
child=DET-SG 3SG-go with dog=DET-SG
'The child goes with the dog.' [S7 IMK]

Stassen (2000) defines two typological classes of languages, those with a comitative strategy which make use of an adposition meaning 'with' for NP(/DP) conjunction ("WITH-languages"), and those like English with a coordinate strategy ("AND-languages"). He notes that many AND-languages also have a comitative construction available for conjunction, and that WITH-languages tend to develop into AND-languages, but not vice versa. Stassen would classify Wamesa as an *and*-language, as it has as coordinating construction, using *=e*, clearly distinct from its comitative preposition. That it also has a prototypical comitative construction, with singular subject agreement on the verb and a post-verbal second participant, is not surprising - AND-languages often allow both, as for example in the English gloss for Example (4.42). What is interesting is that Wamesa appears to have grammaticalized a second coordinate strategy, as in the sentence in (4.42), where the PP headed by *tuti* 'with' appears preverbally and the verb gets dual subject agreement despite the unequal structural rank of the two participants. Stassen (2000: 27) mentions grammaticalization of this pattern as a way in which SVO WITH-languages may develop into AND-languages; in Wamesa it doubles the function of the coordinate =*e* construction.

The colloquial varieties of Indonesian with which I am familiar, including Papuan Malay, make use of the same three types of constructions shown in (4.40a), (4.41), and (4.42). Malay verbs do not agree with their subjects so the singular/dual contrast is not evident, but the word order is the same. The coordinating conjunction is *dan* or *dang*; *dengan*/*dengang* and *sama*[20] are comitative prepositions which can be used conjunctively as well. It is plausible that the conjunctive use of *tuti* in Wamesa is the result of contact with Malay. The sentences in (4.43) show these structures in colloquial Indonesian; they are IMK's translations of the sentences in (4.40a), (4.41), and (4.42). See Kluge (2014: 424-428, 507-9) for parallel examples in PM.

(4.43) a. *Anjing dan kodok berenang.*
dog and frog swim
'The dog and the frog swim'

b. *Anjing sama(/dengan) kodok berenang.*
dog and frog swim
'The dog and the frog swim'

c. *Anak laki-laki pergi sama(/dengan) anjing.*
child male go with dog
'The boy goes with the dog.'

20. *Sama* has a range of uses, including introducing goals, recipients, and patients; see Kluge (2014: 426-27) for details of its usage in Sarmi Coast PM, similar to the PM spoken around Windesi.

One other environment in which =*e* appears frequently is the construction of numerals. Wamesa has an additive number system, described in more detail in §5.7.1. Briefly, it has discrete words for numbers up to five and again for ten and twenty, other numbers are formed by addition using -*e* 'and', as shown in (4.44).

(4.44) a. *rime=**e*** *toru*
five=and three
'eight'

b. *sura=**e*** *rime=**e*** *at*
ten=and five=and four
'nineteen'

4.3.11. Plural Pronouns

The singular, dual, and trial pronouns in Wamesa are fully independent lexical words which bear default stress. The plural pronouns, by contrast, are phonological clitics, and never surface independently. The table in (4.45) gives the full set of Wamesa pronouns.

(4.45) *Wamesa Pronouns*

Person	**Singular**	**Dual**	**Trial**	**Plural**
1incl		*tandu*	*tatoru*	*tata=*
1excl	*yau*	*nandu*	*amatoru*	*ama=*
2nd	*au*	*mandu*	*metoru*	*mia=*
3rd	*andi* (subj/obj) *i* (obj)	*sandu*	*setoru*	*sia=*

Like many languages with rich morphological verbal agreement systems, Wamesa is a *pro*-drop language; subject pronouns are optional. The plural pronouns, when they are used, appear in my data almost exclusively hosted by one of three items: the particle *te*, which adds no extra semantic information; *vura* 'all'; or *kota* 'also'. Examples of these are given in (4.46). Only the dual pronouns can be used with groups of two; the plural pronouns are grammatical with groups of three or more. The trial pronouns are rare. For further discussion of non-plural pronouns, see §5.7.3.

(4.46) a. ***Sia***=*te* *se-sanevesie* *set-unu* *karumas* *kopi.*
they=TE 3PL.HUM-like 3PL.HUM-drink hot.water coffee
'They like to drink coffee.'

b. ***Tata***=*vura* *ta-maye.*
We.incl=all 1PL.INCL-dance
'We all dance.'

c. ***Mia***=*kota* *me-mbavou.*
We.incl=also 2PL-go.home
'You (pl) go home too.'

4.4. Affixes

Putting aside the large number of verbal subject agreement markers, the set of affixes in Wamesa is rather smaller than that of clitics. While most Wamesa clitics are enclitics, the affixes other than those for verbal subject agreement are evenly split between prefixes and suffixes. (4.47) gives a list of the verbal agreement markers, which are discussed in more detail in Chapter 6 (Verbal Infixation). The remaining affixes are listed in (4.48).

(4.47) *Wamesa Verbal Subject Agreement Affixes (§4.4.1, §4.4.7, §6.2)*

Person	Singular	Dual	Plural
1incl		*tur-*	*tat-*
1excl	*i-*	*amur-*	*amat-*
2nd	*bu-*	*mur-*	*met-*
3rd	*di-*	*sur-*	*set-* (human) *si-* (non-human)

(4.48) *Other Wamesa Affixes*

a)	*it-*	applicative	§4.4.2
b)	*on-*	causative	§4.4.3
c)	*ve-*	essive	§4.4.4
d)	*-i*	singular DP/DO	§4.4.5
e)	*-si*	plural non-human DP/DO	§4.4.5
f)	*-tata*	1pl.incl human DP	§4.4.5
g)	*-ama*	1pl.excl human DP	§4.4.5
h)	*-mia*	2pl human DP	§4.4.5
i)	*-sia*	3pl human DP	§4.4.5
j)	*-mu*	inalienable possession (2sg)	§4.4.7
k)	*-mi*	inalienable possession (pl)	§4.4.7

4.4.1. Subject-Verb Agreement

Verbs in Wamesa must agree with their subjects in person, number, and, in the case of the third person plural, animacy. This applies not only to true verbs, but also to adjectives and quantifiers when used predicatively. (See §5.4 for a clarification of these word classes.) Chapter 6 discusses the distribution and behavior of the verbal subject agreement affixes in detail; an overview is given here.

Wamesa has twelve distinct verbal agreement markers. Number distinctions are made between singular, dual, and plural forms,[21] and between inclusive and exclusive forms in the first person non-singular. In the third person plural, a distinction is made between human subjects, which call for the *set-* agreement prefix, and non-human ones, for which *si-* is used. In most cases, these agreement affixes are prefixal, but in the second- and third-person singular, reduced forms of the affixes surface as infixes on consonant-initial verb roots and with the essive prefix *ve-*, the only other C-initial prefix in Wamesa. Chapter 6 gives a full account of verbal infixation. A complete paradigm with both C- and V-initial verb roots is provided in (4.49).

21. Wamesa has a set of trial pronouns, as do several other Cenderawasih Bay languages (Anceaux 1961), but these are morphologically very transparent and take plural verbal agreement.

(4.49) *Wamesa Verbal Agreement Paradigm*

	Prefix	*api* 'to eat'	*pera* 'to cut'
Singular			
1sg	/i-/	*y-api*	*i-pera*
2sg	/bu-/	*bu-api*	*p<u>era*
3sg	/di-/	*di-api*	*p<i>era*
Dual			
1du incl	/tur-/	*tur-api*	*tu-pera*
1du excl	/amur-/	*amur-api*	*amu-pera*
2du	/mur-/	*mur-api*	*mu-pera*
3du	/sur-/	*sur-api*	*su-pera*
Plural			
1pl incl	/tat-/	*tat-api*	*ta-pera*
1pl excl	/amat-/	*amat-api*	*ama-pera*
2pl	/met-/	*met-api*	*me-pera*
3pl hum	/set-/	*set-api*	*se-pera*
3pl NH	/si-/	*si-api*	*si-pera*

The human/non-human contrast is the only gender distinction made in the agreement paradigm, a pattern which contradicts Greenberg's (1966: 58) Universal 37, which states that "A language never has more gender categories in nonsingular numbers than in the singular.". The examples in (4.50) and (4.51) show the identical verbal agreement marking for human and non-human subjects in the singular and dual; example (4.52) shows how they differ in the plural. Very similar patterns, featuring a human/non-human split in the 3pl forms only, are also found in the Biakic languages Biak and Dusner (van den Heuvel 2006: 157; Dalrymple & Mofu 2012: 9).[22]

(4.50) a. *Mararea=pa-i t<**i**>awa.*
child=DET-SG 3SG-fall
'The child falls' [S7 IMK]

22. Roon, another Biakic language, makes an animate/inanimate distinction in all numbers within the verbal agreement paradigm (Gil 2010).

b. *Anggadi=pa-i t<**i**>awa.*
coconut=DET-SG 3SG-fall
'The coconut falls' [S75 IMK]

(4.51) a. *Sinitu=pa sandu **su**-nda.*
person=DET they.two 3DU-go
'The two people go.' [S78 IMK]

b. *Wona=pa-i=e kodo=pa-i=e **sur**-ase.*
dog=DET-SG=and frog=DET-SG=and 3DU-swim
'The dog and the frog swim.' [S77 IMK]

(4.52) a. *Bajak laute=pa-i=ya, setoru **se**-maso warire aiku=pa-i.*
pirate=DET-SG=again 3TRI 3PL.HUM-sit around box-DET-SG
'The pirates again, they three sit around the box.' [S20 IMK]

*Aya=pa-si **si**-maso na ai=pa-i vavo=pa.*
bird=DET-PL 3PL.NH-sit LOC tree-DET-SG top=DET
'The birds sit on top of the tree.' [S20 IMK]

When the subject of the verb is a group consisting of both human and non-human actors (a boy, a dog, a frog, and a turtle, for example), the human agreement marker is used.

Subject agreement is obligatory for any word used predicatively, including not only true verbs as in the examples above, but also predicative adjectives and quantifiers, as well as words of other classes which have changed categories through zero-derivation or by the addition of the essive prefix to function as verbs. Example (4.53a) shows agreement on a predicative adjective, (4.55c) on a numeral with the essive prefix *ve*-, (4.75b) on a predicative quantifier, and (4.53d) on the adverb *manau* 'already', which conveys completive aspect.

(4.53) a. *Wona=pa-i m<**i**>eta.*
dog=DET-SG 3SG-black
'The dog is black.' [S29 IMK]

b. *Mararea* ***se**-mbe-rime=e* *siri.*
child 3PL.HUM-ESS-five=and one
'There are six children; the children are six.' [S20 IMK]

c. *Ririo=wa-si* ***si**-pau.*
leech=DET-3PL.NH 3PL.NH-many
'The leeches are many; there are many leeches' [S73 IMK]

d. *Yau* ***i**-manau* *y-unu* *kambu.*
I 1SG-COMPL 1SG-drink water
'I already drank water; I'm finished drinking water'[23] [S16 IMK]

There are two instances where verbs do not accurately agree with their subjects in all three of person, number, and animacy. Quantifiers specifically often fail to agree in number with their subject. *Pau* 'many' may bear either a singular or plural verbal agreement affix when its subject DP is marked as singular; if the subject DP is marked as plural, *pau* must also bear the plural agreement affix. The difference is that, while the gloss is 'many' in both cases, plural marking as in (4.54b) implies a greater number than singular, as in (4.54a) and (4.54c). Quirky agreement on quantifiers is not unusual cross-linguistically, and manifests, for example, in English constructions like *many a good student* (c.f. e.g. Quirk et al. 1985: 759, Cardinaletti & Giusti 2006 example (63) and endnote 4).

(4.54) a. *Ririo=pa-**i*** *p<**i**>au.*
leech=DET-SG SG-many
'The leeches are many; there are many leeches' (Lit. 'The leech is many.')

b. *Ririo=pa-**i*** ***si**-pau.*
leech=DET-SG PL-many
'The leeches are very many; there are very many leeches' (Lit. 'The leech are many.')

23. Compare this to *Yau inunu kambu manau* with the same meaning but no verbalization of *manau* 'COMPL'.

c. *Ririo=pa-**si*** ***si**-pau.*
leech=DET-PL PL-many
'The leeches are extremely many; there are extremely many leeches' [S73 IMK]

The second exception to obligatory number and animacy agreement is with numerals used predicatively with the essive prefix. The subject DP here is almost always marked as singular, but the predicate always receives the third-person plural human agreement marker. In cases with a human subject DP, such as example (4.55a), this leads to a number mismatch; when the subject DP is non-human, as in (4.55b), there is a mismatch in both number and animacy. Using the non-human 3pl agreement marker is ungrammatical.

(4.55) a. *I-samuai mararea=pa-**i*** ***se**-mbe-rime=e muandu. Mua=pa-**i***
1SG-gather child-DET-SG 3PL.HUM-ESS-five=and two. male=DET-SG
se**-mbe-rime. Vavi=pa-**i ***se**-mbe-muandu.*
3PL.HUM-ESS-five. female=DET-SG 3PL.HUM-ESS-two.
'I had seven children. There are five boys. There are two girls.' (Lit. 'I had a child they are seven. The boy are five. The girl are two.) [S2 IMK]

b. *Kelinci=pa-i* ***se**-mbe-rime.*
rabbit=DET-SG 3PL.HUM-ESS-five
'There are five rabbits; the rabbits are five.' (Lit. 'The rabbit are five.') [S20 IMK]

c. * *Kelinci=pa-i* ***si**-ve-rime.*
rabbit=DET-SG 3PL.NH-ESS-five
'There are five rabbits; the rabbits are five.'[S20 IMK]

4.4.2. Applicative *it-*

The applicative prefix *it-* has two functions: it may introduce an instrument, overt or not, or it may give aspectual information. The first use is more frequent in my data. If overt, the instrument always precedes the *it*-bearing verb. Whether these dual functions indicate a single polysemous morpheme or two homophonous affixes is arguable. Unlike the homophonous but distinct clitic pairs, such as determiner and aspectual *=pa* (§4.3.1, §4.3.5) and topicalizing and directional *=ma* (§4.3.3, §4.3.4), the two uses of *it-* cannot be divided by the type of phrase in which each appears. Unlike the essive prefix *ve-* (§4.4.4), *it-* never appears twice on the same root. The fact that the aspectual and instrumental meanings of *it-* do not co-occur, however, gives some support to the homophony analysis. I will treat *it-* here as a single polysemous affix, but the evidence is not unequivocal on this point.

The examples in (4.56) show *it-* used to introduce an instrument in a range of different relationships to the applicativized verb, including in an equitive clause, as the direct object of a higher clause, as the object of a PP, and with the applicative verb in a relative clause modifying the instrument.

(4.56) a. *Wai=ne-i=ma set-**it**-avakire sasu.*
rope=DET-SG=TOP 3PL.HUM-APPL-hang clothing.
'They use the rope to hang up clothing.' [S30 IMK]

b. *Nini=ma kai=wa-si wura y-**it**-ane ana.*
this=TOP papeda.spoon=DET-PL all 1SG-APPL-eat papeda
'These are all the *kai* I use to eat papeda.' [S41 IMK]

c. *Yau i-rute aivasore=ne-i y-**i**-ndora kamumi=pa-i.*
I 1SG-hold sandal=DET-SG 1SG-APPL-hit mosquito=DET-SG
'I use a sandal to hit the mosquito.' [S41 IMK]

d. *I-maso na wa=pa=i=e y-**i**-mbavou=ra.*
1SG-sit LOC canoe=DET-SG=and 1SG-APPL-go.home=to.there
'I sit in the canoe and use it to go home.' [S68 IMK]

e. *Kai=pa-i y-**i**-susa ana=pa-i t<i>au rawave.*
papeda.spoon=DET-SG 1SG-APPL-turn papeda=DET-SG 3SG-fall down.below
'The *kai* I use to wind up papeda fell down.' [S70 IMK]

One restriction on *it-* is that it may not introduce a human instrument. Animals are acceptable, as in (4.57a), but people and body parts are not, as in (4.57b) and (4.57c).

(4.57) a. *Wona=ne-si y-**it**-awer pimuna=pa-i.*
dog=DET-PL 1SG-APPL-hunt pig=DET-SG
'I use the dogs to hunt the pig.' [S58 IMK]

b. * *Sinitu=wa-i y-**i**-mbori tiket=pa-i.*
person=DET-SG 1SG-APPL-buy ticket=DET-SG
'I use that man to buy the ticket.' [S70 IMK]

c. * *Vara=ne-i y-**i**-ndute ana=ne-i.*
hand=DET-SG 1SG-APPL-hold papeda=DET-SG
'I use my hand to hold the papeda.' [S70 IMK]

Instruments introduced by *it-* may be optionally dropped, particularly when they are already named and salient in the discourse. In this case, there may be some ambiguity as to whether the applicative is being used instrumentally or aspectually. Without an overt instrument to force the instrumental reading, it is contextual cues and the nature of the activity beng described (and its likelihood of involving an instrument) that determine the interpretation.

(4.58) a. *Aivasore=ne-i y-**it**-apai vera do re=wa.*
shoe=DET-SG 1SG-APPL-run towards to land=DET
'I use these shoes to run inland.' [S70 IMK]

b. *Y-**it**-apai vera do re=wa.*
1SG-APPL-run towards to land=DET
'I use (it) to run inland' (if shoes are salient); otherwise 'I'm about to run inland.' [S68 IMK]

c. *Yau y-**i**-mbui.*
I 1SG-APPL-write.
'I use (it) to write; I'm about to write' (equally likely). [S81 IMK]

There are other ways to introduce instruments as well. One is for the instrument to appear as the object of the preposition *tuti* 'with'. This construction can be ambiguous with the comitative, depending on the plausibility of the resulting scenario. As in English, the example in (4.59a) 'I eat fish with rice' can mean either that I used rice to pick up the fish and eat it (as is customary in some parts of Indonesia), or that I ate fish and rice as part of the same dish. The other strategy for instrumentals is to use the verb *rute* 'hold', as in (4.59b). Use of *rute* is not mutually exclusive with use of instrumental *it-*, as in example (4.56c) above.

(4.59) a. *Yau y-ane dia=ne-e tuti pas/sendo=pa-i.*
I 1sg-eat fish=det-sg with rice/spoon=det.
'I eat the fish with rice/the spoon.' [S77 IMK]

b. *Yau i-rute pas/sendo=pa-i y-ane dia=pa-i.*
I 1sg-hold rice/spoon=det-sg 1sg-eat fish=det-sg
'I eat the fish using rice/the spoon.' (Lit. 'I hold rice/the spoon I eat the fish.') [S77 IMK]

In its aspectual reading, *it-* most often has an prospective meaning, translated into Malay as *langsung*. Cowan (1955) reports *it-*, which he misinterprets as *in-* due to its triggering of r-splitting in clusters, as a progressive marker; Saggers (1979) has it as marking imperfective aspect. Both of these readings can arise, given interactions with the verb itself and other TAM markers in the clause, but the primary use is prospective, as in the examples in (4.60).

(4.60) a. *Yau y-**it**-ena.*
I 1sg-appl-sleep.
'I'm about to go to sleep.' [S68 IMK]

b. *Set-**it**-isa pimuna=pa=i.*
3pl.hum-appl-stab pig=det-sg.
'They're about to spear the pig.' [S80 IMK]

In combination with other aspectual markers, *it-* can lend other aspectual meanings. Combined with *=pa* 'not yet' and *manau* 'already, compl', *it-* usually gives a past/completive reading; speakers often reported the event described as having happened 'yesterday'. There are exceptions to this, as in (4.61b), which was translated as happening *sudah sekarang* 'just now already'; whether these exceptions are consistent within individual verbs is as yet unknown. If an instrument is already salient in the discourse, *it-* can simultaneously convey instrumental and aspectual readings, as in (4.61c).

(4.61) a. *Yau y-**i**-ndina kavio Windesi=pa/manau.*
I 1SG-APPL-know language Windesi=PA/already
'I already know the Windesi language.' [S68 IMK]

b. *Yau y-**i**-mbui=pa.*
I 1SG-APPL-write=PA
'I'm just now already writing.' [S68 IMK]

c. *Wa nini=ma yau y-**i**-mbavou=ra. Yau y-**i**-mbavou=pa.*
canoe this=TOP I 1SG-APPL-go.home=to.there. I 1SG-APPL-go.home=PA
'I use this canoe to go home. I used it to go home before.' [S68 IMK]

As mentioned previously in §4.2.2, *it-* also has complex aspectual interactions with its host verb, yielding a range of interpretations. The examples in (4.62) are reproduced from (4.7). Some of these interactions are predictable; predicative adjectives, for example, almost always get a durative, individual-level (rather than stage-level) interpretation, as in (4.63), which is ungrammatical when combined with completive *manau*. Others are lexically determined; no pattern arises based on the verb's *aktionsart*, its transitivity or status as unaccusative or unergative, or other features. For example, *voru* 'die' becomes past/completive with *it-*;[24] the related achievement verb *muna* 'kill' takes the default prospective reading. 'Sleep' and 'know' in examples (4.60a) and (4.61a) are examples of other verbs which, like 'die', take patients as subjects, but which take the prospective reading when other aspectual markers are not present.

24. No instrumental reading is available here, though this may be the result of cultural/religious taboo against suicide.

(4.62) a. *Yau i-nai na Bintuni.*
I 1SG-be.at LOC Bintuni
'I'm in Bintuni.' (unmarked tense/aspect) [S68 IMK]

b. *Yau y-**i**-nai na Bintuni.*
I 1SG-APPL-be.at LOC Bintuni
'I live in Bintuni.' (durative interpretation) [S68 IMK]

c. *Yau y-**i**-nda so Bintuni.*
I 1SG-APPL-go to Bintuni
'I'm about to go to Bintuni.' (prospective interpretation) [S68 IMK]

d. *Yau y-**i**-mboru.*
I 1SG-APPL-die
'I already died.' (completive interpretation) [S68 IMK]

(4.63) *Sia=vura set-**it**-ate.*
they=all 3PL.HUM-APPL-good
'They are all beautiful people.' [S71 IMK]

With stative verbs and predicative adjectives, *it-* can also have an intensifying effect, as in (4.64).

(4.64) a. *Set-**i**-sanevesie.*
3PL.HUM-APPL-happy
'They are very happy.' [S80 IMK]

b. *Set-**i**-ndina.*
3PL.HUM-APPL-know
'They really know; they know well.' [S80 IMK]

4.4.3. Causatives

The caustive prefix *on-* is derived from the verb 'give', also *on.* The use of the prefix is very limited; most constructed examples failed and few natural examples appeared in my corpus. The sentence in (4.65a), one of the few accepted by my consultants, comes from

Cowan (1955: 56).[25] When it is used, the affected entity must be human, never animal or inanimate. By far the preferred way to indicate causation is with the independent verb *on* 'give', as in (4.65c). This is parallel to the equivalent Papuan Malay construction, which uses the verb *kasih* 'give' (rather than the causative suffix *-kan* found in Standard Indonesian) to express causativity.

(4.65) a. *Sur-**o**-ndama.*
3DU-CAUS-come
'They two make (him) come.' [S46 IMK]

b. * *Yau y-**o**-nggasio andi.*
I 1SG-CAUS-angry he/she
'I make him angry.

c. *Yau y-**on**-i k<i>sio.*
I 1SG-give-SG 3SG-angry
'I make him angry.' [S47 IMK]

Causative *on-* can appear together with applicative *it-*. When this occurs, *it-* precedes *on-*; no examples with the order reversed were accepted by my consultants. It isn't obvious what function the applicative prefix is playing in the sentence in (4.66); as there is no clear instrument it is most likely an aspectual use not explicitly translated by the speaker.

(4.66) *Y-it-**o**-kavio kavio Wamesa.*
1SG-APPL-CAUS-speak language Wamesa.
'I cause (him) to speak the Wamesa language.' [S71 IMK]

4.4.4. Essive *ve-*

The essive prefix *-ve* has two main uses. It can create adjectives with the meaning 'having the properties of X' (where X is the root), and it can introduce relative clauses. In the

25. The prefix surfaces as [o-] rather than [on-] in these examples because the /n/ deletes adjacent to the root-initial C as part of cluster simplification (see §2.3.2).

former usage it can attach to a range of word classes. The examples in (4.67) show *ve-* used with adjectives (4.67a) - (4.67b), quantifiers (4.67c), and verbs (4.70c).

(4.67) a. *Wona* ***ve-****meta=pa-i di-ase.*
dog VE-black=DET-SG 3SG-swim
'The black dog swims.'

b. *Yau i-vori sasu sama* ***v****<i>-urar.*
I 1SG-buy clothing buttocks 3SG-VE-red
'I buy red pants." (Lit. 'I buy pants, it is red.')

c. *Suomuse* ***ve-****pau si-nai na rawana=wa.*
shark ESS-many 3PL.NH-be.at LOC sea=DET
'Many sharks are in the ocean.'

The above categories can freely appear attributively or predicatively. Nouns and numerals, by contrast, require the addition of *ve-* to be used predicatively. Subject agreement morphology cannot attach directly to these roots, so *ve-* is required to mediate.

(4.68) a. *Simuti=pa-i* ***v****<i>****e****-buo.*
orange=DET-SG 3SG-ESS-fruit
'The orange (tree) is fruiting/has fruit.' [S60 BMW]

b. *Bajak laute=pa-i se-****mbe****-toru venasia.*
pirate=DET-SG 3PL.HUM-VE-three only
'There are only three pirates; The pirates number only three.' [S20 IMK]

Ve- is not grammatical with other word classes, such as prepositions, as in (4.69).

(4.69) * *Sinitu* ***ve-****tuti=pa-i k<i>opa.*
person ESS-with=DET-SG 3SG-jump
'The accompanying person jumps.' [S80 IMK]

In its capacity as a relativizer, *ve-* can attach to adjectives and verbs. An example of this is given in (4.70a). When *ve-* is used with adjectives and verbs, which use is intended can be ambiguous, as in (4.70b); in this and many related examples the semantic difference between the two readings is quite subtle. In the 'properties of' usage with verbs and adjectives, *ve-* gives an individual-level or habitual meaning to the property being described (Carlson 1977), similar to that found with the aspectual use of *it-* with adjectives, as in (4.70c).

(4.70) a. *Yau y-ane aibuo* ***ve****-nai na meja=pa vavo=pa.*
I 1SG-eat fruit ESS-be.at LOC table=DET top=DET
'I eat the fruit which is on top of the table.' [S42 IMK]

b. *wona* ***ve****-meta=pa-i*
dog ESS-black=DET-SG
'The black dog; the dog which is black'.

c. *wona* ***ve****-kari*
dog ESS-bite
'a bitey dog; a dog that often bites' [S42 IMK]

The property-attributing and relativizing uses of *ve-* can be used simultaneously by double *ve-* prefixation on a single word, with one *ve-* contributing each of the two meanings. The most common usage of this is to create ordinal numbers, literally 'that which has the property of being (i.e.) two', as in (4.71a). Another frequent example involves the word *rawana* 'sea'. Adding the essive prefix gives *ve-rawana* 'having the properties of the sea', an oft-used metaphor for 'blue'. Adding a second relativizing *ve-* yields a word meaning 'which has the properties of the sea', or 'which is blue', as in (4.71b).

(4.71) a. ***Ve-ve****-muandu kota t<i>pu dia katu=pa-i.*
ESS-ESS-two also 3SG-grasp fish small=DET-SG
'The second one is also holding a little fish.' [S20 IMK]

b. *Di-vute aiku* ***ve-ve****-rawana=pa-i, di-siwar-i.*
3SG-guard box ESS-ESS-blue=DET-SG, 3SG-lean-3SG
'He's guarding the blue box, he's leaning on it.' [S20 IMK]

One other use of *ve-* is to make loan words usable as verbs. In the examples in (4.72) and (4.73), *olaraga* is the Malay word for 'exercise' and *spid* is a loan via Malay from Dutch meaning 'speedboat' (see also §2.4.3). Attaching the verbal agreement markers to these directly is ungrammatical; adding *ve-* makes agreement grammatical.

(4.72) a. ***V***<*i*>***e****-olaraga.*
3SG-ESS-exercise
'He exercises.' [S78 IMK]

b. * *Di-olaraga.*
3SG-exercise
'He exercises.'

(4.73) a. *I-**ve**-spid.*
1SG-ESS-speedboat
'I go by speedboat.' [S64 IMK]

b. * *I-spid.*
1SG-speedboat
'I go by speedboat.' [S64 IMK]

4.4.5. Number Marking: NPs and DPs

There are two sets of related but distinct suffixes which take the forms *-i* and *-si*, marking singular and plural number, respectively. These are clearly related to the third-person singular and plural non-human verbal agreement markers *di-*/*-i-* and *si-*, though those are prefixal(/infixal) while these are suffixal. The first pair of *-i*/*-si* suffixes almost exclusively co-occur with the determiners, though occasionally also on nouns, marking number on the DP or NP. The second pair, which co-occur with the verb, are discussed in §4.4.6.

In most cases, the DP must be marked for number. Number marking can be done in either of two ways: by means of a number suffix on the determiner, or by a following numeral. These two strategies are mutually exclusive; use of both together is ungrammatical, as in (4.74c) and (4.74d).

(4.74) a. *wona=pa-**i***
dog=DET-SG

'the dog'

b. *wona=pa-**si***
dog=DET-PL
'the dogs'

c. *wona=pa(*-**i**)* *siri*
dog=DET(*-3PL.NH) one
'the one dog'

d. *wona=pa(*-**si**)* *siaran*
dog=DET(*-3PL.NH) hundred
'the hundred dogs'

Number marking within DPs with a quantifier depends on the choice of quantifier; the universal quantifier *vura* 'all', requires the preceding determiner to be marked for number, while others such as *pau* 'many' prohibit it when used attributively. The syntactic reasons for this distinction were discussed in §4.3.2.

(4.75) a. *wona=pa*(-**si**)* *vura*
dog=DET*(-3PL.NH) all
'all the dogs'

b. *wona=pa(*-**si**)* *pau*
dog=DET(*-3PL.NH) pau
'the many dogs'

Because there is no dual number marking suffix, the only way to indicate duality is by using the numeral *muandu* 'two', as in *wona pa muandu* 'the two dogs'.

In the plural, if number is marked by a suffix on the determiner rather than by a numeral (i.e. *wona pai* 'the dog' rather than *wona pa siri* 'the one dog'), that suffix must agree with the head noun in animacy and, if the head noun is human, in person as well. The number markers used for human subjects are homophonous with the pronouns. Compare (4.76a) - (4.76c) with (4.74a) - (4.74c), their non-human counterparts, above. Example

(4.76e) gives an example of dual number marking, which is possible by suffix rather than numeral when the referent is human.

(4.76) a. *sinitu=pa-**i***
person=DET-SG
'the person'

b. *sinitu=pa-**sia***
person=DET-3PL.HUM
'the people'

c. *sinitu=pa(*-**i**)* *siri*
person=DET(*-SG) one
'the one person'

d. *sinitu=pa(*-**sia**)* *siaran*
person=DET(-*3PL.HUM) hundred
'the hundred people'

e. *sinitu=pa-**mia***
person=DET-2PL.HUM
'you two people'

The main exception to obligatory number marking on the DP is when it is plural and the subject of the clause. When a predicative adjective or quantifier agreeing with the DP directly follows it, marking on both is dispreferred; otherwise it usually occurs. From a phonological point of view, the resulting structure is particularly vulnerable to reduction: in most cases marking on both elements leads to a sequence of two adjacent identical, unstressed, monosyllabic morphemes with the same meaning. While this may explain the higher frequency of affix omission in these cases, it cannot be the sole reason for dropping the agreement marker, as there are cases as in (4.78a) where the *si-* affixes are not linearly adjacent, and others such as (4.77b) where they are not identical.

(4.77) a. *wona=pa(-**si**)* ***si**-pau*
dog=DET(-3PL.NH) 3PL.NH-many

'The dogs are many, there are many dogs'

b. *Sinitu=pa(**-sia**) **se**-pau*
person=DET(-3PL.HUM) 3PL.HUM-many
'The people are many, there are many people'

(4.78) a. *Wona wa(**-si**) **si**-ase.*
dog DET(-3PL.NH) 3PL.NH-swim
'The dogs swim.'

b. *Wona wa pau **si**-ase.*
dog DET many 3PL.NH-swim
'The many dogs swim.'

In the case of the plural, it is sufficient for number to be expressed by the agreement of the verb with its subject. Evidence for this comes from (4.79), where the inability of the predicative adjective *pimasa* (see §5.4) to agree with its subject makes dropping the number agreement on the DP impossible.

(4.79) *Wona wa*(**-si**) pimasa.*
dog DET(-3PL.NH) big
'The dogs are big.'

The one other exception to obligatory DP number marking, and the only case where number marking may fail to occur in the singular, is with locational nouns and their possessors. This includes both nouns denoting a part of something else (*vavo* 'top', *raro* 'inside') and the special class of geographical locational nouns (i.e. *rau* 'sea', *re* 'land', etc., as defined in §5.6.2). Possessors of locational nouns tend to have number marking while the locational nouns themselves tend not to; presence of number marking on the locational implies number marking on the possessor, but not vice versa. Plurality must be overtly marked on the possessor, otherwise it is interpreted as singular.

(4.80) a. *na meja=pa(**-i**) vavo=pa(**-i**)*
LOC table=DET(-3SG) top=DET(-3SG)

‘on top of the table’

b. *na meja=pa*(-**si**) vavo=pa(-**si**)*
LOC table=DET*(-3PL) top=DET(-3PL)
‘on top of the tables’

(4.81) *do rau=wa(-**i**)*
to sea=DET(-3SG)
‘to the sea’

When a determiner is present, it is always the bearer of suffixal number marking. There are, however, cases in which *-i* or *-si* appears on the noun when no determiner is present. These are rare in my data, though not non-existent, and tend to be nouns where number is salient, in need of disambiguation, or surprising. The word for ‘finger’, *varakia*, for example, sometimes appears in the singular with the *-i* suffix as *varakia**i***, but never with *-si* as *varakia**si*** in the plural, as fingers tend to come in groups and so the singular is the marked form. Example (4.82) shows the singular marker *i* on a proper name, Sonya; its appearance here helps to disambiguate the antecedent of the following determiner, which is *sinitu* ‘person’, not *Sonya*, since in this case *sinitu* ‘person’ is plural while *Sonya* is singular.

(4.82) *Sinitu ve-rora Sonya-i=pa-sia se-nda vera wana.*
person ESS-hit Sonya-SG=DET-3PL.HUM 3PL.HUM-go to there
‘The people who hit Sonya went over there.’ [S78 IMK]

Further evidence of the possibility of using number suffixes on bare nouns comes from Cowan (1955), who cites forms such as *anio-si* ‘houses’ and *dian-si* ‘fish’, with the *-si* plural suffix directly on the noun, though these are given in isolation, so there is no way to know whether or where a determiner might have appeared. In my data, a contrast also appears to be fossilized in nouns like *asarua**i*** ‘sea urchin’ versus *asarua**si*** ‘sand dollar’,[26]

26. Why the sand dollar should be plural while the urchin is singular is unclear. This is clearly not synchronically decomposable into a noun *asar* plus determiner *wa-i/si*, as the addition of a determiner is

and in the lexicalized substantive *baba**i*** 'older sibling', presumably derived from *baba* 'big'. This is not possible on other categories of words, such as adjectives, as shown, for example, by the ungrammaticality of (4.83).

(4.83) * *katu-**si***
small-PL
'small (pl)'

That it is the determiner which bears the number marking affix and not normally the noun argues for a DP structure rather than an NP. As Anderson et al. (2006) point out, it is common for a morphosyntactic property assigned to a whole phrase to be marked at only one point in that phrase, generally on the head or at an edge. This describes number in Wamesa, which is marked only once, normally on the determiner. Though in practice this most often puts the number suffix in an edgemost position, this fact is epiphenomenal - when the DP includes a numeral or quantifier, the determiner to which the number suffix attaches is penultimate, as previously shown for example in (**??**). Number inflection is not attracted to the edge of the phrase; it is attracted to the head: the determiner.

4.4.6. Number Marking: Verbs

The second *-i*/*-si* suffix pair appears as a dummy pronoun on many transitive verbs when a following argument, usually the direct object, is not overt or has been raised past the verb, filling its place. In this case the singular marker *-i* is the default; plural *-si* is only used when the plurality of the missing argument is salient, equivalent to saying 'I hit them' in English, rather than 'I hit it' or 'I hit', both of which would require the use of singular *-i*. The examples in (4.84) show the verb *rora* 'hit' both with an overt object and without it, with *-i* and *-si* filling that argument slot.

possible, as in *asaruai pai/pasi* 'the urchin/s'.

(4.84) a. *Yau i-rora kamumi=pa-**i**/**si**.*
I 1SG-hit mosquito=DET-SG/PL
'I hit the mosquito/s.'

b. *Yau i-rora-**i**.*
I 1SG-hit-3SG
'I hit (it/something).'

c. *Yau i-rora-**si**.*
I 1SG-hit-3PL.NH
'I hit (them/something pl).'

In (4.85), *-si* takes the place of DO argument co-referential with a noun appearing earlier in the sentence.

(4.85) *Dia=pa si-kai, y-a-**si** vura.*
fish=DET 3PL.NH-empty 1SG-eat-3PL.NH all
'The fish are all gone, I ate (them) all.' [S77 IMK]

These same suffixes appear not only when the argument is missing entirely, but also when it has moved so as to no longer surface directly following the verb in the linear order. An example of this is given in (4.86), where *dia pai* 'the fish', the direct object of the verb *yisane* 'I stab', is topicalized by raising to the left edge of the sentence. Like *verora*, this verb, too, receives a singular DO marker to fill the place of the moved DP.[27]

(4.86) *Dia=pa-i y-isane-**i**.*
fish=DET-SG 1SG-spear-SG
'The fish, I stabbed (it).' [S10 IMK]

The preceding examples all show these suffixes taking the place of a direct object. While this is the most frequent use of post-verbal *-i* and *-si*, they can also be used to take the place of other types of arguments. In Wamesa, the word *miso* 'sit' requires a locational

27. The sentence in (4.86), while not actually passive itself, is typical of constructions given as translations of Malay passive sentences.

argument; when this is not expressed by a prepositional phrase, *-i* is required to take its place.

(4.87) a. *M<u>so na karapea=pa-i.*
2SG-sit LOC chair-DET-SG
'You sit on the chair.'

b. *M<u>so-**i**.*
2SG-sit-3SG
'You sit.'

This unmarked, indefinite singular argument suffix contrasts with the homophonous full definite 3rd person singular DO pronoun, which bears stress, is set off from the preceding word by a pause and/or glottal stop, which is non-phonemic but may precede V-initial words, and refers to a specific entity.

(4.88) a. *Yau i-sera-**i**.*
[ˈjau iseˈra**j**]
I 1SG-hit-SG
'I see (something).'

b. *Yau i-sera **i**.*
[ˈjau iˈsera ˈ**ʔi**]
I 1SG-hit 3SG
'I see it/him.'

Further evidence that suffixal *-i* and pronominal *i* denoting a singular argument are distinct comes from the plural. If they were the same morpheme able to separate from the verb root, we would expect the same to be possible in the plural. There is, however, no equivalent homophonous plural DO pronoun; a sentence similar to that in (4.88b) but with freestanding *si* is ungrammatical. The pronoun *sia* '3PL.HUM' or number-marked determiner *=pa-si* 'DET-PL', with a null NP head in its DP, must be used instead, as in (4.89).

(4.89) a. *Yau i-sera-**si**.*
[ˈjau iseˈra**si**]
I 1SG-see-PL
'I see (them/something pl).'

b. * *Yau i-sera **si**.*
[ˈjau iˈsera ˈ**si**]
I 1SG-see PL
'I see them.'

c. *Yau i-sera=**pa-si**.*
I 1SG-see=DET-PL
'I see them.'

The nominal and verbal number suffixes can co-occur within a single clause. This is shown in (4.90), where the direct object DP includes a relative clause with a transitive verb and no overt DO. The first *-i* to follow the verb marks the direct object, which is not overtly realized as a NP. The second allows the determiner to agree in number with the head noun, *sinitu.* These are not to be confused with the prefixal *i-* on the main verb *rina* 'know', which marks agreement with a first person singular subject.

(4.90) *Yau i-rina sinitu ve-rora-i=pa-i.*
I 1SG-know person ESS-hit-3SG=DET-SG
'I know the person who wants to hit (someone/thing).' [S42 IMK]

The above account covers most instances of the *-i* and *-si* suffixes, but there are some instances where they appear unexpectedly, or fail to appear where they should, which are so far unaccounted for. The first case involves co-occurrence with the applicative prefix *it-*. In its non-aspectual uses, *it-* introduces an instrument. It is tempting to describe *it-* as changing the valency of the verb on which it appears, promoting the instrument from adjunct to argument. If this were the case, however, we would expect to see number marking on the verb when that instrument is not overt, just as we do for a non-overt direct object or locational argument. This, however, is not the case. As shown in the examples in (4.91), verbs bearing applicative morphology do not recieve an object number suffix with non-overt instruments, shown in the examples in (4.91). They do, however, receive it with a non-overt object, as in (4.92).

(4.91) a. ***Set-i-mbosa.***
3PL.HUM-APPL-paddle
'They use it to paddle.' [S80 IMK]

b. *Yau* ***y-i-mbavou****=pa.*
I 1SG-APPL-go.home=PA
'I already used it to go home.' [S68 IMK]

(4.92) *Sia=vura set-it-kutu-i.*
they=all 3PL.HUM-APPL-cut-3SG
'They use it to cut (something).' [S42 IMK]

It appears that the applicative prefix, while it marks the presence of an instrumental participant in the event, does not actually change the valency of the verb. This behavior is attested in Bardi, with the second of that language's two applicative markers Bowern (2012: 489, 495). Instrumental applicatives in Bardi can surface marked as an adjunct with the instrumental suffix even when the verb bears the applicative affix. This is also the case for the applicative in Abaza, a Northwest Caucasian language (O'Herin 2001).

There are also scattered instances of extra or missing *-i* which are less explicable and do not seem to follow a pattern. These may be speech errors, mis-transcriptions, extreme reductions of a vowel in word-final position, evidence of an additional not-yet-uncovered pattern in the data, or some combination of the above.

4.4.7. Inalienable Possession

As previously discussed, inalienable possession is marked in the first- and third-person singular by means of the definite determiners *=ne* and *=pa*. The second person singular adds the suffix *-mu(i)* to the possessum. The choice between *-mu* and *-mui* seems not to be a principled one; van den Berg (2009) reports too that variation between the two appears essentially random in his data. In the non-singular, the possesum is marked with the same agreement prefixes used for verbal subject agreement, in this case agreeing with the possessor, as well as the suffix *-mi*. A full paradigm for inalienable possession is given in (4.93) for the word *tama* 'father'; with a first person singular possessor, the suppletive form *yai* is used.[28]

(4.93) *Inalienable (Direct) Possession*

Person	Singular	Dual	Plural
1incl		*tu-tama-mi*	*ta-tama-mi*
1excl	*yai=ne-i*	*amu-tama-mi*	*ama-tama-mi*
2nd	*tama-mu(i)*	*mu-tama-mi*	*me-tama-mi*
3rd	*tama=pa-i*	*su-tama-mi*	*setama-mi* (human) *si-tama-mi* (non-human)

The fully-affixed possessum behaves as any other noun within a DP. The possessor may be omitted or may directly precede the possessum DP. Examples of inalienably possessed kin and body part terms used in context are given in (4.94).

(4.94) a. *Wona=ne-si vura* ***si-ru-mi****=pa=si* *si-pota.*
dog=DET-PL all 3PL.NH-head-POSS=DET-PL tpln-sick
'All the dogs' heads hurt.' [S8 IMK]

b. *(Tata=vura)* ***ta-sane-mi****=ne-si* *si-pota.*
(we.INCL=all) 1pl.INCL-stomach-POSS=DET-PL 3PL.NH-sick
'All our stomachs hurt.' [S29 IMK]

c. *(Au)* ***sumo-mui*** *r<i>ane dia.*
(You) aunt-2sg.POSS 3SG-boil fish.

28. The word for 'mother', *sinia*, also has a suppletive form, *avini*, with a first-person singular possessor. *Yai* and *avini* are ungrammatical with other person/number combinations.

'Your aunt boils fish.' [S10 IMK]

As discussed in §4.3.1, inalienable or direct possession is used in Wamesa for all kin terms and most body parts. The inalienable construction is preferred for human possessors and dispreferred for non-human animal possessors, and not grammatical for body parts which humans do not have, such as wings and tails. Alienable (indirect) possession is indicated using the verb *ne* 'have', and is possible with all possessums, including those kin terms and body parts which may be inalienably possessed. This construction is shown in (4.95a).[29] Less common is juxtapositional possession, where the possessum directly follows the possessor in the sentence, with no linking element. This is most often used for partitive possession, particularly with locational nouns, but is available for all possessors and possessums. An example of juxtapositional possession is given in (4.95b).

(4.95) a. ***I-ne wona***=*ne-i b<i>ba=va.*
1SG-have dog=DET-SG 3SG-big=NEG
'I have a small dog; my dog is small.' [S6 IMK]

b. *wonggei=wa-i ponori=pa-si*
cassowary=DET-SG egg=DET-PL
'the cassowary's eggs' [S79 IMK]

Using the alienable, rather than inalienable, does not result in a change of meaning or a change in the sort of relationship implied between possessor and possessum. This is shown in (4.96), where the inalienable construction given first is preferred but the alienable construction using *ne* 'have' is also acceptable.

(4.96) a. *vara=ne-i*
hand=DET-SG
'my hand' (inalienable) [S29 IMK]

29. *biba va* in this sentence literally means 'is not big'; I translate it here as 'is small' because that is the translation provided in this case by my consultant. ('Anjing saya kecil.')

b. *i-ne* *vara*
1SG-have hand
'my hand' (alienable) [S29 IMK]

The option of using the alienable 'have' construction for kin terms and body parts appears to be a new development in the language. Van den Berg (2009) discusses grammatical possession in a number of SHWNG languages; in his data the inalienable construction is obligatory for kin terms and body parts. This shift to optionality may have come about under influence from Malay, which has no alienability distinction.

Chapter 5

Word Classes

5.1. Differentiating Word Classes

An ongoing debate in the typological literature is whether and how languages differentiate between different categories of words. On one side are what Croft (2000) calls the 'lumpers', who argue against the idea of separate noun, verb, and adjective classes as a linguistic universal. Some classic examples used as evidence for this position include the Salish languages, which Kinkade (1983) and others argue distinguishes only between predicates and particles; Quechua (Weber 1989; Hengeveld 1992), which is argued not to disinguish adjectives from nouns; and Tagalog (Gil 1993 et seq.), Riau Indonesian (Gil 1994, 2001), and several Polynesian languages, which are claimed to have only a single word class, S. On the other side are the 'splitters'; for nearly every one of the above languages - van Eijk & Hess (1986) and Davis & Matthewson (1999) for Salish, Aldridge (2009) and Richards (2009) for Tagalog, Floyd (2011) for Quechua - there is an opposing argument that the classes in question can in fact be distinguished, often after taking into account a larger sample of the lexicon or by applying a different set of criteria or level of granularity - or thoroughness - for what counts as 'the same behavior'. Croft (1990: 141) declares that “one of the few unrestricted universals is that all languages have nouns and verbs”, an assertion echoed by Davis & Matthewson (1999) and Evans & Osada (2005), though for the former it is a theoretical necessity that this be so, while the latter see the existence of counterexamples as non-problematic, though they dismiss those previously presented.

Evans & Osada (2005: 366) lay out three criteria for rigorously determining that two prototypical word classes are in fact merged in a language. First, all members of the macro-class must be “distributionally equivalent” both morphologically and syntactically: they must appear in the same environments. Secondly, any differences in meaning which result from the use of one type of root in a slot more usually associated with another type of root (i.e. a Salish word which translates to English as a concrete noun, used predicatively) must be directly derivable through compositional principles. These effects must be bidirectional; to claim that there is one category encompassing nouns and verbs it is not enough that prototypical nouns may be used in verbal environments, but prototypical verbs must be equally usable in nominal environments. Thirdly, these facts must hold across the lexicon, not simply for a conveniently illustrative set of examples. Evans and Osada accept claims that not all languages have a separate class of adjectives, but argue that all existing claims of single-category languages (those which merge at least nouns, verbs, and adjectives into one) have failed on at least one of the above criteria.

These are the criteria which will be used here to distinguish Wamesa word classes, with one major caveat. As Haspelmath (2012: 118) points out, looking for an *exact* match in morphosyntactic distribution of forms to define a class leads to a near-endless division into smaller and smaller categories and subcategories. A broader class of verbs, however useful, would be impossible to define, because certain verbs, systematically or idiosyncratically, have slightly different syntactic and morphological possibilities. Instead we are forced to separate out transitive vs. intransitives, those verbs which can take an applicative or causative prefix vs. those that cannot, those which require a locative argument vs.

those which do not, and all combinations of these (and other) features as exclusive groups. Some of these are useful distinctions to make, such as transitivity, while others, such as morphologically vs. periphrastically causativized verbs,[1] may not be. Strict adherence to this test, however, gives us no flexibility in the matter.

For usefulness' sake, then, rather than innumerable mini-classes, the Wamesa lexicon here will be divided into a small number of classes, some of which are further subdivided into not-necessarily mutually-exclusive subclasses. The choice of when to stop splitting and begin lumping, as Haspelmath points out, lacks an objective basis; Baker (2003: 5-6) calls it an "unanswerable question" whose answer comes down to "taste and terminology". In Wamesa, splits can clearly be made between groups of items resembling English nouns, verbs, and adjectives, so, Eurocentric though it may well be, those are the major categories I will use. An argument could be made for placing adjectives as a subclass of verbs, as their behavior when used predicatively is the same, though a distinction must be drawn at least at the class/subclass level due to the inability of verbs but not adjectives to appear as bare roots modifying the noun. As Haspelmath further points out, the terms 'noun', 'verb', and 'adjective' may not be appropriately applicable to all languages; for convenience and clarity I continue to use those terms, rather than 'Class A, B, & C' or his more-specific 'thing-root', 'action-root', and 'property-root'. While less precise than his terms, they allow us here to consider both a larger proportion of the lexicon and the category-changing processes which allow items to move from one class to another but may in some cases leave us with more than a bare root.

Wamesa, which does treat its nouns, verbs, and adjectives differently, will have little to add to the question of whether languages can exist which do not differentiate between these categories. This is an important question, as Baker (2001) and Evans & Osada (2005)

1. See §4.4.3.

point out, as a requirement that languages must distinguish these classes puts a significant restriction on what is and is not a possible natural language. A language with only one supercategory is easy to imagine; an example would be predicate calculus, as used in logic and formal semantics. What Wamesa can speak on is the question of whether words are underlyingly specified as to their class or only assigned based on the constructions or syntactic positions in which they appear. Like English, Wamesa is fairly flexible as to word class; zero-derivation, particularly into the category of verbs, is common. But there are some hard restrictions: nouns, for example, cannot appear predicatively without the addition of the essive prefix *ve-*; adjectives, verbs, and at least some adverbs and prepositions can. Other types of words - question words, determiners, pronouns - cannot be used predicatively at all. Davis & Matthewson (1999) use a similar pattern in English, the inability of many adjectives to appear in a nominal construction without derivational morphology such as *-ness*, to argue against functional rather than intrinsic determination of word classes. If word class in Wamesa were under-specified underlyingly, we would expect to be able to find any type of root in any type of construction, with a purely compositional meaning; that this is not the case points towards intrinsic specification of word class.

5.2. Nouns

The class of nouns in Wamesa includes both concrete and abstract entities, as well as locational nouns, which denote parts of other objects (top, inside, bottom) and are discussed in more detail in §5.6.2. Nouns can act as the subject of the sentence, as indicated by the agreement morphology on the verb (see §4.4.1). They can participate in possessive constructions, either as possessor or possessum (see §4.4.7). And finally they can serve as the head of a Noun Phrase, modified by any adjectives and directly preceding the determiner within the DP, if a determiner is present. Pronouns and proper nouns can appear without accompanying material in the DP; other nouns usually appear with an overt determiner. Generic plurals do not require a following determiner, and sometimes the determiner is omitted even though it is required by the grammar, likely under influence from Malay.

These properties allow us to test for noun-hood in several ways. If a word followed by no more than a determiner is grammatical as the subject of a sentence, where the verb agrees with that lexical item in person and number, then it is a noun. If it can appear as part of a possessive construction, again without any modifying elements other than a determiner and any appropriate possessive morphology, then it must be a noun.

The examples in (5.1) - (5.2) show these tests applied to various lexical items. In the first example, *kodo* 'frog' is shown to be a noun, as it can function as the subject of the clause as part of an NP with the determiner *pai*. The preposition *tuti* cannot do so; no change in the person or number markings on the determiner or verb will render this sentence grammatical. The second example shows *ai* 'tree' as the possessor of *vavo* 'top', both nouns. Note that both are directly followed by a determiner. This construction is not acceptable when the possessum is replaced by a verb *sipope* 'they fly'.[2]

(5.1) *Nouns as subjects*

a. ***Wanggar***=*pa-i k<i>opa.*
rat=DET-SG 3SG-jump
'The rat jumps.' [S80 IMK]

b. * ***Tuti***=*pa-i k<i>opa.*
With=DET-SG 3SG-jump
'The with jumps; the accompanying one jumps' [S80 IMK]

(5.2) *Nouns in possessive constructions*

2. This sentence actually does have a grammatical interpretation, in which *pa* is an aspectual marker and *sipope pa* is a separate clause rather than part of a possessive construction. In this interpretation, the sentence means 'The birds sit in the tree; they have flown.'

a. *Aya=pa-si si-maso na* ***ai****=pa-i* ***vavo****=pa.*
bird=DET-PL 3PL.NH-sit LOC tree=DET-SG top=DET
'The birds sit on the top of the tree.' [S20 IMK]

b. * *Aya=pa-si si-masoi na* ***ai****=pa-i* ***si-pope****=pa.*
bird=DET-PL 3PL.NH-sit LOC tree=DET-SG 3PL.NH-fly=DET
'The birds sit on the flight of the tree.' [S80 IMK]

If a form bears non-nominal morphology, such as verbal agreement prefixes (§4.4.1), applicative marking (§4.4.2), causative marking (§4.4.3), or a directional enclitic such as *=ma*[3] or *=ra* directly on the root, then it can be ruled out as a nominal root. Example (5.3) shows how the verb *roi* 'sing' can take verbal agreement, applicative, and directional morphology, but these are ungrammatical on the noun *ranu* 'song'. Unlike several other word classes, nouns can appear predicatively bearing verbal morphology only when the essive prefix *ve-* is also present, as in (5.4), reproduced from (4.68a).

(5.3) a. *Set-i-****roi****-ra.*
3PL.HUM-APPL-talk-to.there
'They use it to sing to (people) over there.'

b. * *Set-i-****ndanu****-ra.*
3PL.HUM-APPL-**song**-to.there
'They use it to song to over there.'

(5.4) a. *Simuti=pa-i* ***v<i>e-buo****.*
orange=DET-SG 3SG-ESS-fruit
'The orange (tree) is fruiting/has fruit.' [S60 BMW]

b. * *Simuti=pa-i* ***b<i>uo****.*
orange=DET-SG 3SG-fruit
'The orange (tree) is fruiting/has fruit.'

A form can change classes to become a noun after the addition of this non-nominal

3. The homophonous focus particle *=ma* does appear directly following the NP. These are likely historically derived from the same source but are not synchronically related.

morphology, however. Example (5.5) show the verbal root *ra* 'go' with the 3rd-person singular agreement prefix and a following determiner, this derives the noun 'road'. The same process is attested with the verb *unu* 'drink', deriving the noun 'a drink'.[4]

(5.5) a. ***R<i>a=pa-i*** *n<i>ai na wana.*
3SG-go=DET-SG 3SG-be.at LOC there
'The road is over there.' [S76 IMK]

b. *r<i>a*
3SG-go
'he goes'

Nouns expressing types of places - different from locational nouns, which are described below - can be derived by attaching the locative preposition *na* as a suffix to an agreement-bearing verb. These can then appear in the DP as would any other noun. Three examples are given in (5.6). *Raro* in the first sentence is a locational noun meaning 'inside; interior', though it translates most fluently in English to a preposition. The first example was produced by speaker BAK as part of a frog story; the second two examples occurred during elicitation with IMK. Note that while the subject agreement on the verb is always 3rd person, it can be singular or plural, resembling the construction used for passive-like constructions.[5]

(5.6) a. *Sandu sung-gubi-re raro=ne-i, raro...* ***seng-gubi-na****=ne-i.*
they.two 3DU-bathe-DUR inside=DET-SG inside... 3DU-bathe-LOC=DET-SG
'The two of them swam inside (the water), inside... the swimming place.' [S50 BAK]

b. ***p<i>ote-na****=ne-i*
3SG-go.fishing-LOC=DET-SG

4. Rendered as *minuman* in Malay, the contact language, where the nominalizing *-an* suffix renders word class more clear-cut.

5. See §4.4.6.

‘the fishing place; the place where one fishes’ [S61 IMK]

c. ***set-api-na***=*ne-i*
3PL.HUM-eat-LOC=DET-SG
‘the eating place; the place for eating’ [S61 IMK]

Locational nouns are slightly different from the rest of the noun class in their behavior in possessive constructions. (See §4.4.7 for a more detailed discussion of possession in Wamesa.) Only a limited group of nouns, including kinship terms and certain body parts, can be inalienably possessed. Possessive relationships which are unmarked for alienability - that is, those which can involve all nouns, whether or not they are eligible for the inalienable possessive construction - can be described using juxtaposition or by using the verb *ne* ‘have’. *Ne*-possession, the most common construction, is available to all sub-classes of nouns *except* for locationals. Juxtapositional possession, with the order $[_{DP}$ Possessor$][_{DP}$ Possessum$]$ is by far the less common of the two options, and occurs most often with locationals. The example in (5.7) includes both kinds of alienable (or more precisely, non-inalienable) possession. The verb *ne* ‘have’ is used to express the speaker’s possession of their garden, while *rawesi* ‘side’, a locational noun, is linked to *romi* ‘garden’ through juxtapositional possession.

(5.7) *Kambu pe-si n<i>ai na* ***i-ne*** ***romi=wa*** ***rawesi=wa***
water DET.INDEF-PL 3SG-be.at LOC 1SG-have garden=DET side=DET
‘There is a stream beside my garden.’ (Lit. ‘The waters are at my garden’s side.’) [S76 IMK]

5.3. Verbs

Verbs in Wamesa take a range of affixes not available to other word classes, including subject agreement markers (without an accompanying inalienable possessive suffix), causative prefixes, and directional enclitics.

(5.8) *Subject agreement:*

Yau ***i-mune*** *koro=wa-i.*
I 1sg-kill snake=det-sg

'I kill the snake.' [S77 IMK]

(5.9) *Causative prefix:*

Esi=pa-i ***di-o-mbavou*** *so n<i>e anio.*
one=det-sg 3sg-caus-go.home to 3sg-have house.

'The one had to go home to his house.' [S60 BMW]

(5.10) *Directional enclitics:*

Se-kopa=ra *sinitu=ne-i.*
3pl.hum-jump-to.there person-det-sg

'They jump over there, towards the person.' [S50 BAK]

As exemplified earlier in §4.4.1, some other types of roots - notably excluding nouns - can surface with verbal morphology. Wamesa is what Evans & Osada (2005: 365) call a "rampant zero conversion language" with high categorial lability, particularly when it comes to deriving verbs; forms from other classes frequently undergo zero-derivation to verbhood to be used as predicates. In this case, subject agreement is required and other verbal morphology may also apply. The examples from (4.53) are reproduced below as (5.11) to illustrate. Evidence that this is indeed verbal marking and not just marking on a predicate comes from a small subclass of adjectives which can be used predicatively but which cannot take verbal morphology; these are discussed further in §5.4.

(5.11) a. *Wona=pa-i* ***m<i>eta****.*
dog=det-sg 3sg-black

'The dog is black.' [S29 IMK]

b. *Mararea* ***se-mbe-rime****=e siri.*
child 3pl.hum-ess-five=and one

'There are six children; the children are six.' [S20 IMK]

c. *Ririo=wa-si* ***si-pau****.*
leech=det-3pl.nh 3pl.nh-many

‘The leeches are many; there are many leeches’ [S73 IMK]

d. *Yau* ***i-manau*** *y-unu kambu.*
I 1SG-COMPL 1SG-drink water
‘I already drank water; I’m finished drinking water’ [S16 IMK]

Verbs and nouns are the only word classes in Wamesa which may appear adjacent to another member of their own class within a clause (excluding list environments). With nouns, this occurs when one noun modifies the other, as in (5.12a), or in a possessive construction without an intervening determiner, as in (5.12b).

(5.12) a. ***Aya tomboro****=pa-i s<i>uvi na kamberei.*
bird owl=DET-SG 3SG-exit LOC hole
‘The owl emerges from the hole.’ [S7 IMK]

b. ***Maniai karaini****=pa-i n<i>ai na* ***awini madiawi****=pa-i*
bee nest=DET-SG 3SG-be.at LOC mother younger.sibling=DET-SG
n<i>e romi=wa-i.
3SG-have garden=DET-SG
‘The bees’ nest is in my mother’s younger sibling’s garden.’ [S76 IMK]

Serial verb constructions with two consecutive verbs are common in natural Wamesa speech. In elicitation, sequences of up to five consecutive verbs were tested and accepted, suggesting that in principle there is no upper limit to the number of instances of VP repetition which may be grammatical, other than that imposed by the memory capacity of listener and speaker. In these constructions, all of the verbs must refer to the same event, and all must agree with their subject. Subjects are shared between the verbs, but not other arguments. They are produced with no intonational breaks between the verbs, as would occur in non-serial constructions. The examples in (5.13) were produced by speakers IMK and BAK, respectively, during frog story narration; the examples in (5.14) were invented by me and confirmed as grammatical by IMK during elicitation. In some of these examples, the serial verbs are embedded under either *o* ‘want’ or *kamberei* ‘not want’, making the number of consecutive verbs larger than the number which can be said to be serial.

(5.13) a. *Mararea=ne-i di-ase ma re,* ***di-osa k<i>o-re n<i>e***
child=DET-SG 3SG-swim to land 3SG-stand 3SG-hold-DUR 3SG-have

sarera=ne-i.
fishing.pole=DET-SG

'The child swims to land, he stands up holding his fishing pole.' [S29 IMK]

b. ***T<i>au=wa t<i>pure-i.***
3SG-fall=down 3SG-hold-3SG

'He falls down holding him.' [S50 BAK]

(5.14) a. *Yau* ***i-kamberei i-ra y-ase i-sere*** *suomuse.*
I 1SG-not.want 1SG-go 1SG-swim 1SG-see shark.

'I don't want to go swim and see sharks.' [S39 IMK]

b. *Andi* ***di-o r<i>a p<i>ote n<i>unu di-a*** *dia=pa-i.*
He/she 3SG-want 3SG-go 3SG-go.fishing 3SG-cook 3SG-eat fish=DET-SG

'He wants to go and catch, cook, and eat the fish.' [S39 IMK]

5.4. Adjectives

Verbs and adjectives closely resemble one another in Wamesa. The major distinguishing factor is that adjectives can be used attributively as a bare stem, while verbs cannot. When modifying a noun within the DP, verbs must be affixed with the essive marker *ve-*, which derives an adjective or introduces a relative clause, as described in §4.4.4. Adjectives can take *ve-* in this position, but are grammatical without. There is a meaning difference between attributive adjectives with and without *ve-*, so the two cases are unlikely to be unifiable under a single relative clause analysis. The behavior of these two classes in this position is contrasted in (5.15) and (5.16).

(5.15) *Adjectival root:*

rebuki ***baba****=pa-si*
stone big=DET-PL

‘the big rocks’ [S14 IMK]

(5.16) *Verbal root:*

a. *mararea* ***ve-rina****=pa-sia*
child ESS-know=DET-3PL.HUM
‘the children who know’ [S80 IMK]

b. * *mararea* ***rina****=pa-sia*
child know=DET-3PL.HUM
‘the knowing children’ [S80 IMK]

Distributionally, attributive adjectives can appear in the NP directly preceded by the noun and followed by a determiner, as in the preceding examples. When used predicatively, adjectives behave exactly like stative verbs; it could be argued that in this position they have undergone zero-derivation to become verbs. Here they follow the DP subject, agree with it in person, number, and, where appropriate, animacy, and can take other verbal morphology such as the applicative prefix *it-*. This is not simply an *aktionsart* distinction; other non-adjectival stative verbs, such as *sanepaya* ‘like, be happy’, pattern with regular verbs, not with adjectives.

The examples in (5.17) demonstrate this behavior. In (5.17a) the adjective *baba* ‘big’ is attributive, appearing directly after the noun and without person and number marking. In (5.17b) the same adjective is used predicatively; it agrees with the subject *mararea* ‘child’ in person and number, just as the verb *tau* ‘fall’ does in the sentence before.

(5.17) a. *Mararea* ***baba****=pa-i t<i>au=wa.*
child big=DET-SG 3SG-fall=down.
‘The big child falls’

b. *Mararea=pa-i* ***b<i>ba****.*
child=DET-SG 3SG-big.
‘The child is big.’

As with verbs, several predicative adjectives can appear consecutively, shown in (5.18). This is not true, however, of bare-stem attributive adjectives. Only one bare attributive adjective may appear per DP; any additional adjectives must be marked with *ve-* to create a relative clause, as in (5.19).

(5.18) *I-newe bolpen* ***k<i>ke b<i>ba k<i>rakuai.***
1SG-have pen 3SG-green 3SG-big 3SG-strong
'My pen is big, green, and strong.' [S4 IMK]

(5.19) a. *I-ne wona* ***pimasa ve-meta****=wa-i.*
1SG-have dog big ESS-black=DET-SG
'I have a big black dog.' [S76 IMK]

b. * *I-ne wona* ***baba kariria****=wa-i.*
1SG-have dog big evil=DET-SG
'I have a big mean dog.' [S76 IMK]

There is a small subclass of 'true' adjectives whose roots cannot directly bear subject agreement affixes. When used predicatively, these fall into two additional categories. The first group, including color terms such as *nukuai* 'yellow', *avu* 'grey, brown' and *kumuar* 'black, dark', take *ve-* plus subject agreement, as in (5.20). The second group, which has only two members so far attested, *katu* 'small' and *pimasa* 'big', take no agreement morphology at all. They constitute one of the very few instances in the language where the predicate does not include subject agreement, and the prime evidence that subject agreement is a property of the verb head itself and not simply of the predicate construction.

(5.20) a. *Dia karu=pa-i=ma* ***v<i>e-nukuai.***
fish water=DET-SG=TOP 3SG-ESS-yellow
'The fish is yellow' [S12 IMK]

b. * *Dia karu=pa-i=ma* ***n<i>ukuai.***
fish water=DET-SG=TOP 3SG-yellow
'The fish is yellow'

(5.21) a. *Wonggei=pa-si* ***pimasa***.
cassowary=DET-PL big
'The cassowaries are big.'

b. * *Wonggei=pa-si* ***si-pimasa***.
cassowary=DET-PL 3PL.NH-big
'The cassowaries are big.'

5.5. Prepositions

Prepositions in Wamesa can be distinguished on functional and distributional grounds. Functionally, prepositions are a closed class of words regulating a relationship between syntactic constituents, often but not necessarily DPs. Prepositions in Wamesa include, for example, *so* 'to, for', *tuti* 'with', *na* 'on, at, from', and *to* 'until'. These linearly precede their objects; only one possible postposition, *kasau* 'between' is attested in my data. The complete set of attested prepositions is listed in (5.22).

(5.22) *Prepositions*

Form	Meanings	Objects
kasau	between	common nouns
maso	towards	common nouns
na(*na*)	at, on, from	common nouns
so	to, for	common nouns
tuti	with, instrumental	common nouns
to	until	common nouns
tua	past	common nouns
susa	down into	common nouns
warire	around	common nouns
do	to	geographic nouns
ma	to here	geographic nouns
ra	to there	geographic nouns
re	at	geographic nouns

Examples of various uses of the most common prepositions are given in (5.23) - (5.25). The first of these prepositions, *so*, is usually glossed as 'to', but performs a wider range of functions than just denoting a geographic destination. (5.23a) gives an example of this basic usage, where the object of *so* is endpoint of a physical journey. The sentence in (5.23b) has a somewhat more abstract usage, where the object of the preposition is the recipient of a greeting. In (5.23c) the path is even less concrete; *so* here describes the transfer of the property of being green to the fingernails. Example (5.23d) demonstrates *so*'s benefactive function, with the object of the preposition acting as the person for whose benefit the action of the VP is carried out.

(5.23) *So: 'to, for'*

a. *Yau y-apai=re* ***so*** *Bintuni.*
I 1SG-run=DUR to Bintuni
'I run to Bintuni.'

b. *Yau y-oyo "diru vesie"* ***so*** *Sutri.*
I 1SG-say night good to Sutri.
'I say "good night" to Sutri.' [S10 IMK]

c. *Set-it-one=pa-i ve-kake* ***so*** *se-mbara-kiai-dire=pa-si*
3PL.HUM-APPL-give=DET-SG ESS-green to 3PL.HUM-hand-digit-nail=DET-PL
sanope "kutek".
name nail.polish
'The thing they use to make their fingernails green is called "nail polish".' [S71 IMK]

d. *V<i>ori serei* ***so*** *vinie=pa-i.*
3SG-buy cloth for wife=DET-3SG
'He buys cloth for his wife.' [S70 IMK]

Tuti denotes a comitative or instrumental relationship. In addition to its most common usage as 'with' it also can carry out a function which is best translated into English as 'and', as shown in (5.24b). The sentence in (5.24c) shows its instrumental usage.

(5.24) *Tuti: 'with'*

a. *Su-mai* ***tuti*** *sinitu vata=ne-i.*
3DU-play with person good=DET-SG
'They play with the good boy.' [S50 BAK]

b. *Sinitu vata=ne-i n<i>e kodo=ne-i* ***tuti*** *wona=ne-i.*
person good=DET-SG 3SG-have frog=DET-SG with dog=DET-SG
'The good boy has a frog and a dog.' [S50 BAK]

c. *Yau y-o y-ane anambet* ***tuti*** *kai.*
I 1SG-want 1SG-eat cold.sago with utensil
'I want to eat cold sago using *kai*.' [S41 IMK]

The preposition *na* has a wide range of locative uses. It can mean simply ‘at’, as in (5.25a) and (5.25b). It can also means specifically ‘from’, denoting movement or place of origin, as in (5.25c). This sense of ‘from’ can be used in statements such as ‘I am from America’ in addition to sentences such as the one given below which entail more immediate movement of the subject. *Na* can also have an instrumental meaning, as in (5.25d). This final function is equivalent to that of *tuti* in (5.24c) above, and is far less common than the others.

(5.25) *Na*

a. *Wona=pa-i m<i>so* ***na*** *n<i>e aipata=pa-i.*
dog=DET-SG 3SG-sit on 3SG-have bed=DET-SG
‘The dog sits on his bed.’ [S7 IMK]

b. *Yau i-nai* ***na*** *Bintuni.*
I 1SG-be.at at Bintuni.
‘I’m in Bintuni.’ [S68 IMK]

c. *Yau i-kopa* ***na*** *nina vera yana.*
I 1SG-jump from here to there
‘I jump from here to there.’

d. *Mararea=wa-sia se-so aya=wa-i* ***na*** *rebuki.*
child=DET-3PL.HUM 3PL.HUM-throw.at bird=DET-SG with stone
‘The children pelt the bird with stones.’

5.6. Other Spatial Constructions

When describing spatial relationships, a distinction must be made in Wamesa between locational nouns, directional adverbs, post-verbal particles, relators, and prepositions. All of these word types express location, but all have slightly different grammatical properties.

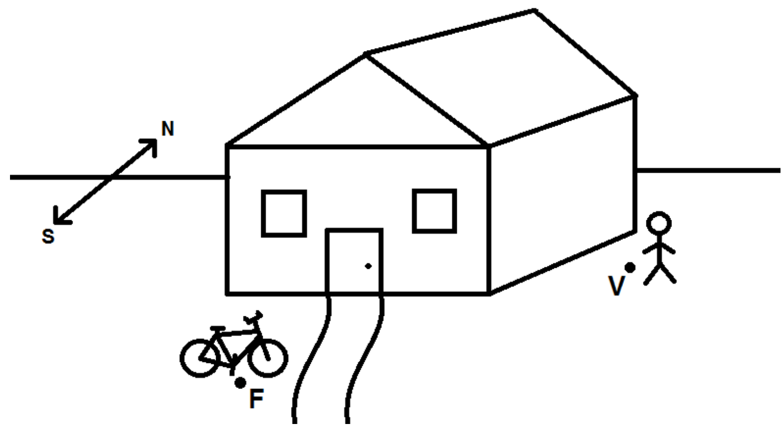


Figure 5.1: Frames of Spatial Reference. N: North; S: South; F: Figure, V: Viewpoint.

5.6.1. Frames of Reference

One property which distinguishes the various categories of locational items is the reference system in which they participate. Levinson (1996a,b) lays out three distinct frames of reference (defined below) which languages may use to describe spatial relations: intrinsic, relative, and absolute. Wamesa makes use of all three of these systems, with particular emphasis on the intrinsic and absolute orientations. The following section is based on Levinson's work and its application by van den Heuvel (2006) and Ross (2007) to related languages.

Figure 5.1, adapted from van den Heuvel (2006), will be used here to illustrate these three spatial reference systems. First, though, it is necessary to define some basic terminology. Take for example the sentence in (5.26).

(5.26) The bicycle is south of the house.

There are two main arguments in this sentence: 'the bicycle', the *Figure*, and 'the house', the *Relatum*. The Figure is that thing which the sentence is locating in space (the NP which the following PP modifies), while the Relatum is that entity in relation to which the Figure is located (the NP object of the preposition). The *Origo* is the center point or origin of the directional system, in this case the house, from which the position of the Figure is reckoned.

In an *intrinsic* reference system, the position of the Figure is described in relation to some part of the Relatum which remains the same regardless of orientation, speaker location, and other factors. Here the Origo is the volumic center of the Relatum (van den Heuvel 2006). In an intrinsic system, the scenario in Figure 5.1 can be described as 'The bicycle is in front of the house' even when this is not so from the viewpoint of the speaker (standing at point V) because (in English) the side of the house facing the street and with the main door is considered to be inherently the front. As Levinson (1996b) points out, what constitutes the front, back, side, or even top or bottom of an object[6] is often not inherent at all but culturally determined, and not universal. (It is hard, for example, to imagine in English what the inherent front of a tree might be, but in Chamus one exists (Levinson 1996b).)

A *relative* directional system gives the location of the Figure in relation to the Relatum based on the location of an external *Viewpoint*, often that of the speaker, which also serves as the Origo. To say that 'the bicycle is in front of the house' in Figure 5.1 would be untrue in a relative framework, given the Viewpoint at point V. With the speaker in that location, a more accurate statement would be 'the bicycle is beside the house' or 'the bicycle is to the left of the house'. Were the Viewpoint to move to the base of the front path, the statement 'the bicycle is in front of the house' would be true in both an intrinsic and a relative system.

An *absolute* reference framework uses fixed coordinates, such as the cardinal directions in English, to describe the position of the Figure. In this case, the statement 'the

6. The top of a box of cereal is still its top even if the box happens to be lying on its side.

bicycle is south of the house' would be true in an absolute framework. This statement retains its truth value regardless of the orientation of the house and the position of the Viewpoint/speaker.

5.6.2. Locational Nouns

Outside of a particular subset to be discussed later, most locational nouns in Wamesa work within the intrinsic framework. The relational noun describes a part of a reference object (its top, its interior) or a space in relation to that reference object (below, above). In order to express the location of a second object in relation to the reference object, the locational noun must co-occur with a preposition, most often locative *na(na)* 'on, at, from', or *so* 'to', as in (5.28). Locational nouns most often appear in possessive constructions ('the table's top'), in which the locational noun is the possessum and the reference noun is the possessor, as in (5.27a) and (5.27b). Locational nouns can also appear unpossessed, as in (5.28), where it is not explicitly stated what the dog has fallen down to or from.[7] Locational nouns are less likely than other common nouns to co-occur with a determiner, and when they do that determiner is unmarked for number.

(5.27) a. *meja=pa-i* ***diu****=wa*
table=DET below=DET
'below the table'

b. *ai=pa-i* ***vavo****=pa*
tree=DET-SG top=DET
'on top of the tree'

(5.28) *Wona=pa-i t<i>au=wa so* ***diu****=wa.*
dog=DET-SG 3SG-fall=down to below=DET
'The dog falls down (to below).' [S7 IMK]

7. In this case, the dog fell out of a window to the ground below.

Prepositions can be distinguished from locational nouns by their position relative to the head noun. In simplest terms, a preposition precedes its object, while a locational noun will directly follow the DP possessor (if one exists) and will usually be directly followed itself by a determiner. Two prepositions may not appear in a row, but a preposition may directly precede a locational noun.

(5.29) a. *Si-maso* ***na vavo***.
3PL.NH-sit LOC top
'They sit on top.' [S20 IMK]

b. * *Si-maso* ***na tuti*** *mararea mua=pa-i.*
3PL.NH-sit LOC with child male=DET-SG
'They sit with the boy.'

There is a subclass of geographic locational nouns which are never possessed. These nouns refer to geographic locations, and function within an absolute directional framework, similarly to compass points in English. Rather than referring to universal directions such as north and south, these terms apply to the specific landscape in which Wamesa society exists, and refer to direction/location relative to land, sea, and elevation, with a label for each salient zone. Given the landscape, 'seawards' means downhill/out of the mountains and 'landwards' means uphill/into the mountains; this is reflected in the terminology. The standard village layout is also integral to the system: *po* 'in front' specifically means the area of the village between the two rows of houses, which face each other; *pui* 'behind' refers to the wilder, less cultivated landscape behind the houses, and by extension to any other towns or villages along the coast of Cenderawasih Bay (though not those farther inland). Less obviously absolute is *ri* 'outside (the salient area)'. *Ira do ri wa* 'I go outside' can, depending on context, mean that the speaker is leaving a house or leaving the Wamesa area, to a relatively distant destination such as Sorong or Jakarta. Unlike English compass points, which designate a single direction, terms which function within the absolute directional framework in Wamesa can encompass a wider range of actual trajectories: *rau* 'seawards' is generally eastwards, since Cenderawasih Bay is to the east, and *re* 'landwards', *wi* 'uphill, towards the interior', and *ye* 'upriver, inland uphill' are generally westwards, but these can encompass nearly a 180-degree directional range. *pui* 'behind' and *po* 'in front' can mean either of two directions within a given village, though which two directions those are will depend on the orientation of that village, and which is meant in a particular utterance depends on where in the village the one being located originates, and therefore which direction is followed to end up in the fringes of the village (*pui*) or the central area (*po*). When used to describe longer trips, *pui* can mean either north or south along the coast. *Ri* encompasses trips in any direction at all, so long as sufficient distance is traveled to exit the relevant zone. All of these, however, are defined in terms of features of the landscape rather than as parts of a Relatum or in terms of a variable Viewpoint.

In addition to their locational functions, some of these terms also function as common nouns. *Wi* also means 'mountain'; *rau* 'sea', and *re* 'land, shore'. *Po* has broadened its meaning to a more general 'in front', with a relative orientation, as in (5.30f).

These nouns do not occur as objects of the usual prepositions *so* and *na*, but rather a separate set specific to this subclass of nouns. The combination of preposition + geographic noun acts adverbially, and can be made more precise with the addition of a PP specifying the destination, as in (5.30b). They are also only attested in my data with the distal determiner *wa*, never with *ne* or *pa*, which express location closer to the speaker. Geographic locational nouns can be reduplicated to denote an increased distance in the direction indicated, as in *do rau* 'seawards' versus *do raurau* 'far out to sea'.[8]

8. *Karduse* 'box' in (5.30f) is a loan from Malay; see §2.4.5 for the adaptation and retention of consonant clusters in borrowed words.

(5.30) a. *I-ra* ***do rau****=wa.*
1SG-go to ocean=DET
'I go seawards.' [S75 IMK]

b. *I-ra* ***do rau*** *so nu Mansinam.*
1SG-go to ocean to island Mansinam
'I go seawards to Mansinam Island.' [S64 IMK]

c. *I-ra* ***do re****=wa.*
1SG-go to land=DET
'I go landwards.' [S66 IMK]

d. *I-ra* ***ra*** *ye.*
1SG-go to.there inland.uphill=DET
'I go into the mountains.' [S41 IMK]

e. *I-maso-i* ***re po****.*
1SG-sit-3SG at front
'I sit out front.'

f. *Buku=pa-i n<i>ai na* ***do po*** *karduse=pa-i.*
book=DET-SG 3SG-be.at at to front box=DET-SG
'The book is in front of the box.' [S12 IMK]

5.6.3. Locational Relators

Ross (2007) defines a locational relator as 'a preposition-like morpheme which differs in its distribution from a preposition in that it precedes either a prepositional phrase or a local noun' (268). He identifies these as occuring in the Oceanic languages of New Britain and New Ireland, and in Longgu, spoken on Guadalcanal in the Solomon Islands. In these languages, the ablative and allative relators are derived from the Proto-Oceanic verbs **mai* 'come' and **ua* 'go towards hearer' (Ross 2003). The example sentences in (5.31),[9] from

9. *Additional abbreviations from Ross:*
IRR irrealis
O: object pronominal enclitic or suffix
PREP preposition
R relator
S: subject pronominal enclitic or suffix

Longgu, show the relator *vu*, from POc **ua*, in pre-nominal and pre-prepositional position.

(5.31) a. *...m-ara la maʔa* ***vu*** *masuʔu.*
and-s:3pl go perfective r bush
'...and they went into the bush.'

b. *amalu ho la* ***vu*** *ta-na malaba-i ni um^{w}ani-a.*
d:1pl.excl irr go r prep-p:3sg garden-sg in.order.to weed-o:3sg
'We will go into the garden to weed it.' (Ross 2007: 268)

On the opposite end of New Guinea, a similar lexeme is found in Wamesa: *vera* 'to, towards', decomposable into the essive prefix *ve-* and the directional particle *ra* 'to there; movement away from speaker'. Though its etymology is different from the Oceanic relators - Wamesa *ra* is cognate with POc **la* 'go (to), go away from speaker' rather than **ua* 'go towards addressee' - its function and distribution closely mirrors that found in the above examples. As with Longgu *vu*, *vera* can appear directly preceding either a noun, as in (5.32a) and (5.32b), or a preposition, as in (5.32c). Its meaning is roughly equivalent to that of *so* but somewhat less precise. Use of *vera* does not entail that the destination was reached, only that movement occurred towards that direction. *Vera* and *so* may appear together, as in (5.32c).

(5.32) a. *Nini=ma s<i>e* ***vera*** *wana.*
this=top 3sg-see toward over.there
'This one looks over there.' [S20 IMK]

b. *Set-apai* ***vera*** *kambu.*
3pl-run toward water
'They run towards the water.' [S29 IMK]

c. *Yau i-ra nai na Amerika* ***vera*** *so Indonesia.*
I 1sg-go be.at at America towards to Indonesia.
'I go from America to Indonesia.'

5.6.4. Directional Adverbs

Only three freestanding directional adverbs are attested in my corpus, along with three adverbial enclitics which indicate the direction or distance of motion on a verb. The free-standing adverbs are *mesu* 'downwards', *maye* 'upwards', and *rawa* 'downwards; over there'. The second of these appears to be decomposable into=*ma* 'movement towards the speaker' plus the geographic noun *ye* 'inland uphill', and third into *ra* 'movement away from speaker' plus =*wa* 'downwards, into'. In the case of *maye* this involves a broadening of its specific geographic use described above.

(5.33) a. *Padamara=ne-i v<i>viare* ***mesu*** *so meja=ne-i.*
lamp=DET-SG 3SG-shine downwards to table=DET-SG
'The lamp shines down onto the table.' [S30 IMK]

b. *Koro=wa-i s<i>uve naina kamberei=wa-i* ***maye*** *k<i>ripe yau.*
snake=DET-SG 3SG-exit LOC hole-DET-SG upwards 3SG-bite I
'The snake came up out of the hole and bit me.' [S77 IMK]

c. *Su-tau* ***rawa*** *so kambu.*
3DU-fall downwards to water
'They fell down into the water.' [S7 IMK]

d. *Kodo=pa-i m<i>so* ***rawa*** *nina, ama ai=ne-i vavo=ne-i.*
frog=DET-SG 3SG-sit over.there this, umm tree=DET-SG top=DET-SG
'The frog sits over there, umm, on top of the log.' [S50 BAK]

The enclitics are =*ra* 'movement away from speaker', =*ma* 'movement towards speaker', and =*wa* 'movement down or into'. These forms attach to the verbs of motion, and were discussed in §4.3.4.

(5.34) a. *Setoru se-mbavou=**ra**.*
they.three 3PL.HUM-go.home=to.there
'The three of them go home (to a place away from the speaker).' [S50 BAK]

b. *Su-mbavu=**ma**.*
3DU-go.home=to.here
'They come home (to a place near the speaker).'

c. *Su-tau rawa so diu=wa. Mapa=pa-i su-tau mesu:*
3DU-fall downwards to below=DET valley=DET-SG 3DU-fall downwards
*wona=pa-i t<i>au=**wa**, mararea=pa-i t<i>au=**wa**.*
dog=DET-SG 3SG-fall=down child=DET-SG 3SG-fall=down
'They fall down to below. They fall down into the valley: the dog falls, the boy falls.' [S7 IMK]

5.6.5. Locational Deictics

Finally, Wamesa has three locational deictics. These forms are distinguished by the extent of the distance they denote between the speaker and the object of the deixis. This three-way split corresponds directly to that found in the determiners (§4.3.1), and it is likely that there is some historical connection between the determiner and deictic forms, as well as with the demonstratives, discussed below in §5.7.2.

(5.35) *nina* 'here'
yana 'there (middle-distance)'
wana 'there (far distance)'

(5.36) a. *Se-mbori aibuo **nina**.*
3PL.HUM-buy fruit here
'They buy fruit here.' [S60 BMW]

b. *Varo sinitu ve-maso na **yana**.*
not.exist person ESS-sit LOC there
'There's no one sitting there.' [S20 IMK]

c. *Nini=ma s<i>e vera **wana**.*
this=TOP 3SG-see towards there
'This one looks over there.' [S20 IMK]

5.7. Additional Categories

5.7.1. Numerals

Numerals form their own class in Wamesa, appearing following the determiner in the DP, as discussed in §4.3.2. They share a position with other quantifiers in the linear word order, but their behavior when used predicatively is different. Where non-universal quantifiers can bear subject agreement directly on the root, numerals require the essive prefix *ve-* to be used predicatively. The examples in (5.37) illustrate this contrast. The universal quantifier *vura* 'all' cannot be used predicatively at all.

(5.37) a. *Ririo=wa-si* ***si**-pau.*
leech=DET-3PL.NH 3PL.NH-many
'The leeches are many; there are many leeches' [S73 IMK]

b. *Se-pote dia* ***se**-mbe-at.*
3PL.HUM-go.fishing fish 3PL.HUM-ESS-four
'They caught four fish; they caught the fish they are four.' [S20 IMK]

Wamesa has an additive number system, which is quinary-decimal (Comrie 2013) at least as high as twenty. Under Hammarström's (2010: 15) definition,[10] five and ten are bases in the Wamesa system. Discrete terms exist for numbers one through five; six through nine are expressed by addition, using *=e* 'and'. 'Six', for example, is *rime=e siri* 'five and one'. Ten is again discrete, and eleven through 19 are again additive. It is difficult to reliably elicit numerals higher than 10; speakers either don't know or disagree with each other. IMK, for example, used *utin* to mean 'twenty', while TLB and BWM used it for 'hundred'. Even when speaking Wamesa in the village, speakers used Malay numerals

10. "The number *n* is a base iff:
1. the next higher base (or the end of the normed expressions) is a multiple of *n*;
and
2. a proper majority of the expressions for numbers between *n* and the next higher base are formed by (a single) addition or subtraction of *n* or a multiple of *n* with expressions for numbers smaller than *n*."

exclusively when referring to numbers over ten. Wamesa numbers through 19 are given in (5.38).

(5.38) *Wamesa Numerals*

One:	*siri* 'one'	Eleven:	*sura=e siri* 'ten=and one'
Two:	*muandu* 'two'	Twelve:	*sura=e muandu* 'ten=and two'
Three:	*tiga* 'three'	Thirteen:	*sura=e toru* 'ten=and three'
Four:	*at* 'four'	Fourteen:	*sura=e at* 'ten=and four'
Five:	*rime* 'five'	Fifteen:	*sura=e rime* 'ten=and five'
Six:	*rime=e siri* 'five=and one'	Sixteen:	*sura=e rime=e siri* 'ten=and five=and one'
Seven:	*rime=e muandu* 'five=and two'	Seventeen:	*sura=e rime=e muandu* 'ten=and five=and two'
Eight:	*rime=e toru* 'five=and three'	Eighteen:	*sura=e rime=e toru* 'ten=and five=and three'
Nine:	*rime=e at* 'five=and four'	Nineteen:	*sura=e rime=e at* 'ten=and five=and four'
Ten:	*sura* 'ten'		

5.7.2. Determiners and Demonstratives

The definite determiner clitics are discussed in detail in §4.3.1. In addition to these, Wamesa has one indefinite determiner, *pe*. This form is not a clitic and does not trigger stress shift, but otherwise behaves in the same way as the definite determiners, appearing in the head of the DP and bearing the number affixes for the phrase.

(5.39) a. *Vavi ambe* ***pe****-i t<i>au=wa.*
woman foreign DET-SG 3SG-fall-down.
'A foreign woman falls down (off of her bicycle).' [S60 BMW]

b. *Raria* ***pe****-si i-vori rau, raria* ***pe****-si i-vori.*
day DET-PL 1SG-buy leaf, day DET-PL 1SG-buy NEG
'Some days I buy vegetables, some days I don't.' [S31 IMK]

As with the definite determiners and locational deictics, the Wamesa demonstratives show a three-way distance-based split. The table (5.40) gives the corresponding forms for these three categories.

(5.40)

	Determiner	Deictic	Demonstrative
'Here'	*=ne*	*nina*	*nini*
'There'	*=pai*	*yana*	*yani*
'There (far)'	*=wa*	*wana*	*wani*

Distributionally, however, the demonstratives resemble determiners but are not identical to them. Dryer (1992) cites Welsh and the Austronesian language Dehu as examples of other languages in which determiners and demonstratives cannot be considered part of a single class. Wamesa DPs including a demonstrative are far less likely than those without to have an overt determiner, but the two categories do sometimes co-occur as in (5.41a), suggesting that they do not occupy the same structural position. Demonstratives can also co-occur with adjectives as in (5.41b). Demonstratives surface following nouns and adjectives but preceding determiners.

(5.41) a. *S<i>e maso aya* ***nini****=pa-i.*
3SG-see towards bird this=DET-SG
'He looks towards this bird.' [S7 IMK]

b. *Rabo baba* ***nini*** *v<i>e-buo.*
log big this 3SG-ESS-fruit.
'This big tree is fruiting.' [S30 IMK]

In subject position, particularly when not modifying an overt noun, the demonstratives are very often marked with the topic particle *=ma*, especially in equational constructions like that in (5.42).

(5.42) ***Nini****=ma niau=pa-i, m<i>so meja=ne-i vava=ne-i.*
this=TOP cat=DET-SG, 3SG-sit table=DET-SG under=DET-SG
'This is a cat, sitting under the table.' [S30 IMK]

5.7.3. Pronouns

Wamesa pronouns, like the subject agreement prefixes, distinguish inclusive and exclusive forms for the 1st-person non-singular, human versus non-human in the 3rd-person plural, and singular, dual, and plural number for all persons. The full pronoun paradigm is given in (5.43), reproduced from (4.45).

(5.43) *Wamesa Pronouns*

Person	Singular	Dual	Trial	Plural
1incl		*tandu*	*tatoru*	*tata=*
1excl	*yau*	*nandu*	*amatoru*	*ama=*
2nd	*au*	*mandu*	*metoru*	*mia=*
3rd	*andi* (subj/obj) *i* (obj)	*sandu*	*setoru*	*sia=*

The third-person singular pronoun has two forms: *andi*, which may be used in subject or object position, and *i*, which appears in object position only. Both are independent lexical words, and contrast with the object suffix *-i* (§4.4.6); the use of *andi* in object position is more emphatic and implies a human antecedent, while *i* can be used with any level of animacy, encompassing all three of 'he/she/it'. The 3pl pronoun *sia=* is likewise used only with a human antecedent; non-human subjects must be referred to using a full noun, and objects with a noun or a determiner. The trial pronouns are rarely used and transparently composed of the plural verbal agreement prefix plus the numeral *toru* 'three', equivalent to a construction such as *we three* in English. The same can be done with larger numerals, though this never occurs in natural speech in my corpus. The plural pronouns, which are proclitics, are discussed in more detail in §4.3.11.

Wamesa is a pro-drop language, so use of overt pronouns is optional, and they are often omitted. Pronouns resemble their full nominal counterparts in that they can fulfil the same set of semantic and syntactic roles. However, they permit far fewer modifiers. Pronouns can co-occur with numerals, quantifiers such as *vura* 'all', and the modifier *kota* 'also', but never appear in my corpus modified by determiners, adjectives, or relative clauses.

5.7.4. Adverbs

Wamesa has a limited number of manner adverbs. The only three which are attested in my data are *saira* 'quickly', *nanaria* 'slowly', and *rusara* 'repeatedly'. A fourth adverbial phrase, *tuti vesie* 'well' is composed of the preposition *tuti* 'with' plus the adjective *vesie* 'good'. No other compositional forms in this pattern are attested. Manner information is rarely used in Wamesa discourse, and when it is it must be encoded periphrastically.

(5.44) a. *Yau i-kavio* ***nanaria.***
I 1SG-speak slowly.
'I speak slowly.' [S18 IMK]

b. *Yau i-vori rau* ***saira.***
I 1SG-buy leaf quickly.
'I buy vegetables quickly.' [S31 IMK]

c. *K<i>vio* ***rusara.***
3SG-speak repeatedly
'He speaks again and again.' (Used to describe a dog's repeated barking in a frog story.) [S7 IMK]

d. *Andi r<i>oi* ***tuti vesie.***
he/she 3SG-sing with good
'She sings well.'

There are three other adverbs which appear frequently in my corpus: *kota* 'also', *venasia* 'only', and *ariri* 'again'.

(5.45) a. *Rusa=ne-i* ***kota*** *di-pai.*
deer=DET-SG also 3SG-run
'The deer also runs.' [S7 IMK]

b. *Se-mbe-toru* ***venasia.***
3PL.HUM-ESS-three only
'They are only three; there are only three of them.' [S20 IMK]

c. *I-topa* ***ariri.***
1SG-narrate again
'I'll tell it again.'[11] [S50 BAK]

The majority of Wamesa adverbs are temporal. Because tense and aspect are not in most cases marked directly on the verb, that information often comes from an adverb, or from temporal nouns such as *kausapa* 'tomorrow' and *ravinie* 'yesterday' used adverbially. Some examples are given below. These usually appear clause-finally in natural speach, but can also appear clause-initially or directly preceding the verb.

(5.46) a. *Yau i-vori kavaru maki=pa-i* ***varani.***
I 1SG-buy bean mung=DET-SG not.yet
'I haven't bought the mung beans yet.' [S5 IMK]

11. This is the formulaic opening line used in storytelling. The word for 'story', *setopa*, is derived from the verb *topa* used here and literally means 'they narrate'.

b. *Yau i-vori kavaru maki=pa-i* ***tomanau***.
I 1SG-buy bean mung=DET-SG already
'I already bought the mung beans.' [S5 IMK]

c. ***Kausapa nani*** *yau i-vori kavaru maki=pa-i.*
tomorrow later I 1SG-buy bean mung=DET-SG
'Later tomorrow I'll buy the mung beans.' [S5 IMK]

5.7.5. Interrogatives

Polar questions are formed with the clause-final question clitics *=e* and *=te*. Using *=te* makes no assumptions about the expected answer, while *=e* presupposes that the statement from which the question is formed if true, as in tag questions using 'right?' or 'didn't he?' in English, or *ka?* in PM. The particles *=e* and *=te* are placed invariably clause-finally.[12]

(5.47) a. *Sasu sama=pa-i v<i>e-mahal=**te**?*
clothing buttocks=DET-SG 3SG-ESS-expensive=Q
'Are the pants expensive?' [S29 IMK]

b. *Au sane-mu=pa-i p<i>ota=**e**?*
You stomach-POSS 3SG-sick=Q
'Your stomach hurts, right?' [S29 IMK]

The two question markers are associated with very different intonational contours. *=Te* is characterized by a falling contour on the previous word followed by a high pitch, shown in Figure 5.2. *=E*, on the other hand, is characterized by a reletively high preceding pitch, followed by a marked fall, as in Figure 5.3.

Content WH-words appear *in situ* in Wamesa. The table in (5.48) gives a list of the Wamesa interrogatives and their English equivalents.

12. My corpus does not include examples of the relative ordering of these question markers and the other clause-final clitic, negative *=va*, though I would predict that *=va* would precede *=e* and *=te*.

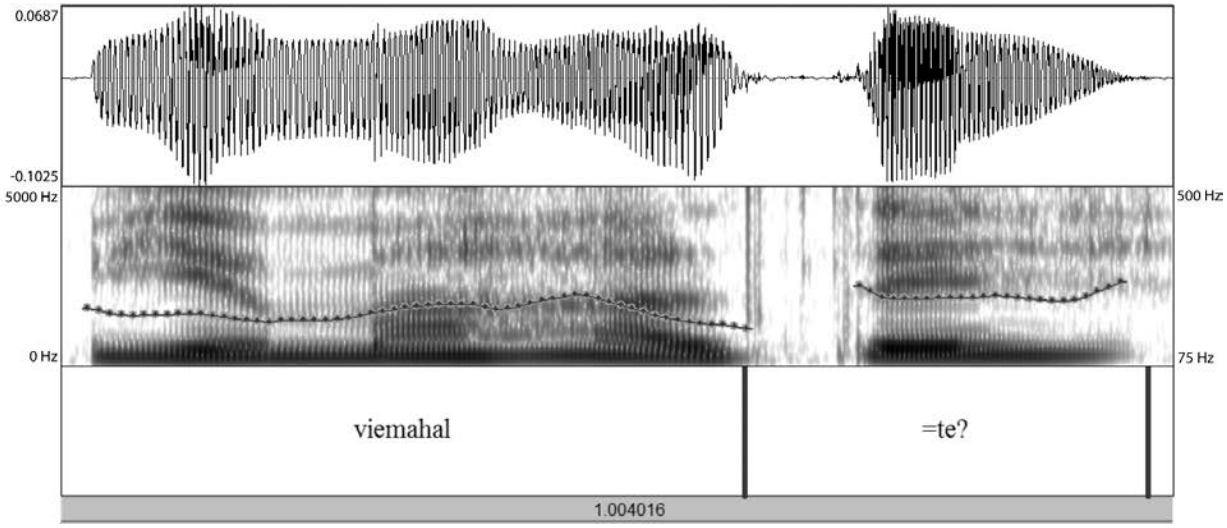


Figure 5.2: High Pitch on *=te*

(5.48) *Interrogatives*

tei	'who'
toni	'which'
vitoi	'what, how'
toi	'where, how many'
toine	'how many'
topina	'where to'
otopi	'why'

There appears to be a *wh* morpheme *to* which forms the base for most of these words. Some of the forms are morphologically transparent; *topina* 'where to' could plausibly be broken down into /to-pi-na/, literally 'Q-thing-LOC'. These forms have the same distribution as the content words for which they stand in: *tei* 'who' and *vitoi* 'what' can be followed by a determiner, as in (5.49b); *toine* 'how many' in (5.49f) takes the same morphology as a numeral; and so forth.

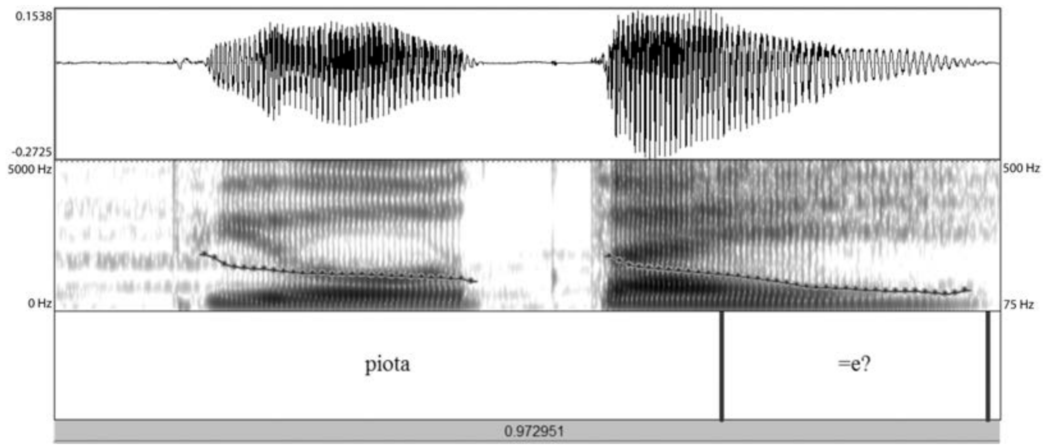


Figure 5.3: Falling Pitch on *=e*

(5.49) a. *Au v<u>ori vitoi?*
you 2sg-buy what
'What did you buy?' [S13 IMK]

b. ***Tei**=ne-i v<i>ui buku=pa-i?*
who=det-sg 3sg-write book=det-sg
'Who wrote the book?' [S13 IMK]

c. *Sasu sama **toni** v<i>e-mahal?*
clothing buttocks which 3sg-ess-expensive
'Which pants are expensive?' [S13 IMK]

d. *N<u>ai na toi=ne-i?*
2sg-be.at loc where=det-sg
'Where do you live?' [S13 IMK]

e. *Au=ma bu-o r<u>a **topina**?*
you=top 2sg-want 2sg-go to.where
'Where do you want to go?' [S13 IMK]

f. *Nomu mararea=pa se-mbe-**toine**?*
2sg.have child=det 3pl.hum-ess-how.many
'How many children do you have?' [S13 IMK]

Otopi, 'why', appears to be a compound composed of the verb *o* 'want' plus *to* 'Q' and *pi* 'thing'; a literal translation might be 'want what thing' or 'want what reason'. Accordingly, it recieves verbal subject agreement, as would the plain verb *o* 'want'.

(5.50) a. *Andi* ***di-otopi*** *r<i>a* *vera* *sasi rau?*
he/she 3SG-why 3SG-go to beach
'Why does she go to the beach?' [S13 IMK]

b. ***Set-otopi*** *se-mbori* *buku=pa-i?*
3PL.HUM-why 3PL.HUM-buy book=DET-SG
'Why do they buy the book?' [S13 IMK]

Chapter 6

Verbal Infixation: Synchronic & Diachronic

6.1. Introduction

In describing a linguistic phenomenon, it is important to be able to model the competence of an individual speaker, as for example Chapter 3 does for stress assignment. Speakers do not have access to the history of their language - in Lass' (1984:178) words, "a segment does not know where it came from" - and therefore their individual mental grammars must be able to produce the attested forms of that language on purely synchronic terms, without recourse to knowledge of prior states of the language. However, in some cases, accounting for how a pattern arose diachronically may provide a richer understanding of the phenomenon, and more importantly a motivation for its existence. If this is the case, then having both types of explanation is not necessarily redundant, contra Blevins (2004). This is arguable especially true for patterns which are cross-linguistically marked or dispreferred - though not impossible - under a given theory. V/r/k splitting, discussed in §2.3.3, was one such case. Wamesa verbal infixation is another.

Verbal agreement in Wamesa is in most cases expressed by the addition of a prefix to the verb root. In certain instances, however, it surfaces instead as a vowel infixed after the root-initial consonant. This placement leads to cases of vowel hiatus which could easily be avoided if prefixation rather than infixation were to apply. This is unexpected: infixation here creates marked structures - onsetless syllables and discontinuous morphemes - without any obvious phonological or morphological motivation. The first half of this chapter will give a synchronic account of Wamesa infixation using alignment constraints, after McCarthy & Prince (1993) and McCarthy (2003), which can be used to produce the attested outputs.

While this theory is descriptively adequate, it does not explain the existence of such a phontactically unnatural pattern; all it can do is stipulate a ranking of constraints which produces the attested output. For more explanatory power I argue that we must look for a diachronic story as well. In this instance, while the evidence is scarce, a case can be made that coarticulation of the prefix-final vowel with the root-initial consonant in certain environments, reinforced by improved discrimination of contrasts in a prominent position, led eventually to full metathesis, instantiated synchronically as infixation.

The second half of the chapter gives an articulatorily-based historical account for how infixation arose. Bermúdez-Otero (2006: 6) notes that a theory of local, myopic sound change such as that set forth by Ohala (1992) and espoused here "predicts that phonologization is blind: it is driven by local phonetic properties and operates without regard for its global effects on the phonological system... [A] sequence of blind changes could easily lead to the violation of a universal markedness law," for example that requiring syllable onsets. This is precisely what I argue happened here. Certain phonetic properties of the affixed verb in an earlier form of the language, in this case higher degrees of coarticulation of the prefix vowel and root-initial consonant, which originally were simply part of the normal range of variation, were phonologized so that they themselves became the target pronunciation. This led to ambiguity as to the linear order of the vowel and consonant, already strongly coarticulated, and created a situation of metathesis. Each of these changes was locally improving in terms of the perception of the specific segments which they affected, but when taken in the larger context of the word created a more marked structure.

6.2. Wamesa Verbal Agreement: Background and Data

Wamesa verbal agreement markers usually appear as prefixes. The exception is in the 2nd and 3rd person singular, where the marker appears as a CV prefix ([bu-] and [di-], respectively) on vowel-initial verb roots, and as a -V- infix ([-i-] and [-u-]) on consonant-initial roots. Infixation is entirely predictable based on the shape of the verb root; there are no exceptions to this pattern of which I am aware. The full inflectional paradigms for two Wamesa verbs are given below in (6.1), reproduced from (4.49).

(6.1) *Wamesa Verbal Agreement Paradigm*

	prefix	*api* 'to eat'	*pera* 'to cut'
Singular			
1sg	/i-/	j-api	i-pera
2sg	/bu-/	bu-api	p<u>era
3sg	/di-/	di-api	p<i>era
Dual			
1du incl	/tur-/	tur-api	tu-pera
1du excl	/amur-/	amur-api	amu-pera
2du	/mur-/	mur-api	mu-pera
3du	/sur-/	sur-api	su-pera
Plural			
1pl incl	/tat-/	tat-api	ta-pera
1pl excl	/amat-/	amat-api	ama-pera
2pl	/met-/	met-api	me-pera
3pl hum	/set-/	set-api	se-pera
3pl NH	/si-/	si-api	si-pera

This pattern is unexpected for several reasons. First, infixation creates marked syllable structures by adding violations of the Onset constraint without any concomitant improvement of faithfulness or reduction of violations of other markedness constraints. The unattested form **dipera*, with prefixation rather than infixation in the 3rd person singular, is phonotactically less marked than attested *piera*; where the former has simple CV syllable structure throughout, the latter shows an instance of vowel hiatus, as well as the interruption of the stem by other morphological material. Second, infixation does not appear across the board for all CV- agreement affixes. The 3rd person plural non-human marker, /si-/, the only other CV agreement marker, always surfaces as a prefix, never as an infix, as does the first person singular marker /i-/. This suggests that /bu-/ and /di-/ specifically are picked out for infixation, whether with a diacritic, targeted constraints, or some other mechanism, where the other affixes in the paradigm are not.

6.3. Synchronic Accounts

6.3.1. Syllable Wellformedness Accounts

Many accounts of infixation, particularly those within the Austronesian language family, rely on syllable-wellformedness constraints to motivate the placement of the affix away from the root edge. Infixes reconstructable to Proto-Austronesian **um* and **in* (Dahl 1976; Ross 2002; Blust 1993b) appear in a number of modern Austronesian languages and are widely discussed in the literature. The modern reflexes generally appear attached near the left edge of the word, though the details of the shape and placement of the affix vary language to language (Yu 2007). Examples of reflexes of **um* in two modern languages, Chamorro and Tagalog, are shown in (6.2).

(6.2) a. *Chamorro*

epanglo	‘hunt crabs’	um-epanglo	‘to look for crabs’
gupu	‘to fly’	g<um>upu i paharu	‘the bird flew’
tristi	‘sad’	tr<um>isti	‘becomes sad’

(Anderson 1992: 208; Yu 2007: 91)

b. *Tagalog*

aral	um-aral	‘teach’
sulat	s<um>ulat	‘write’
gradwet	gr<um>adwet	‘graduate’

(French 1988; McCarthy & Prince 1993: 21)

As Yu (2007: 21) puts it, infixes are often taken to be underlyingly prefixes which "‘migrate’ only when the infixed outcome yields a ‘better’ surface form". The Tagalog data, for example, is accounted for by McCarthy & Prince (1993) as follows. The *um-* affix is taken to be underlyingly a prefix, as specified by an Align constraint indexed specifically to *um-*. On consonant-initial roots, however, the affix is realized after the root-initial consonant in order to satisfy a more highly-ranked NoCoda constraint. This interaction is demonstrated in (6.3), adapted from their (49). The Tagalog data and the syllable-based analysis are echoed for example by Zuraw (1996).

(6.3) *Tagalog um-*

/um+root/			No-Coda	Align-*um*
a.		um.grad.wet	***!	
b.		gum.rad.wet	***!	*
c.	☞	grum.ad.wet	**	**
d.	☞	u.ma.ral	*	
e.		a.ru.mal	*	*!*

A phonotactically-based analysis along these lines clearly won't hold for Wamesa. Infixation doesn't lead to a 'better' surface representation; instead, it creates an instance of vowel hiatus which would be avoided if the agreement marker appeared as a prefix instead. Wamesa is certainly not the only example of infixation which violates phonotactic universals, even in Austronesian. Example (6.4) shows instances of another modern instantiation of Proto-Austronesian **um* in Atayal, a Formosan language (Egerod 1965; Yu 2007), and an (unrelated) infixing nominalizer in Leti, a Central Malayo-Polynesian language spoken in the Moluccas (Blevins 1999).

(6.4) a. *Atayal animate actor focus*

qul	q<m>ul	'snatch'
hŋuʔ	h<m>ŋuʔ	'soak'

b. *Leti nominalizer*

kaati	'to carve'	k<ni>aati	'carving'
kini	'to kiss'	k<n>ini	'kissing'
mai	'to come'	m<i>ai	'coming'

Both of these cases are like the Wamesa pattern in that infixation occurs even though it violates both syllable markedness constraints and faithfulness ones. To account more satisfactorily for these types of patterns, we must move away from the use of phonotactic constraints and instead rely entirely on McCarthy & Prince's (1993) theory of Generalized Alignment, not only to move the affix towards the left word boundary but also to pull it rightwards.

6.3.2. Aligning the Affix

As discussed previously in §3.4, while Align constraints as originally conceived incur more violations the farther apart the two specified edges are, McCarthy (2003) points out that gradient alignment constraints are problematic. He makes the case that evaluating the violations of such constraints categorically solves the difficulties caused by gradient evaluation while still producing the desired output. He does this by specifying for each constraint not only the edges to be aligned, but also the maximum amount of material - nothing at all, one segment, one syllable, etc. - which that constraint will allow to surface between them. Any amount of material above the given threshhold will incur exactly one violation, regardless of its extent. In cases of infixation, separate, independently rankable constraints violated by intervening material of a segment or more, a syllable or more, a foot or more, etc. between the edges in question are necessary to avoid infixation too deep within the root. Here I will use mainly classic ALIGN constraints, but they will be evaluated categorically; the constraints which allow more than a zero amount of intervening material will be specified as such.

Looking for the moment only at the 2nd and 3rd person singular affixes on consonant-initial verb roots, the first step in the analysis is to motivate infixation. In GA, an affix appears as a prefix if there is a sufficiently highly ranked ALIGN prefix requiring that it appear at the left edge of the stem. Infixation occurs when a more highly ranked constraint requires another element, in this case the verb root, also to occur at the left edge of the stem. For now the change in the form of the affix from CV to V will be taken for granted; §6.3.3 will explore the motivation for deletion of the consonant.

(6.5) ALIGN-ROOT: Align(Root, L, Stem, L); requires the left edge of the root to coincide with the left edge of a stem, working against prefixation.

(6.6) PREFIX(agr): Align($[\text{AGR}]_{\text{Af}}$, L, Stem, L); requires the left edge of the agreement affix to align with the left edge of the verb stem, working against infixation. This constraint will be revised in §6.3.8 to account for the non-infixing agreement markers.

Example (6.7) gives a tableau for the verb root *pera* to demonstrate the interaction of these two constraints. For the sake of brevity, I will only use the 3rd person singular in this section; the violations for the 2nd person singular will be identical. In this and all further tableaux, affixal material will be bolded for clarity.

(6.7) *Alignment Constraints I*

/di + pera/	Align-Root	Prefix(agr)
a. ☞ p**i**era		*
b. **di**pera	*!	

These constraints are sufficient to ensure that the agreement affix does not surface as a prefix, but they make no predictions as to where it will in fact appear. Prefix/σ(agr) prevents the affix from moving too far to the right. As infixation interrupts the verb root, a Contiguity constraint barring discontinuous morphemes (McCarthy & Prince 1995) must be low-ranked, as any candidate which satisfies it will violate either Align-Root or Prefix/σ(agr), as shown in the tableau in (6.9).

(6.8) a. Prefix/σ(agr): like Prefix(agr) above, this constraint is optimally satisfied by aligning the left edge of the agreement affix to the left edge of the stem. But where Prefix(agr) assigns a violation mark for the presence of any intervening material regardless of its size, this version is violated when the affix is misaligned by a syllable or more.

b. Contiguity: Assigns a violation mark for any discontinuous morpheme.

(6.9) *Alignment Constraints II*

/di + pera/	Align-Root	Prefix(agr)	Prefix/σ(agr)	Contiguity
a. ☞ pi**e**ra		*		*
b. **di**pera	*! W	L		L
c. pe**di**ra		*	*! W	*
d. pera**di**		*	*! W	L

The tableau in (6.9) shows, for each row but that of the winner, whether that constraint prefers that row's candidate or the winner, if it differentiates at all. These comparisons show that Align-Root must be ranked over Prefix(agr) and Contiguity, and that Prefix/σ(agr) must be ranked above Contiguity. No other ranking arguments can be made at this point.

6.3.3. Cluster Simplification

In the attested candidate, *piera*, the agreement affix, underlyingly /di-/, is reduced to a single vowel. The inverse analysis, with underlying /-i-/ becoming surface [di-] with V-initial roots through consonant epenthesis, is implausible. First, the initial consonant is different for each affix, though it may not be diachronically coincidental that rounded [u] is preceded by labial [b] in the second person and high front [i] by apical [d] in the third. Secondly, a survey of related languages reveals that quite a few use the same consonants as Wamesa.[1] The lexically-specific nature of the consonants and their uniformity across the subgroup argue in favor of their being underlying, not epenthetic.

In order to account for the deletion of the initial /d/, a constraint barring consonant clusters in onset position is required. This is supported by the general phonotactics of the

1. See (6.28) for a comparison of agreement affixes in 14 Cenderawasih Bay languages.

language, as discussed in §2.2.3 and §2.2.1.

*ComplexOnset has three possible repairs: either the affix-initial consonant or the root-initial consonant could be deleted, or an epenthetic vowel could be added to break up the cluster. In the attested candidate, the affix-initial consonant is deleted while the root-initial consonant is retained. While this goes against the universal tendency to preserve the first of two segments in a cluster during simplification (Wilson 2000; McCarthy 2008), a root-initial segment may take precedence due to its prominent position and thereby avoid deletion, as pointed out by McCarthy (2008: 297). (See also Alderete 1995; Jun 1995; Beckman 1998). Deletion of either segment violates Max. The Max constraint is defined in (6.10) in terms of Correspondence Theory (McCarthy & Prince 1995, 1999).

No specific Max(Root) constraint is needed here to rule out a candidate such as **diera*, where it is the root-initial consonant which is deleted to avoid a cluster, as this form violates Align-Root, which is so far undominated. This is demonstrated by candidate (c) in Tableau (6.17) below. Max(Root) is unrankable with regards to the other constraints, and will be omitted in future tableaux.

Candidate (f), with epenthesis of a schwa to break up the cluster, is ruled out by Pfx/σ(agr); no Dep constraint is necessary at this point. Max must be ranked below Align-Root, Prefix/σ(agr), and *Complex. Rather than attempt to capture these complex ranking relationships within the tableau, one possible total ranking is given there. (6.12) gives a Hasse diagram showing the complete set of precedence relationships as they stand so far.

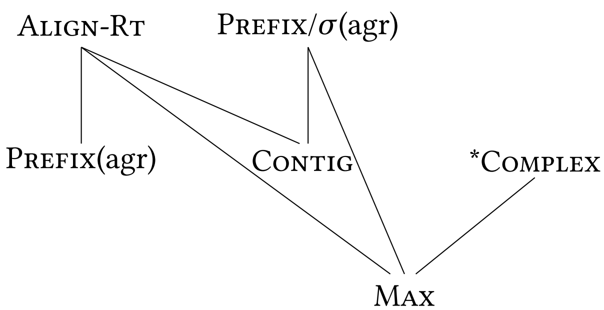
(6.10) a. *ComplexOnset: Assigns a violation mark for each instance of a consonant cluster in onset position.

b. Max: Assigns one violation mark for every segment in the input which lacks a corresponding segment in the output.

(6.11) *Cluster Simplification*

/di + pera/	Align-Rt	Pfx(agr)	Pfx/σ(agr)	Contig	*CO	Max	Max(Rt)
a. ☞ p**i**era		*		*		*	
b. **di**pera	*! W	L		L		L	
c. **di**era	*! W	L		L		L	* W
d. pera**di**		*	*! W	L		L	
e. p**di**era		*		*	*! W	L	
f. pə**di**era		*	*! W	*		L	

(6.12) *Constraint Rankings*



6.3.4. Parse-Morpheme

The next component of the analysis to ensure that the affix is realized at all; Align-Root in (6.17) is best satisfied when the affix is deleted entirely and the root can align with the left edge of the stem with root contiguity intact. Since the Prefix constraints evaluate a candidate based on the presence or extent of the material between the target positions of the affixes and their actual positions, these constraints are vacuously satisfied in the case where the morpheme fails to be realized at all. Max is too low-ranked to rule out the candidate with full affix deletion. This problem is solved by a high-ranking Parse-

Morph constraint which ensures the agreement affixes have some audible exponence in the output (Akinlabi 1996).

(6.13) Parse-Morph: “A morph must be realized in the output” (Akinlabi 1996:9).

This is supported by an independent aspect of Wamesa phonology. Vowel sequences of a high vowel plus [a] are avoided in the language in derived environments, i.e. when created at the boundary between an agreement affix and a root. Roots such as *adia* ‘fire’ and *kamuavu* ‘starfish’ exist and surface intact regardless of the position of word stress, but derived /$V_{[\text{high}]}$a/ sequences only surface if the /a/ is stressed. (See §2.3.4.) In these cases, it is the vowel of the affix which survives, defying the tendency in consonant clusters to preserve the segment belonging to the root over that belonging to the affix. In vowel-initial roots, where the agreement marker is a single infixed vowel, privileging the root vowel over the infix would lead to the complete disappearance of the affix in the surface form. This, however, is not what happens. The forms in (6.14) ilustrate this deletion in a consonant-initial root *saserai* ‘look for’, alongside hypothetical ungrammatical forms without /a/-deletion.

(6.14) *saserai ‘to look for’*

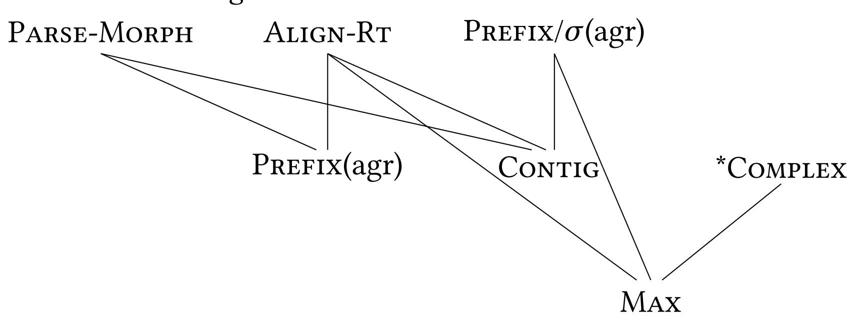
1sg	i-saserai	
2sg	s<u>serai	*s<u>aserai
3sg	s<i>serai	*s<i>aserai

Tableau (6.15) demonstrates the interaction of ParseMorph with the existing constraints and an updated Hasse diagram. Candidate (g) is the bare verb root, with no realization of the morpheme. In order to rule out this candidate, Parse-Morph must be ranked above Prefix(agr) and Contiguity.

(6.15) a. *Ensuring Morpheme Realization*

/di+pera/	Parse-Morph	Align-Rt	Pfx(agr)	Pfx/σ(agr)	Contig	*Complex	Max
a. ☞ p**i**era			*		*		*
b. **di**pera		*! W	L		L		L
c. pe**di**ra			*	*! W	*		L
d. pera**di**			*	*! W	L		L
e. p**di**era			*		*	*! W	L
f. pə**di**era			*	*! W	*		L
g. pera	*! W		L		L		** W

b. *Constraint Rankings II*



6.3.5. Vowel Hiatus

Finally, the winning candidate [piera] violates Onset by creating vowel hiatus in between the [i] and the [e]. A candidate such as (h) *[piʔera]with an epenthetic consonant breaking up the hiatus would fare better on that constraint than the attested form, as would candidate (i) *[pira], with deletion to prevent hiatus. Wamesa does not have a phoneme /ʔ/, and onsetless syllables are tolerated both word-initially and -medially.[2] To rule out

2. A glottal stop is sometimes added to phrase-initial onsetless syllables after a pause or a vowel-final word.

this repair we must include a standard Dep constraint, ranked above Onset. Onset must be ranked below Max.

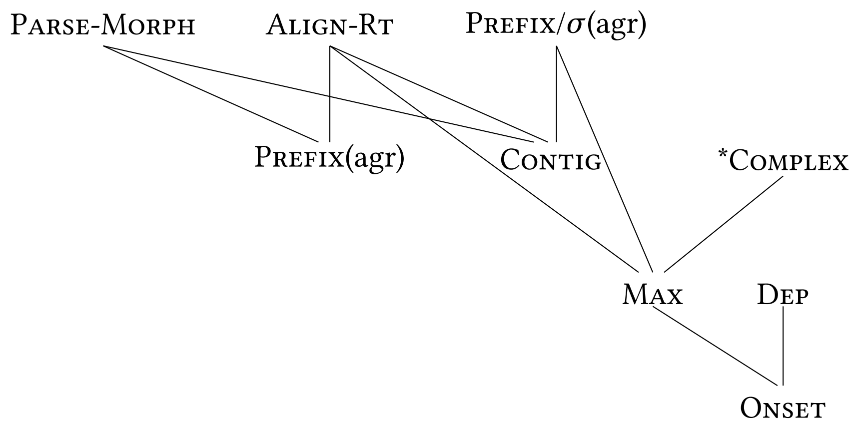
(6.16) a. Onset: Assigns a violation mark to any syllable not beginning with a consonant.

b. Dep: Assigns one violation mark for every segment in the output which does not correspond to a segment in the input.

(6.17) *Cluster Simplification*

/di+pera/	P-M	Align-Rt	Pfx(agr)	Pfx/σ(agr)	Contig	*Complex	Max	Dep	Onset
a. ☞ p**i**era			*		*		*		*
b. **di**pera		*! W	L		L		L		L
c. pe**di**ra			*	*! W	*		L		L
d. pera**di**			*	*! W	L		L		L
e. p**di**era			*		*	*! W	L		*
f. pə**di**era			*	*! W	*		L	* W	*
g. pera	*! W		L		L		**! W		L
h. p**i**ʔera			*		*		*	*! W	L
i. p**i**ra			*		*		**! W		L

(6.18) *Constraint Rankings III*



6.3.6. Vowel-Initial Roots

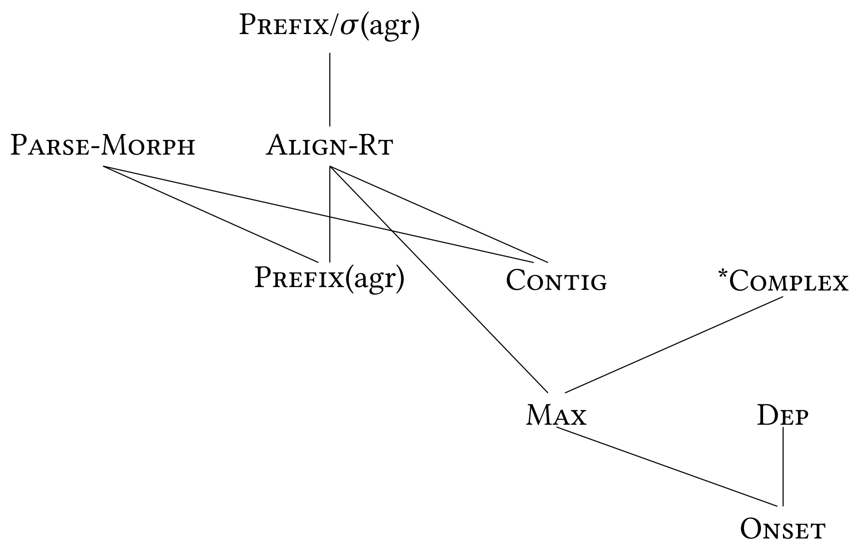
The current set of active constraints, as shown in (6.19), is sufficient to account for vowel-initial roots as well. For the sake of clarity, I refer here only to those constraints regarding affix placement, not shape; the agreement marker appears in its full CV form in this environment.

(6.19) *Vowel-Initial Roots: [diapi]*

/di + api/	PREFIX/σ(agr)	ALIGN-ROOT	PREFIX(agr)	CONTIGUITY
a. ☞ **di**api		*		
b. a**di**pi	*! W	L	* W	* W
c. api**di**	*! W	L	* W	

The above tableau provides an additional ranking argument for PREFIX/σ(agr). On vowel-initial roots, the agreement marker surfaces as a prefix, not an infix as with consonant-initial roots. In order to rule out infixed candidates such as (b), a constraint barring infixation must be ranked above ALIGN-ROOT. Furthermore, this constraint must only apply to infixation on vowel-initial roots, allowing it to occur on consonant-initial ones. That constraint is Prefix/σ(agr). Infixing the agreement markers on a C-initial root only offsets them by a single non-syllabic segment, which does not incur a violation of Prefix/σ(agr). On V-initial roots, however, if the agreement marker appears infixed after a single segment, that segment must be a vowel, which in itself constitutes a minimal syllable and violates Prefix/σ(agr). (6.20) gives an updated Hasse diagram of the constraint rankings.

(6.20) *Constraint Rankings IV*



Another candidate which must be taken into consideration is *[dapi], with the /di-/ prefix truncated to a single consonant. This candidate would fare better than the attested candidate [diapi] on Align-Root/σ, analogous to Prefix/σ(agr), and on general Onset. Align-Root/σ is defined in (6.21).

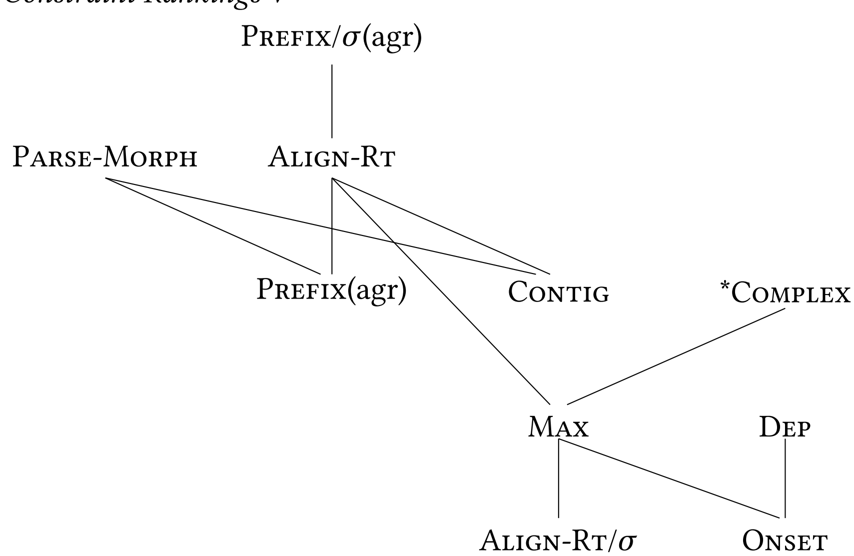
(6.21) Align-Root/σ: Assigns a violation mark to any material intervening between the left edge of the root and the left edge of the stem which constitutes a syllable or more.

We know that Onset is active in the language because of obligatory reduction of high vowels to glides word-initially and intervocalically, as discussed in §2.3.6. However, onsetless syllables do exist in Wamesa, and they are never avoided through outright deletion of a vowel. Onset therefore cannot be undominated, as demonstrated in §6.3.3. To produce the attested output, Align-Root/σ and Onset must both be ranked below Max. Tableau (6.22a) incorporates these additional constraints. Here I return again to providing a single consistent total ordering in the tableau, with an accompanying Hasse diagram.

(6.22) a. *Avoiding Truncation on V-Initial Roots*

/di + api/	Prefix/σ(agr)	Align-Rt	Prefix(agr)	Contig	Max	Onset	Align-Rt/σ
a. ☞ **di**api		*				*	*
b. a**di**pi	*! W	L	* W			*	L
c. api**di**	*! W	L	* W	* W		*	L
d. **d**api		*			*! W	L	L

b. *Constraint Rankings V*



6.3.7. Interim Summary

These constraints, in the partial ordering given by the Hasse diagram in (6.22b), fully account for the infixation patterns seen with 2nd and 3rd person singular agreement affixes on both consonant- and vowel-initial verb roots. The summary tableaux in (6.23a) and (6.23b) show how the complete set of constraints interacts to give the attested outputs.

(6.23) a. *3sg Summary Tableau: C-Initial Roots*

/di + pera/	Prefix/σ(agr)	P-M	Align-Rt	Prefix/(agr)	Contig	*Complex	Max	Dep	Onset	Align-Rt/σ
a. ☞ pi**i**era				*	*		*		*	
b. **di**pera			*! W	L	L		L		L	* W
c. pe**di**ra	*! W			*	*		L		L	
d. pera**di**	*! W			*	L		L		L	
e. p**di**era				*	*	*! W	L		*	
f. pə**di**era	*! W			*	*		L		*	* W
g. pi**ʔ**era				*	*		*	* W	L	
h. pera		*! W		L	L		** W		L	
i. p**i**ra				*	*		**! W		L	

b. *3sg Summary Tableau: V-Initial Roots*

/di + api/	Prefix/σ(agr)	P-M	Align-Rt	Prefix/(agr)	Contig	*Complex	Max	Dep	Onset	Align-Rt/σ
a. ☞ **di**api			*						*	*
b. a**di**pi	*! W		L	* W	* W				*	L
c. api**di**	*! W		L	* W	* W				*	L
d. **d**api			*				*! W		L	L
e. api		*! W	L				** W		*	L

6.3.8. Beyond 2nd and 3rd Singular

The final issue to be addressed is that this set of constraints is not yet sufficient for other person/number combinations. Given the current constraint set, all agreement affixes are incorrectly predicted to surface as infixes on C-initial roots. In reality, however, it is only the 2nd- and 3rd-person singular forms which ever appear as infixes. This is demonstrated in (6.24) with the 2nd person plural. The attested form is [sipera], but this analysis predicts [piera] instead. As above in §6.3.6, only the alignment constraints are initially shown here.

(6.24) *Incorrect Prediction: [sipera] '3pl.nh-cuts'*

/si + pera/	Prefix/σ(agr)	Align-Rt	Prefix/(agr)	Contig
a. (☞) **si**pera		*!		
b. ☞! p**i**era			*	*
c. pe**si**ra	*!		*	*
d. pera**si**	*!		*	

In order for /si-/ to appear as a prefix, there must be a constraint ranked above ALIGN-ROOT which demands that it do so. PREFIX/σ(agr) will not suffice in this case, as these affixes prefix even to consonant-initial roots, which does not create a full syllable's misalignment. Given the resemblance of /si-/ and /di-/, as well as 1st-person singular /i-/, it is hard to say that infixation only takes place in a particular phonetic environment without losing locality and, given the existence and exclusively prefixing behavior of /i-/, phonetic motivation.

The solution to this problem is to make the alignment constraints affix-specific. Rather than a single PREFIX/(agr) constraint governing all of the agreement markers, we instead need two: one governing the placement of /bu-/ and /di-/, ranked as before, and a second governing all of the other agreement affixes, ranked above ALIGN-ROOT. Under Generalized Alignment, the position of every affix must already be specified by some alignment constraint, and because Wamesa has both prefixes and suffixes, the grammar cannot simply specify the entire class of affixes as left-aligned in the word. Therefore it is not too large of an additional burden to rank the ALIGN constraints for some affixes differently than those for others. (6.25) gives the new affix-specific ALIGN constraints, and (6.26) gives an updated tableau for *sipera* incorporating them.

(6.25) a. PREFIX(2sg, 3sg): Align([AGR(2sg, 3sg)]$_{\mathrm{Af}}$, L, Stem, L); This is a cover constraint for two individual constraints, one of which requires the left edge of the 2nd person singular agreement affix to align with the left edge of the verb stem, and the other requiring the same of the 3rd person singular affix.

b. PREFIX(elsewhere): Align([AGR(1sg, Du, Pl)]$_{\mathrm{Af}}$, L, Stem, L); This is a cover constraint a family of individual constraints requiring the left edge of the all agreement affixes except the 2nd and 3rd person singular to align with the left edge of the verb stem.

(6.26) *Split Alignment Constraints*

/si + pera/	Pfx(1sg,du,pl)	Pfx/σ(agr)	Align-Rt	Pfx/(2,3sg)	Contig
a. ☞ **si**pera			*		
b. p**i**era	*! W				* W
c. pe**si**ra	*! W	* W	L		* W
d. pera**si**	*! W	* W	L		* W

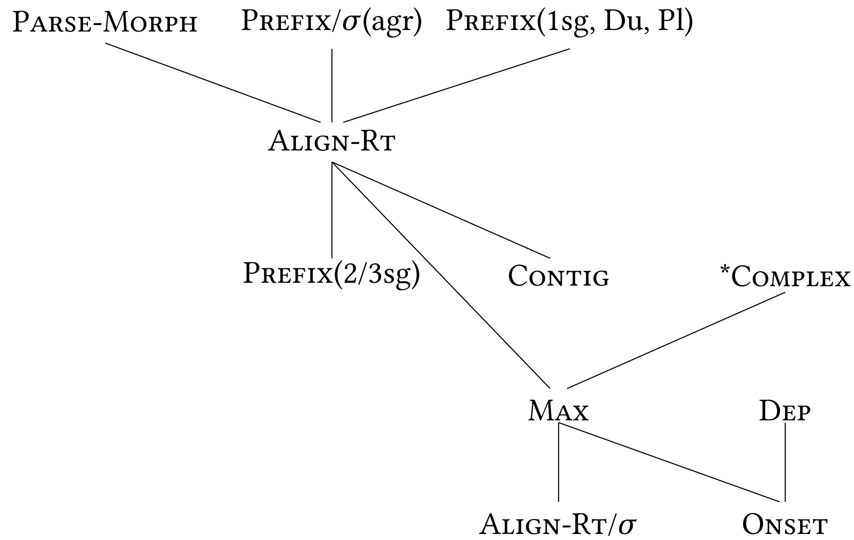
6.3.9. Summary Tableau and Final Constraint Rankings

The summary tableau for *sipera* in (6.27a), which includes the affixless candidate (e), gives us additional ranking information for Parse-Morpheme. For this form, it must be ranked above Align-Root. As Align-Root is already known to outrank Contiguity and Prefix(2sg, 3sg) (formerly Prefix(agr)), by the transitive property it continues to outrank them, as required by the 2nd and 3rd person affixes when attached to consonant-initial roots (as demonstrated in §6.3.4). This final constraint ranking is given in (6.27b).

(6.27) a. *Summary Tableau for [sipera]*

/si + pera/	P-M	Prefix(1sg, du, pl)	Prefix/σ(agr)	Align-Rt	Prefix/(2,3)	Contig	*Complex	Max	Dep	Onset
a. ☞ **si**perai				*				*		*
b. p**i**era		*! W		L		* W		*		*
c. pe**si**ra		*! W	* W	L		* W		*		*
d. pera**si**		*! W	* W	L				L		*
e. pera	*! W			L				** W		*

b. *Final Constraint Rankings*



6.3.10. Conclusion

As shown above, a combination of ALIGN constraints, when ordered correctly, can produce the attested outputs even for a pattern with phonotactically marked results, such as the Wamesa data discussed here. This is crucial: speakers of the language must have some way for their synchronic grammar to produce the attested outputs in order for them to be the attested outputs. But while the synchronic account is descriptively adequate, it is not particularly explanatory in any broader sense. What the above tableaux tell us is that the 2nd and 3rd person singular agreement affixes surface as infixes in order to satisfy a ranking of constraints which were expressly chosen and ordered to produce the infixed forms. There seems to be no benefit at all, either in terms of language-specific or cross-linguistic generalizations, to realizing the agreement marker as an infix; the alignment constraints simply tell us to put it there.

To explain why such a pattern might come to exist in the first place, I now turn to the set of historical changes which created the synchronic state of affairs. I argue that each of these changes led to a more natural articulatory or perceptual output, at least within their narrow scope, and that the interaction of these small beneficial changes led to a more unnatural pattern on a broader scale.

6.4. Historical Approaches

6.4.1. The Distribution of Infixation

According to Blust (1993a), verbs in Proto-Central-Eastern Malayo-Polynesian (PCEMP) showed agreement with their subject and direct object by means of a set of pronoun clitics. This is still the case in many CEMP languages, as Klamer (2002b) points out. Within CEMP, the Cenderawasih Bay (CB) languages show these clitics have been morphologized into affixes, and a subset of these, including Wamesa, Dusner (Dalrymple & Mofu 2012), Ambai (Silzer 1983), Biak (van den Heuvel 2006), Wooi (Sawaki in prep), Yerisiam, Yaur, and Yeretuar (David Kamholz p.c.), and a number of others (Anceaux 1961; Silzer 1983) show analogous patterns of infixation. Though the low-level relationships between the Cenderawasih Bay languages have not yet been determined in any detail, the majority of those mentioned above are generally classified as Yapen or Biakic (Lewis et al. 2013). A family tree of the languages in question, based largely on the classifications in Lewis et al. (2013), is presented in Figure 6.1. Subgroup names are in small caps; infixing language names are in italics, and non-infixing language names are in plain text. The lists of languages within each family are not necessarily exhaustive; only those languages mentioned in this paper are included. Tandia, an ungrouped language within Cenderawasih Bay (see example (1.2)), is excluded here due to lack of data.

It appears that a daughter language of Proto-Cenderawasih Bay existed which was a common ancestor of the modern infixing languages. In this proto-language, which I will call Proto-Biak-Yapen (PBY)[3] after the two largest subgroups of its daughter languages,

3. More reconstruction is needed to establish PBY and its development; preliminary comparison of the

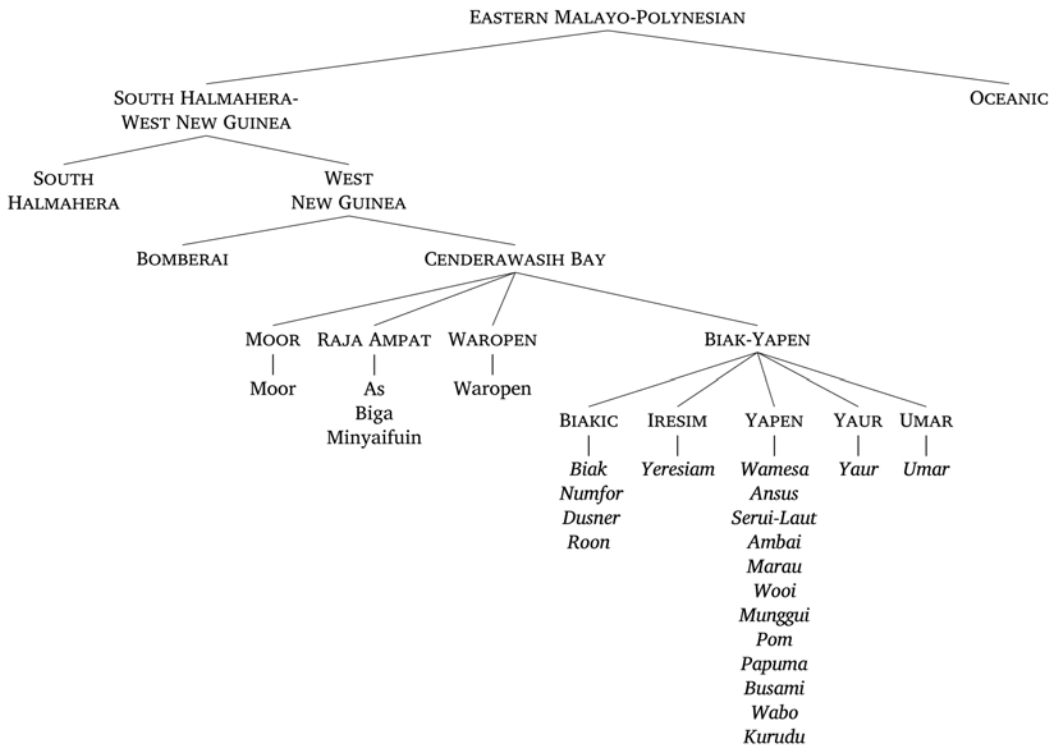


Figure 6.1: Eastern Malayo-Polynesian and the Distribution of Infixation

the current pattern of infixation in the 2nd and 3rd person singular arose. I argue that this happened via a long-term process of metathesis of the affix-final vowel and the root-initial consonant, motivated by increased acoustic discrimination, followed by a reanalysis of the location of the vowel and a simplification of the resulting cluster.

6.4.2. A Note on Data Sources

The Cenderawasih Bay languages, located as they are in a remote and politically volatile province of Indonesia, are severely underdocumented. Therefore the amount and reliability of the available data varies greatly language to language. The Ethnologue (Lewis

modern languages suggests a number of possible lexical innovations, including the words for 'tongue', 'feather', 'to live', 'smoke', 'lightning', and a collection of animal terms.

et al. 2013) lists 32 CB languages in total. For some of these, such as Meoswar, there is very little data at all beyond a few wordlists appearing in early collections, and no way of telling directly whether infixation exists in the language. Anceaux (1961) gives comparative wordlists and verbal paradigms, some more complete than others, for 13 languages. This data was collected in the first half of the 20th century by a range of linguists, explorers, missionaries, and Dutch colonial administrators. Greenhill et al. (2008) provide 210-item basic wordlists for nine CB languages, drawn from a number of sources. Ambai and Biak are each the subject of a full reference grammar submitted as a dissertation (Silzer 1983; van den Heuvel 2006). Silzer's Ambai grammar also includes a table of agreement affixes from 13 languages. Dusner is the subject of a recent sketch grammar (Dalrymple & Mofu 2012). Yusuf Sawaki provided me with a sketch of the agreement marking system of Wooi, the subject of his current fieldwork and upcoming dissertation, and David Kamholz provided verbal paradigms for Moor, Yerisiam, Yaur, and Umar. David Gil contributed data on Roon.

6.4.3. Cognacy

The first step is to establish that the agreement prefixes in the infixing languages did indeed descend from a common ancestor. In this case the strangeness of the pattern works in our favor; infixation is cross-linguistically relatively rare to begin with (Yu 2007), and the chances that it would arise independently across this many closely-related languages, and only in the 2nd and 3rd person singular, are vanishingly slim. Furthermore, the affixes in question closely resemble each other in all of these languages. Silzer's (1983) table of full pronouns and the singular verbal agreement markers as they appear with non-infixing and infixing verb roots, respectively, in 13 Cenderawasih Bay languages is reproduced in

(6.28).[4] Data from Roon (Gil 2010) has been added in the final row.[5]

(6.28) *Cenderawasih Bay Verbal Agreement Affixes (Singular)*

	1 sg		2 sg		3 sg	
	Pron.	Affix	Pron.	Affix	Pron.	Affix
Wooi	yau	y-/i-	au	bu-/-u-	i	ty-/-i-
Munggui	yau	y-/i-	au	w-/-u-	i	ty-/-i-
Pom	yau	y-/i-	au	w-/-u-	i	di-/-i-
Papuma	yau	y-/e-	au	w-/-u-	i	t-/-i-
Busami	yau	ya-	au	w-/-u-	i	s-/-i-
Wamesa	yau	y-/i-	au	bu-/-u-	i	di-/-i-
Ansus	yau	y-/e-	au	bu-/-u-	i	d-/-i-
Serui	yau	y-/i-	wau	bu-/-u-	i	d-/-i-
Ambai	yau	y-/i-	wau	bu-/-u-	i	d-/-i-
Wabo	aya	ay-/a-	awa	b-/-o-	i	d-/-i-
Kurudu	aya	ay-/a-	awa	b-/-u-	i	d-/-i-
Biak	aya	y-/ya-	au	w-/-u-	i	d-/i-
Waropen	ya	y-/ya-	auo	au-/a-	i	i(y)-/i-
Roon	ya	y(a)-/i-	aw	w(a)-/-u-	(t)i	t-/-i-

The tables in (6.29) below provide illustrative examples of these affixes as they are used in the CB languages other than Wamesa. They show verb conjugations in the singular from three infixing and one non-infixing language.

(6.29) a. *Dusner (Dalrymple & Mofu 2012)*

	ors 'to stand'	*man* 'to see'
1sg	**y**-ors	man
2sg	**w**-ors	m<**u**>an
3sg	**ndi**-ors	m<**i**>an

4. Some language names here are altered slightly to reflect current practice and for consistency. Original transcriptions using <y> to reflect the high front glide [j] and <ng> for the velar nasal [ŋ] are preserved.

5. Roon also has an animacy distinction in the third person singular. Animate forms are in the table; the inanimate forms are non-infixing.

b. *Yerisiam (Kamholz p.c.)*

	areki 'to see'	*ra* 'to go'
1sg	ne-**j**-areki	ne-ra
2sg	a-**gu**-areki	a-r<**u**>a
3sg	i-**di**-areki	i-r<**i**>a

c. *Wooi (Sawaki p.c.)*

	ihang 'to structure'	*ra* 'to go'
1sg	**y**-ihang	ra
2sg	**bu**-ihang	r<**u**>a
3sg	**t**-ihang	r<**i**>a

d. *Non-Infixing: Waropen (Anceaux 1961)*

	ano 'to eat'	*ra* 'to go'
1sg	**y**-ano	**ya**-ra
2sg	**au**-ano	**a**-ra
3sg	**iy**-ano	**i**-ra

These affixes and their interactions with the verb roots bear a striking resemblance to one another cross-linguistically. Of these languages, all but Waropen display infixation of the 2nd and 3rd person singular on at least some consonant-initial verb roots.[6] Based on these similarities, that these forms and their behavior are inherited features can hardly be doubted.

6.4.4. Historical Paths to Infixation

Yu (2007) cites five possible historical sources of infixation. These sources are as follows: 1) pre-existing infixation; 2) entrapment of an affix between two formerly independent,

6. Some languages, such as Biak (van den Heuvel 2006), show extremely complex patterns of agreement marking overall, and only a subset of the possible surface forms are presented here. The additional variations are suggested to be subsequent developments in the language.

now fused morphemes; 3) mutation of reduplication, in which later changes render the reduplication opaque; 4) morphological excrescence, predicated on the accidental similarity between internal syllables of unrelated words; and 5) metathesis. Of these five, only one, metathesis, can possibly have given rise to the Cenderawasih Bay pattern.

Strictly speaking, infixation in Wamesa is attributable to a pre-existing infixation pattern in its ancestor language, Proto-Biak-Yapen. This does not, however, explain its emergence in PBY itself, which is our focus here. It is impractical to suggest that infixation was inherited from any more distant ancestor than PBY. If it were innovated in Proto-Cenderawasih Bay, for example, we must explain not only how infixation emerged, but also why the process was reversed and prefixation restored in so many of the daughter languages. The problem is only exacerbated by placing infixation's source higher and higher up the family tree, as it then must have been lost in all branches but one, PBY. Far more likely than repeated undoing of the process at each split of the tree is that infixation was innovated in the one branch in which it appears, Biak-Yapen.

This leads us back to the question of how it arose. Entrapment occurs when an a morpheme *a*, often an affix, clitic, preverb, etc, becomes fused over time to a following morpheme *b*, a root. If a third morpheme *c* appears between *a* and *b* at the stage in which they are independent, it may still appear in that position after fusion, even though *a* and *b* are no longer analyzed as separate. Thus *c* becomes 'entrapped' and is synchronically an infix. There is no way this can be true of the PBY case, since the root-initial consonant after which the infixed vowel surfaces was not originally an independent morpheme, as comparison with cognates across the family attests.

A third possibility, after plain inheritance and entrapment, is reduplication mutation. This arises when a productive pattern of imperfect reduplication - that is, reduplication which does not entail perfect identity between the base and reduplicant - is reanalyzed as infixation, such as when a sound change renders the reduplication opaque. Yu gives the example of Hopi, citing data from Jeanne (1982). In traditional Hopi, plurals are formed by prefixing reduplication of the initial CV of the noun, followed by a shortening of the vowel of the base if it is long. A process of lenition changes a base-initial /p/ to [v], leading to the alternations *poosi* 'eye (sg)' ~ *poovosi* 'eyes (pl)', *patŋa* 'squash (sg)' ~ *paavatŋga* 'squash (pl)', and so forth. For younger speakers of Hopi, however, [v] is no longer an allophone of /p/ but a separate phoneme. Therefore the [vV] syllable of the base has been reanalyzed as an infix which, in the case of /p/-initial roots, copies the vowel of the root preceded by a [v] and inserts it after the first syllable.

Again though, there is no reason to believe that reduplication was involved in PBY. The infixed vowel is constant regardless of the other vowels in the root,[7] and no other segmental material is infixed which might be derived from the stem.

Finally, morphological excrescence, where a coincidental resemblance of certain forms is reanalyzed as an infix and made productive. This is exemplified by 'Homeric infixation' in American English, the insertion of the syllable *-ma-* to create colloquialisms such as *edu<ma>cate* and *saxo<ma>phone*. Yu argues that this pattern arose in English from filler words like *thingamajig* and *whatchamacallit*, each of which has *-ma-* after an initial trochee. This overlap of material, which he claims is purely accidental, was reanalyzed by speakers as a meaningful morpheme conveying imprecision or casualness, and its use was extended to other lexical items of the appropriate prosodic shape.

This too can be easily ruled out as a source of the PBY pattern. It is clear from the larger paradigms that the infixes did not arise from an accidental resemblance of unrelated forms, but from the prefixes which continue to surface as such on vowel-initial roots.

Instead, infixation in PBY was a classic case of what Blevins & Garrett (1998) call Perceptual Metathesis (PM). PM occurs when the acoustic cues associated with a given

7. Putting aside cases of ablaut, as in Yaur, which cause changes in the vowel but are attributable to assimilation of neighboring vowels, not reduplication.

segment have a particularly long duration. This persistence creates ambiguity regarding the origin of the cues in the word, leading to reinterpretation of the origin of the cues in question in a non-historical position. Blevins & Garrett list a number of features whose acoustic cues persist across a CV or VC domain, and which are thus predicted to participate in metathesis, including laterality, rhoticity, aspiration, glottalization, pharyngealization, and, relevantly for PBY, palatalization and labialization. In their view, metathesis is an extreme example of the case where loss of a high vowel leads to secondary articulation of the (formerly) adjacent consonant. Examples of methathesis of high vowels come from genetically diverse languages. As discussed by the authors, Greek shows metathesis of a high front vowel over consonants, with earlier **VCi* surfacing as *VjC* in the modern language, where the intervening C is a coronal. The Bantoid languages Aghem and Noni do the same for labialization; a reconstructed class 3 noun prefix **u* in the proto-language is realized here as an infixed glide [w], with the round vowel prefix persisting on one language as [o-] and lost in the other. PBY and its daughter languages combine these two patterns, with the persistent cues from both the [i] and the [u] leading to metathesis with the following consonant, in a case of hypo-correction (Ohala 1992, 1993).

6.4.5. Metathesis as a Morpheme-Specific Process

A further comparison of nine Cenderawasih Bay languages plus PCEMP[8] suggests that the metathesis of high vowels with a following consonant which yielded the modern infixation patterns was a morpheme-specific change, or at the very least did not occur within monomorphemic forms, even those of the appropriate phonemic shape. Basic vocabulary wordlists from the Austronesian Basic Vocabulary Database (Greenhill et al. 2008) were used to compare four infixing languages (Numfor, Ambai, Marau, and Wamesa) and five non-infixing ones (Moor, As, Biga, Minyaifuin, and Waropen). Of the 210 lexical items

8. As reconstructed by Blust (1993a).

examined, only one cognate set stands out as a possible case of metathesis of a high vowel within a morpheme: in non-infixing Biga, the word for 'male' appears as *wa-man*, with a bilabial glide before the [m], but this form is cognate with *mua* and *muaŋ*, with the corresponding vowel [u] after the [m], in the infixing languages Wamesa and Ambai, respectively. This however, is not a regular pattern: Biga *wa-bin* 'woman, female' is cognate with Ambai *vivi*, and Wamesa *vavi(n)*, with no corresponding [u] in the root . The origin of the [u] in the Wamesa and Ambai words for 'male' is unclear.

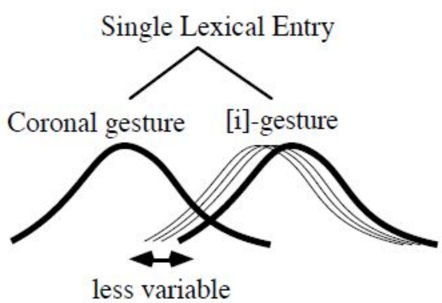
Further counter-evidence for a language-wide metathesis process comes from the word for 'kill', reconstructed by Blust as **bunuq* for PCEMP. Were metathesis possible within a morpheme this would appear to be an ideal form in which for it to take place, as the first syllable has precisely the same shape as the 2nd person singular prefix in most of the PBY daughter languages. With metathesis, we would expect the **u* of the first syllable to coalesce with that in the second syllable and initial **b* to disappear due to cluster reduction, yielding something like *nu(:)* in the modern infixing languages. The non-infixing languages should retain the initial syllable, with a modern form along the lines of *bunu*. In reality, while the modern cognates for the non-infixing languages do appear as *-bun* (As and Biga), *pun* (Minyaifun), *muna* (Waropen) and *muná* (Moor), the word remains unreduced in the infixing ones, appearing as *muni* (Ambai), *(ma)mun* (Numfor), and *mun* (Wamesa). The final **q* of the PCEMP form drops in all of the modern languages, and the now-final **u* of the second syllable surfaces alternately as *u*, *a*, *i*, or ∅ following regular patterns. The initial consonant is retained in all languages, nasalizing or devoicing in some, and the **u* of the first syllable remains unchanged in all instances. Another word of similar shape, CEMP **bulan* 'moon', surfaces in infixing Marau as *hura*), also failing to show any evidence of metathesis. The same is true of all other comparable words in these languages. This is strong evidence against metathesis occurring as a general process in Proto-Biak-Yapen.

This does not, however, preclude the occurrence of metathesis over a morpheme boundary. The account which follows is based largely on the timing relationships between the final vowel of the affix and the initial consonant of the verb root, and the timing of a heteromorphemic VC sequence may be quite different than from of a tautomorphemic one. Several authors (Ladefoged 1992; Byrd 1994; Fougeron & Steriade 1997) have claimed that intergestural timing is specified in the lexicon at the level of either the phonological word or the morpheme. Cho (1998a,b, 2001), using instrumental articulatory studies, provides evidence that timing is encoded in the lexicon at the morphemic level. In the former experiment, he examines the degree of variability in the timing of hetero- versus tautomorphemicic consonant clusters and [pi] sequences in Korean, comparing their timing in lexicalized versus non-lexicalized compounds. In both cases, he finds significantly greater deviations in the relative timing of the two segments in the heteromorphemic cases than the tautomorphemic ones, suggesting that within a morpheme, the relative timing of articulatory gestures is specified in the lexicon and therefore subject to less variation than the heteromorphemic sequences, which are not so specified.

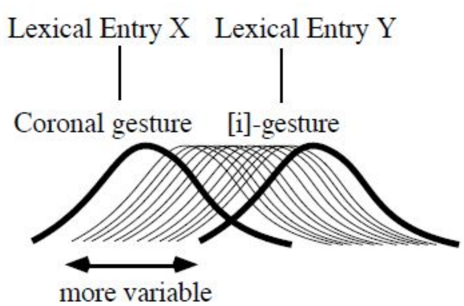
In the second study, Cho investigates the amount of gestural overlap between /t/ or /n/ and a following /i/ both within a morpheme and across a morpheme boundary but within a single word. His findings from the earlier experiment were confirmed here, in that the variation across a boundary was again greater than that within a morpheme. Additionally, Cho finds here that the temporal overlap of the gestures is actually greater between morphemes than within them, to a statistically significant extent. This holds both for /t/, which only palatalizes in Korean before a heteromorphemic /i/, and for /n/, which palatalizes before any /i/, though the difference is greater for /t/. Based on the increased gestural overlap in cases not specified for timing in the lexicon, Cho concludes that the preference of the grammar is actually for higher levels of overlap, and therefore more efficient transmission of cues.

Example (6.30) reproduces a schematic from Cho (1998b) illustrating the differences in variability between tauto- and heteromorphemic gestures demonstrated by his experimental data. In (6.30a), the two gestures, being elements of the same morpheme, have minimal variability in their degree of overlap. (6.30b) illustrates the increased variability of 'normal' articulations across a morpheme boundary.

(6.30) a.



b.



Extending these facts to Wamesa, we can say that change occurred only in the derived environment precisely because it was derived[9], and therefore unspecified in the lexicon with regards to the timing relationships, and subject to both greater variability and possibly greater default overlap. The order of vowel and consonant in Wamesa are reversed from that shown in the diagrams in (6.30), but the pattern is the same. As discussed in greater detail below, increased overlap of the articulatory gestures of the affixal high vowel[10] and the root-initial consonant - whether caused by increased variability of re-

9. Cho's findings have broader implications as well, providing a historical/articulatory motivation for some of the various patterns of synchronic phonology which fail to apply morpheme-internally.

10. No non-agreement prefixes end underlyingly in a high vowel. The applicative prefix *it*- surfaces as

alization, a preference in the grammar for higher overlap unless otherwise specified, or a combination of the two - led over time to metathesis of the segments. Within a root, the timing is specified in the lexicon, and therefore less variable and less conducive to metathesis.

6.4.6. Articulatory Motivations for Infixation

Where metathesis did occur, it was driven by an evolving set of articulatory and perceptual motivations. As Ohala (1992) points out, "sound change is not teleological"; it proceeds in a narrowly local fashion without regard for its effects on the phonological (or morphological) system as a whole, such that improvement in one area can lead to confusion in another, triggering further change. In this case, improved discrimination of segments encouraged greater coarticulation of the affixal vowel and root-initial consonant, which led in turn to ambiguity of the source of the high vowel's cues, and therefore to metathesis.

I suggest that metathesis here was driven initially by a subset of verb roots and subsequently generalized to the rest of the lexicon. Perceptual metathesis can occur over a wide range of segments, but cues for labialization and palatalization will pass more strongly over some classes of segments than others. As demonstrated by the Greek case above, palatalization will pass most readily over apical segments. In Aghem and Noni, metathesis occurred over all initial segments, but labial cues are most effectively passed over velar consonants (Silverman 2006). It is in these environments that metathesis was most favored, and these cases which drove the process in the rest of the lexicon.

The sequence of changes, in brief, was this: the first step towards metathesis was increased coarticulation of the high vowel with the following consonant, leading to round-

---

[i-] before a C-initial root due to cluster simplification, but there is reason to believe, based on van Balen's (1915a) bible translation and other early sources, that heterosyllabic clusters were preserved until fairly recently, and almost certainly remained at the point in history when infixation arose.

ing or palatalization of that consonant. As coarticulation continued to increase, the secondary articulations of the consonant developed into a full offglide. In these forms, the segmental source of the palatalization or labialization was ambiguous, and was reanalyzed as originating solely after the consonant. This process may have been reinforced by the unstressed nature of the agreement affixes - Blevins & Garrett (2004) discuss a type of 'compensatory metathesis' in which the features of an unstressed vowel move into a neighboring stressed syllable - but as the stress patterns of PBY are as yet unknown,[11] this remains speculative. Pressure from the apical-initial roots, which showed metathesis of /i/, and velar-initial forms, with metathesis of /u/, caused the paradigm to regularize (to a certain extent) and allow metathesis on roots with any initial consonant. This metathesis created an illegal cluster, which was then simplified to comply with the phonotactic requirements of PBY. The outcome of this sequence of events was the modern pattern of infixation.

6.4.7. Offglide Formation

For the 2nd person singular, I suggest that metathesis was most strongly motivated on velar-initial stems. To support this claim, let us first turn to a similar case in Trique, a Mixtecan language of Mexico.

As discussed by Silverman (2006), Trique has sequences of the form [uk^{w}a] and [uta], but not *[uka] or *[ut^{w}a]. (Trique has very few labial consonants.) The voiced series of stops follows the same pattern. In other words, velars in the environment /u_a/ always show a labial offglide, while alveolars never do. The [uk^{w}a] and [ug^{w}a] sequences in Trique

11. It is plausible that PBY had largely penultimate stress and disyllabic roots, but this has not been definitively shown. If this is the case, the agreement affixes would often appear in pre-stress position, and compensatory metathesis may well be a factor. PCEMP verb roots are reconstructed almost exclusively as disyllabic; the CB languages favor disyllabic roots but show a fair amount of diversity nonetheless. Stress throughout Austronesian is predominantly penultimate (Klamer 2002a; Blust 2009), but information on stress in the languages of West Papua is scarce. In Wamesa, stress on verb roots is largely penultimate but not exclusively so (see §3); Biak also shows a preference for penultimate stress but is similarly mixed.

are reconstructable as simply [uka] and [uga] (Longacre 1957), and appear as such in other Mixtecan languages. For example, the Trique word [ʒukʷa][12] 'snake' descends from Proto-Mixtecan (PM) **ʒu* 'animal' and **kɔ* 'snake, lizard'. PM *[kɔ] and *[ka] merged into [ka], followed by compounding and labialization to yield the modern form (Silverman 2006).

In the production of any word, adjacent segments will be coarticulated to a greater or lesser degree. Silverman argues that in Proto-Trique, coarticulation of the [u] with the following velar enhanced the perceptual contrast between an [uKa] sequence (where K can be either velar stop) and an [uTa] one, leading to improved lexical discrimination. Again, plain [uKa] sequences are not perceptually problematic - due to certain properties of the response of the auditory nerve, inter-vocalic stops are some of the most perceptible segments in a word (Wright 2004) - but more rounded [uKʷa] sequences are slightly better for discrimination. Silverman presents experimental data to support this claim.

Velars are a particularly good candidate for coarticulation with an [u] vowel. Articulatorily, as discussed by Silverman, the tongue shape necessary to produce an [u] brings the dorsum farther back in the oral cavity, bringing it closer to the target of the [k] gesture and making the distance traveled by the dorsum in order to achieve closure for the stop relatively short. This reduces the time needed to achieve the target closure, leaving less time to unround the lips if the velar is to be non-round. Further, since the velar and labial rounding gestures make use of different articulators, there is no physical impediment to coarticulation. From the perceptual side, it has previously been shown that labial and velar articulations have mutually-reinforcing effects on the acoustic signal, improving ease of discrimination on the part of listeners, most relevantly those children acquiring the language (Jakobson et al. 1961).

If greater coarticulation leads to better lexical discrimination, productions of an item

12. Tones omitted.

with more coarticulation will be correctly interpreted a higher percentage of the time.[13] It has repeatedly been shown that speakers match the frequency of different variants in their own productions to those in the language they hear with remarkable accuracy (see for example Labov 1994; Coleman & Pierrehumbert 1997; Zuraw 2000; Albright & Hayes 2003; and Liberman 2002, among many others). Thus if more strongly coarticulated productions of [uKa] are less often misheard as, say, [uta], over time they can be expected to make up an increasingly large proportion of actual tokens of [uKa]. Further, as the target pronunciation of the velar becomes rounder, outlying productions will also become rounder, and if these continue to improve lexical discrimination, as Silverman argues they do, gestural overlap of the [u] and the [K] will continue to increase, leading to an evolution in the target production of the sequence from original *[uKa] through intermediate *[uK̹a], with a rounded velar, finally to modern Trique [uK^{w}a], with a labial offglide.

It bears emphasizing that this path of change does not require any altruism (in the sense of Kingston 2002) or effort towards clear speech on the part of the speaker. Rather, of the range of degrees of coarticulation naturally produced by a speaker, certain productions are more often correctly identified by the listener and therefore slightly overrepresented compared to their actual proportion of tokens produced. Given what Hayes et al. (2009) call the Law of Frequency Matching, these more distinct, more coarticulated productions will thus be produced slightly more often, in a self-perpetuating cycle leading incrementally to the emergence of the full offglide.

This same story can be applied to Proto-Biak-Yapen as well. While in Trique the [uK] was morpheme-internal, in PBY it straddled the morpheme boundary between the 2nd person singular prefix **bu-* and a velar-initial verb root. Therefore we can posit an intermediate stage in which the 2nd person singular verb marking included both the *bu-*

13. See for example Scarborough (2004) for a laboratory investigation of coarticulation and confusability.

prefix and labialization of a root-initial velar.

An analogous change took place in the 3rd person singular, where coarticulation of the [i] of the prefix led over time to a full [^{j}] offglide on the following consonant. This may have been reinforced by analogy with the [^{w}] offglide in the 2nd person. Though palatalization is most often caused by a following high vowel, Bettoni-Techio & Koerich (2010), for example, has shown that for Brazilian learners of English, a preceding high vowel was sufficient to cause palatalization of word-final alveolar stops, and there are languages where this pattern is phonologized. The result is similar to the Greek case of **VCi* ⟶ *VjC* described by Blevins & Garrett (1998), but with the order of the segments reversed.

6.4.8. Full Metathesis and Spread by Analogy

The Trique change has not proceeded past this stage, but in PBY it appears to have continued on. At this point, the language was in the second of Blevins & Garrett's three proposed stages of diachronic metathesis, with the features of the vowel spreading over the features of the consonant and creating ambiguity as to their source. Increased overlap with the root-initial consonant would have already shortened the duration of the affixal high vowel; this, combined with the presence of the off glide after the root-initial consonant, means that the intrinsically persistent acoustic cues for rounding and palatalization will have been present over a lengthy stretch of the affixed word, and their source in the linear order of segments is unclear. Learners of PBY misparsed these cues as originating after the consonant rather than before it, as historically was the case, causing full VC metathesis.

Around the same time as metathesis, the pattern was regularized to include all verb roots in the language, not just the apical- and velar-initial ones. The analogical pressure came from two sides, with metathesis on apical roots after /i/ and on velar-initial roots after /u/, leaving only the bilabial-initial roots entirely unaffected before regularization. Further, given the persistence of these vowels' cues over all segments, magnified by the increased overlap of gestures over the morpheme boundary, even the initially unaffected segments would have been in an environment prone to metathesis, making this expansion unsurprising. The move towards distinctiveness for apical- and velar-inital roots which set this change into motion was overridden by the gain in paradigm uniformity. The full progression of steps from prefixation through metathesis to regular infixation is laid out in (6.31) below using the modern Wamesa roots *kopa* 'jump' and *topan* 'prepare' as examples, though presumably these forms have changed somewhat from PBY.

(6.31) *Steps to Infixation*

Original prefixation:	*bu-kopa, *di-topan
Increased coarticulation:	*buk̹opa, *dit̟opan
Offglide formation:	*buk^{w}opa, *dit^{j}opan
Full metathesis:	*bkuopa, *dtiopan
Cluster simplification:	kuopa, tiopan
Analogical regularization:	kuopa, tiopan, ruora, puera, etc.

Some evidence for the posited intermediate stage before the spread of infixation throughout the lexicon comes from Roon (Gil 2010), a Biakic language spoken on Roon Island, just north of the Wandamen Peninsula and adjacent to the Wamesa-speaking area. Consonant-initial verbs in Roon fall arbitrarily into two conjugation classes, one of which undergoes infixation, while the other does not. Apical- and velar-initial roots can be found in both classes; if infixation originally spread cleanly throughout these environments, it has since been undone in some cases by subsequent changes. Labial-initial infixing verbs, however, are rare. Though a small handful are attested in the modern form of the language, they are seriously underrepresented in that class (David Gil p.c.). This suggests that infixation never spread to labials in Roon as it did in languages such as Wamesa, and that only later developments in the language led to the inclusion of those few labial-initial items we do find in the infixing class.

6.4.9. Cluster Simplification

One final step is needed to take us to the forms attested in the modern infixing languages of Cenderawasih Bay. The metathesis of the prefix-final vowel and stem-initial consonant described above does not account for what happened to the prefix-initial [b] (2nd person) and [d] (3rd person). In most of the languages involved, consonant clusters are dispreferred to a greater or lesser degree. In Ambai, for example, only homorganic NC clusters are allowed, and only word-medially (Silzer 1983). The complex clusters of Biak, Numfor, and Dusner appear to have developed at a later stage, at least in part as the result of vowel deletion in certain environments in Proto-Biakic.[14] If we assume, as seems warranted, that the phonotactic structure of PBY resembled that of its non-Biakic daughter languages, then the cluster formed by the prefix-initial stop and the root-initial one after metathesis of the vowel would be illegal. The root-initial consonant was preserved in the cluster simplification, due to the prominence of its position (Beckman 1998; Jun 1995).

14. Some representative cognate sets are as follows. Moor is a non-infixing CB language. Ambai and Wamesa are members of the (infixing) Yapen branch of CB. Numfor and Biak are members of the (infixing) Biakic branch of CB. ([k] in Biakic languages corresponds regularly to [t] in other CB languages and PCEMP.)

Gloss	PCEMP	Moor	Ambai	Wamesa	Numfor	Biak
‘back’	-	-	kuruu	karu	kru[ri]	-
‘belly’	-	sine wariʔa (intestines)	ene	sane	sne[ri]	sne
‘to fear’	*ma-takut	-	matai	matai[t]	mkāk	mkák
‘to laugh’	*malip	mariʔa	miri	mari	mbrif	mríf
‘thick’	*kapal	-	-	-	kpor	kpor

(Greenhill et al. 2008; van den Heuvel 2006)

This final step brings us to the modern state of affairs, with infixation on consonant-initial roots in the 2nd and 3rd person singular and prefixation everywhere else.

6.4.10. 1st Person Singular *i-* ~ *j-*

The scenario described above brings up the question of why metathesis only occurred with the 2nd and 3rd person singular affixes and not any of the other affixes, particularly the the 1st person singular, which surfaces alternately as [i-] and [j-] in Wamesa, and the 3rd person singular non-human, [si-]. The essive prefix *ve-* did not trigger metathesis as it does not involve a high vowel, and is not part of the verbal paradigm which underwent regularization as the last step of infix development. The plural forms too lack a high vowel, and in any case the final consonant found in the plural and dual forms would have blocked metathesis, which happened only over a morpheme boundary and not over the clusters that still would have been present with C-initial forms. (See footnote 10.)

This leaves the 1st-person singular and 3rd-person plural non-human prefixes. Both of these morphemes end in a high vowel or glide in their modern forms and would appear to be subject to the same pressures, yet neither can appear as an infix in any of the languages in question. The answer here is that neither affix existed in a form with such a final high vowel at the point at which metathesis applied, and therefore neither was subject to it.

Returning to the table of affixes in (6.28) above, we see that the 1st person singular prefix surfaces as [ya-] in at least some environments in Busami and Biak;[15] the same is true of Dusner (Dalrymple & Mofu 2012: 10). In Wabo and Kurudu, the allomorph which appears with consonant-initial verb roots is simply [a-]. These languages are distributed across the Biakic and Yapen subgroups, so we cannot posit the addition of an [a] to the affix as an innovation within a single branch of CB. In fact, the Biakic languages have *lost*

15. as well as Waropen, a non-infixing language.

the vowel */a/ in many initial syllables, as mentioned in footnote 14. More likely, then, is that this prefix can be reconstructed as *[ja] in PBY (and probably Proto-Cenderawasih Bay, for that matter), with a final low vowel which was subsequently dropped in many of the daughter languages.

The loss of /a/ after a high vowel or glide does not appear to have been a regular change across the lexicons of the languages in question. There are instances in the wordlists of the sequences [ja] and [ia] in PCEMP and those languages with 1sg [ja-] appearing cognate with [a] or [ja] in languages with 1sg [j ~ i], such as CEMP **maya*, WAD *mamaya* 'shy, ashamed'; CEMP **tian*, WAD *sane* 'stomach'; and As, Minyaifun *yas*, Wamesa, Ambai *adia* 'fire'.[16] This reflects maintenance of the sequence or loss of the [j], respectively.

However, **a*-deletion after **i* does occur in two pronominal forms. The PCEMP 3rd person plural **sida*, reduced by the time of PBY to **sia*, appears in the infixing languages, including Wamesa, as *se-* or *si-*, with loss or coalescence of the **a* after an **i*. Similarly-shaped PCEMP **s-ia* appears across the infixing languages as *i* '3sg pronoun', again with loss of the **a*. There is evidence cross-linguistically that reduction of clitics and affixes in agreement paradigms can be unpredictable and irregular. Donohue (2003), for example, describes the development of the verbal agreement morphology in Skou, a non-Austronesian language of north-eastern Papua, which is highly irregular both synchronically and diachronically. In particular, the Proto-Skou 1st person singular prefix **ŋ-* now surfaces as ∅ on most verbs, but three verbs take *k-*, and one takes *n-*, with no apparent conditioning environment for the total loss of the segment versus just the loss of nasality versus fronting of the place of articulation.

As the agreement affixes are clearly cognate across languages, the alternative to

16. These forms were all judged to be cognate in the ABVD; correspondence sets were not double-checked by me.

slightly irregular dropping of the inherited affixal *a* is to posit the independent innovation of the same vowel in a number of languages across several branches of CB, clearly a less plausible hypothesis, especially given the tendency of agreement affixes to reduce over time rather than augment. This is further supported by the fact that the agreement prefixes appear to ultimately be descended from the same ancestor forms as the full pronouns, a common pattern cross-linguistically. Across the infixing languages, and in some non-infixing CB languages as well, the first person singular pronoun includes the final [a], appearing as [jau], [ja], or [aja]. The most plausible scenario, then, is that the full prefixes were reduced in their affixed state, first to *[ja-] by the point of PCBY, then further to [j-][17] in a subset of the modern languages. If this is the case, the high vowel would not have appeared adjacent to the root-initial consonant, and so would not have been in a position to trigger metathesis in PBY.

6.4.11. 3rd Person Plural Non-Human *si-*

The final piece of the puzzle here is the 3rd person plural non-human marker *si-*. Though it has a very similar shape, *si-* is unlike *bu-* and *di-* (and like *i-*) in that it never appears as an infix, regardless of the shape of the root to which it is attached. This is because it was most likely a later innovation in the language and therefore was not present when the process of metathesis was under way. Only a small subset of the Cenderawasih Bay languages have a separate agreement prefix for non-human or lower animate subjects in the 3rd person plural (or any other person/number combination, for that matter), namely Wamesa and the Biakic languages Dusner, Roon, and Biak. While complete paradigms including the plural are only available for a small number of languages, in Anceaux's (1961) section on the verbal systems of the Yapen languages he sees fit to point it out as an unusual feature of Wamesa. As none of the other Yapen languages appear to show this form, I suggest

17. In Wamesa at least, [i] and [j] are non-contrastive; see §2.3.6.

that it developed independently in Proto-Biakic and Wamesa, perhaps occurring first in one language and later as a contact-induced change in the other.[18] The agreement marker in Wamesa is *si-*, probably ultimately derived from the same source as the 3rd person plural human marker *set-*; in Dusner, Roon, and Biak it is non-cognate *na-*, suggesting an independent innovation. Finally, *si-* is the only agreement marker in the plural which does not end in /t/. This suggests that, while /t-/ may have been an independent plural number marker at an earlier stage, by the point at which *si-* was added to the language it had been reanalyzed as simply a part of the main agreement prefix for the other plurals. The same holds for the final /r/ of the dual agreement prefixes.

6.4.12. Historical Summary

Infixation is not uncommon in Austronesian languages - see for example the **-um-*, **-in-* prefixes reconstructed for Proto-Austronesian (Dahl 1976) and their related forms across the modern languages - but infixation which creates vowel hiatus is. Over the development of a subset of the Cenderawasih Bay languages, however, this tendency has been overcome by articulatory and perceptual pressures, yielding a typologically marked pattern which is nonetheless widespread in this group. While more work, both historical and documentary, is needed to establish the internal structure of this subgroup, infixation suggests a common ancestor language of the Yapen languages, the Biakic languages, Yerisiam, Yaur, and Umar, in which the verbal agreement prefixes present in many West New Guinea languages migrated to become infixes on verb roots of the appropriate form.

18. The Dusner-speaking village is surrounded by Wamesa territory, and bilingualism between the two was historically common. The three remaining speakers of Dusner all speak Wamesa (Dalrymple & Mofu 2012).

Bibliography

2013. Speed, n. *OED Online* Oxford University Press. <http://www.oed.com/view/Entry/186147?redirectedFrom=speedboat>.

Adriani, N. & A.C. Kruyt. 1914. *De Bare'e-sprekende Toradja's van Midden Celebes. Derde deel Taal- en letterkundige schets der Bare'e taal en overzicht van het taalgebied: Celeber-Zuid-Halmahera.* Batavia: Landsdrukkerij.

Akinlabi, Akinbiyi. 1996. Featural affixation. *Journal of Linguistics* 32. 239–289. [ROA-185].

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90. 119–161.

Alderete, John. 1995. Faithfulness to prosodic heads. ROA-94-0000.

Aldridge, Edith. 2009. Minimalist questions for the nominalist analysis of Tagalog syntax. *Theoretical Linguistics* 35(1). 51–62.

Anceaux, J.C. 1961. *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea.* M. Nijhoff.

Anceaux, J.C. 1992. *Irian Jaya source materials: The J.C. Anceaux collection of wordlists of Irian Jaya*, vol. 1 & 2 (B 1). Leiden/Jakarta: Department of Languages and Cultures of South-East Asian and Oceania, Leiden University.

Anderson, Anne, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth M. Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine

Sotillo, Henry S. Thompson & Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34. 351–366.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Anderson, Stephen R., Lea Brown, Alice Gaby & Jacqueline Lecarme. 2006. Life on the edge: There's morphology there after all! *Lingue e Linguaggio* 5. 33–48.

Bach, Emmon & Robert T. Harms. 1972. How do languages get crazy rules? In R.P. Stockwell & R. Macaulay (eds.), *Linguistic change and generative theory*, 1–21. Bloomington: Indiana University Press.

Baker, Mark. 2001. *The atoms of language: The mind's hidden rules of grammar*. Basic Books.

Baker, Mark. 2003. *Lexical categories: Verbs, nouns, and adjectives*, vol. 102, Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

van Balen, J.A. 1915a. *Bijbelsche geschiedenissen in de Windessische taal*. Amsterdam.

van Balen, J.A. 1915b. Windesische verhalen met vertaling en woordenlijst. *Bijdragen tot de Taal-, Land- en Volkenkunde* 70. 441–554.

Bammesberger, Alfred. 1982. *A handbook of Irish*, vol. 1: Essentials of Modern Irish. Heidelberg: Carl Winter Universitätsverlag.

Beckman, Jill N. 1998. *Positional faithfulness*. Ph.D. thesis, University of Massachusetts, Amherst.

Bennett, Ryan. 2012. *Foot-conditioned phonotactics and prosodic constituency*. Ph.D. thesis, UC Santa Cruz.

van den Berg, René. 2009. Possession in South Halmahera-West New Guinea: Typology and reconstruction. In Alexander Adelaar & Andrew Pawley (eds.), *Austronesian his-*

torical linguistics and culture history: A festschrift for Robert Blust, chap. 20, 327–357. Canberra: Pacific Linguistics.

Bermúdez-Otero, Ricardo. 2006. Phonological change in Optimality Theory. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 497–505. Oxford: Elsevier.

Bermúdez-Otero, Ricardo & Kersti Börjars. 2006. Markedness in phonology and in syntax: the problem of grounding. *Lingua* 116(5). 710–756.

Bettoni-Techio, Melissa & Rosana Denise Koerich. 2010. Preceding phonological context effects on palatalization in Brazilian Portuguese/English interphonology. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies* 55. 63–81.

Bianchi, Valentina. 2000. The raising analysis of relative clauses: A reply to Borsley. *Linguistic Inquiry* 31(1).

Bink, J.L. 1891. Lijstje van telwoorden en eenige zelfstandige naamwoorden enz. van Wandamen. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 34. 62–71.

Blevins, Juliette. 1997. Rules in Optimality Theory. In Iggy Rocca (ed.), *Derivations and constraints in phonology*, chap. 7, 227–260. Oxford: Clarendon Press.

Blevins, Juliette. 1999. Untangling Leti infixation. *Oceanic Linguistics* 38(2). 383–403.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Blevins, Juliette & Andrew Garrett. 1998. The origins of consonant-vowel metathesis. *Language* 74(3). 508–556.

Blevins, Juliette & Andrew Garrett. 2004. The evolution of metathesis. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 117–156. New York: Cambridge University Press.

Bloomfield, L. 1933a. *Language history.* New York: Holt, Rinehart and Winston.

Bloomfield, Leonard. 1933b. *Language.* New York: Holt, Rinehart and Winston.

Blust, Robert. 1978. Eastern Malayo-Polynesian: A subgrouping argument. In S.A. Wurm & Lois Carrington (eds.), *Second international conference on Austronesian linguistics: Proceedings*, 181–254. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

Blust, Robert. 1993a. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32. 241–293.

Blust, Robert. 1993b. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42. 438–478.

Blust, Robert. 2009. *The Austronesian languages.* Canberra: Pacific Linguistics.

Boersma, Paul & David Weenink. 2013. Praat: Doing phonetics by computer [computer program]. Version 5.3.74.

Booij, Geert. 1995. *The phonology of Dutch*, The Phonology of the World's Languages. New York: Oxford University Press.

Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide.* New York: Palgrave MacMillan.

Bowern, Claire. 2012. *A grammar of Bardi*, Mouton Grammar Library. Berlin: Mouton de Gruyter.

Bowern, Claire, Barry Alpher & Erich Round. 2013. Yidiny stress, length, and truncation reconsidered. Poster presented at the *44th North East Linguistic Society* (NELS 44), UConn, 18-20 October 2013.

Bowern, Claire, Barry Alpher & Erich Round. in prep. Yidiny stress, length, and truncation reconsidered.

Broselow, Ellen. 1999. Stress, epenthesis, and segment transformation in Selayarese loans. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 25. 311–325.

Broselow, Ellen. 2004. Unmarked structures and emergent rankings in second language phonology. *International Journal of Bilingualism* 8(1). 51–65.

Broselow, Ellen, Su-I Chen & Chilin Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20. 261–280.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–252.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica* 49. 155–180.

Browman, Catherine P. & Louis M. Goldstein. 1995. Gestural syllable position effects in American English. In Fredericka Bell-Berti & Lawrence J. Raphael (eds.), *Producing speech: Contemporary issues (for Katherine Safford Harris)*, 19–34. AIP Press.

Bruening, Benjamin. 2001. *Syntax at the edge: Cross-clausal phenomena and the syntax of Passamaquoddy*. Ph.D. thesis, MIT. Photos available for download at http://udel.edu/ bruening/scopeproject/scopeproject.html.

Buckley, Eugene. 1992. Kashaya closed-syllable shortening and prosodic syllabification. In Dawn Bates (ed.), *Proceedings of the tenth west coast conference on formal linguistics*.

Buckley, Eugene. 1997. Optimal iambs in Kashaya. *Rivista di Linguistica* 9(1). 9–52.

Buckley, Eugene. 2009. Locality in metrical typology. *Phonology* 26(3). 389–435.

Buckley, Eugene. 2013. Kashaya extrametricality and formal symmetry. *Supplemental Proceedings of Phonology 2013* To be published on-line by the Linguistic Society of America.

Büring, Daniel. 1999. Topic. In P. Bosch & R.van der Sand (eds.), *Focus – linguistic cognitive and computational perspectives*, 142–165. Cambridge: Cambridge University Press.

Burung, Willem. 2011. Wordlist for languages in Papua. Pusat Studi Bahasa dan Budaya Papua dan Center for Endangered Languages Documentation.

Byrd, Dani. 1994. *Articulatory timing in English consonant sequences*. Ph.D. thesis, University of California, Los Angeles.

Cardinaletti, Anna & Giuliana Giusti. 2006. The syntax of quantified phrases and quantitative clitics. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, vol. V, chap. 71, 23–93. Malden, MA: Blackwell.

Carlson, Gregory. 1977. *Reference to kinds in English.* Ph.D. thesis, UMass Amherst.

Carnie, Andrew. 2013. *Syntax: A generative introduction.* Malden, MA: Wiley-Blackwell, 3rd edn.

Chafe, Wallace. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex. Video available for download at http://pearstories.org/.

Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic fieldwork. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, chap. 7, 152–165. Cambridge: Cambridge University Press.

Cho, Taehong. 1998a. Intergestural overlap and timing in Korean palatalization. In David J. Silva (ed.), *Japanese/Korean linguistics*, vol. 8. Stanford: Stanford Linguistics Association.

Cho, Taehong. 1998b. *Specification of intergestural timing and gestural overlap: EMA and EPG studies.* Master's thesis, UCLA.

Cho, Taehong. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica* 58. 129–162.

Cinque, Guglielmo. 2000. On Greenberg's Universal 20 and the Semitic DP. In Laura Brugè (ed.), *University of Venice working papers in linguistics 10*, 2:45–61. University of Venice, Dipartimento di Scienze del Linguaggio.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36(3). 315–332.

Coetzee, Andries & Joe Pater. 2011. The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The handbook of phonological theory. 2nd edition*, 401–434. Oxford: Blackwell.

Coleman, John & Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational phonology, 3rd meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Comrie, Bernard. 1976. Irregular stress in Polish and Macedonian. *International Review of Slavic Linguistics* 1. 227–240.

Comrie, Bernard. 2013. *Numeral bases*. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL http://wals.info/chapter/131.

Cowan, H.K.J. 1953. *Voorlopige resultaten ven een ambtelijk taalonderzoek in Nieuw Guinea.* Den Haag: M. Nijhoff.

Cowan, HKJ. 1955. Notes on Windesi grammar. *Oceania* 26. 42–58.

Croft, William. 1990. *Typology and universals*, Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes*, Empirical Approaches to Language Typology vol. 23, 65–102. Berlin: Mouton de Gruyter.

Crowhurst, Megan. 1996. An optimal alternative to conflation. *Phonology* 13(3). 409–424.

Crowley, Terry. 2007. *Field linguistics: A beginner's guide*. Oxford: Oxford University Press.

Dahl, O.C. 1976. *Proto-Austronesian.* London: Curzon Press, 2nd edn.

Dalrymple, Mary & Suriel Mofu. 2012. *Dusner* (Languages of the World/Materials 4887). Munich: LINCOM GmbH.

Davis, Henry & Lisa Matthewson. 1999. On the functional determination of lexical categories. *Revue québécoise de linguistique* 27(2). 29–69.

Diessel, Holger. 2013. *Distance contrasts in demonstratives.* Leipzig: Max Planck Institute for Evolutionary Anthropology. URL http://wals.info/chapter/41.

Dixon, R.M.W. 1977. Some phonological rules in Yidiny. *Linguistic Inquiry* 8. 1–34.

Dixon, R.M.W. 1990. Reassigning underlying forms in Yidiny - a change during language death. In P Austin, R. M. W. Dixon, T Dutton & I White (eds.), *Language and history: Essays in honour of Luise A. Hercus*. Canberra: Pacific Linguistics.

Dixon, R.M.W. & Alexandra Aikhenvald. 2002. Word: A typological framework. In *Word: A cross-linguistic typology*, chap. 1, 1–41. Cambridge: Cambridge University Press.

Donohue, Mark. 1999. *A grammar of Tukang Besi*, vol. 20, Mouton Grammar Library. Berlin / New York: Mouton de Gruyter.

Donohue, Mark. 2003. Agreement in the Skou language: A historical account. *Oceanic Linguistics* 42(2). 479–498.

Donohue, Mark & Yusuf Sawaki. 2007. Papuan Malay pronominals: Forms and functions. *Oceanic Linguistics* 46(1). 253–276.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.

Dupoux, Emmanuel, Christophe Pallierb, Núria Sebastián-Gallés & Jacques Mehler. 1997. A destressing "deafness" in French? *Journal of Memory and Language* 36. 406–421.

Dupoux, Emmanuel & Sharon Peperkamp. 2002. Fossil markers of language development: phonological 'deafnesses' in adult speech processing. In J. Durand & B. Laks (eds.), *Phonetics, phonology, and cognition*, 168–190. Oxford: Oxford University Press.

Dupoux, Emmanuel, Sharon Peperkamp & Núria Sebastián-Gallés. 2001. A robust method to study stress "deafness". *Journal of the Acoustical Society of America* 110(3). 1606–1618.

Egerod, Søren. 1965. Verb inflexion in Atayal. *Lingua* 15. 251–82.

van Eijk, J.P. & T. Hess. 1986. Noun and verb in Salish. *Lingua* 69(4). 319–331.

Eisner, Jason. 1997. What constraints should OT allow? Paper presented at the Annual

Meeting of the Linguistic Society of America (LSA), Chicago, January 1997.

Elenbaas, Nine & René Kager. 1999. Ternary rhythm and the lapse constraint. *Phonology* 16(3). 273–330.

Esser, S.J. 1938. Talen. In The Hague (ed.), *Atlas van tropisch Nederland*, Het Koninklijk Aardrijkskundig Genootschap in Samenwerking met den Topografischen Dienst in Nederlandsch-Indië. Nijhoff.

Evans, Nicholas & Toshiki Osada. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9. 351–390.

Fehri, Abdelkader Fassi. 1988. Agreement in Arabic, binding and coherence. In Michael Barlow & Charles A. Ferguson (eds.), *Agreement in natural language: Approaches, theories, descriptions*, chap. 6, 107–158. Stanford, CA: Center for the Study of Language and Information.

Fields, Phil. 2010. Papuan Colloquial Indonesian. *SIL Electronic Working Papers* 5.

Flaming, Rachel. 1983a. Cohesion in Wandamen narrative. *Nusa: Linguistic Studies in Indonesian and Languages in Indonesia* 15. 41–49.

Flaming, Rachel. 1983b. Wandamen kinship terminology. In William R. Merrifield, Marilyn Gregerson & Daniel C. Ajamiseba (eds.), *Gods, heroes, kinsmen: Ethnographic studies from Irian Jaya, Indonesia*, vol. 17, International Museum of Cultures Publication, 245–53. Dallas: International Museum of Cultures.

Floyd, Simon. 2011. Rediscovering the Quechua adjective. *Linguistic Typology* 15. 25–63.

Fougeron, Cécile & Donca Steriade. 1997. Does deletion of French schwa lead to neutralization of lexical distinctions? In *5th European Conference on Speech Communication and Technology*. Rhodes, Greece.

Franks, Steven. 1989. The monosyllabic head effect. *Natural Language and Linguistics Theory* 7. 551–563.

Frascarelli, Mara. 2007. Subjects, topics and the interpretation of referential *pro*: An interface approach to the linking of (null) pronouns. *Natural Language and Linguistic Theory* 25. 691–734.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, vol. 100, Linguistik Aktuell/Linguistics Today, 87–116. Amsterdam: John Benjamins.

French, Koleen Matsuda. 1988. *Insights into Tagalog reduplication, infixation and stress from non-linear phonology*. Master's thesis, University of Texas, Arlington.

Fries, C. & K Pike. 1949. Coexistent phonemic systems. *Language* 25. 29–50.

Garrett, Andrew & Juliette Blevins. 2009. Analogical morphophonology. In Kristin Hanson & Sharon Inkelas (eds.), *The nature of the word: Essays in honor of Paul Kiparsky*, 527–545. Cambridge, MA: MIT.

Gick, Bryan. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In John Local, Richard Ogden & R. A. M. Temple (eds.), *Papers in laboratory phonology VI: Phonetic interpretation*, 222–236. Cambridge University Press.

Gick, Bryan, Fiona Campbell, Sunyoung Oh & Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34(1). 49–72.

Gil, David. 1993. Tagalog semantics. In *Proceedings of the nineteenth annual meeting of the Berkeley Linguistics Society: General session and parasession on semantic typology and semantic universals*, 390–403. Berkeley: Berkeley Linguistics Society.

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17(2). 179 – 200.

Gil, David. 2001. Escaping Eurocentrism: Fieldwork as a process of unlearning. In P. Newman & M. Ratliff (eds.), *Linguistic fieldwork*, 102–132. Cambridge: Cambridge University Press.

Gil, David. 2010. Word classes in Roon. In *Second TRIPLE international conference: Word classes: Nature, typology, computational representation*. Rome: Roma Tre University. March 26th.

Givón, Thomas. 1983. Topic continuity in discourse: An introduction. In T. Givón (ed.), *Topic continuity in discourse: A quantitative crosslanguage study*, 5–41. Amsterdam: John Benjamins.

Goedemans, Rob. 1996. An optimality account of onset-sensitive stress in quantity-insensitive languages. *The Linguistic Review* 13. 33–47.

Goldsmith, John. 1976. *Autosegmental phonology*. Ph.D. thesis, MIT, Cambridge, MA. Published by Garland Press, New York, 1979.

González, Carolina. 2005. Phonologically-conditioned allomorphy in Panoan: Towards an analysis. In Jeffrey Heinz, Andrew Martin & Katya Pertsova (eds.), *UCLA working papers in linguistics*, vol. 11 Papers in Phonology 6, 39–56. University of California, Los Angeles.

González, Carolina. 2007. Typological evidence for the separation between stress and foot structure. In Mattti Miestamo & Bernhard Wälchli (eds.), *New challenges in typology: Broadening the horizons and redefining the foundations*, Trends in Linguistics, 55–75. Berlin: Mouton de Gruyter.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.

Green, Thomas. 1995. The stress window in Pirahã: A reanalysis of rhythm in Optimality Theory. Massachusetts Institute of Technology, Cambridge.

Green, Thomas & Michael Kenstowicz. 1995. The Lapse constraint. Massachusetts Institute of Technology, Cambridge.

Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. The MIT Press.

Greenhill, Simon J., Robert Blust & Russell D. Gray. 2008. The Austronesian Basic Vocabulary Database: From bioinformatics to lexomics. *Evolutionary Bioinformatics* 4. 271–283.

Grijzenhout, Janet & Martin Krämer. 2000. Final devoicing and voicing assimilation in Dutch derivation and cliticization. In Barbara Stiebels & Dieter Wunderlich (eds.), *Lexicon in focus*, vol. 45, Studia Grammatica, 55–82. Berlin: Akademie Verlag.

Grimes, Charles E. 1990. *Comparative wordlist for Austronesian languages of Maluku, Nusa Tenggara and Irian Jaya*. Ambon: Unpati SIL.

Hale, Kenneth. 1992. On endangered languages and the safeguarding of diversity. *Language* 68. 1–3.

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23(3). 387–429.

Hammarström, Harald. 2010. Rarities in numeral systems. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory*, 11–61. Mouton de Gruyter.

Haraguchi, Shōsuke. 1977. *The tone pattern of Japanese: An autosegmental theory of phonology*. Tokyo: Kaitakusha.

Harriehausen, Bettina. 1990. *Hmong Njua: Syntaktische Analyse einer gesprochenen Sprache mithilfe daten-verarbeitungstechnischer Mittel und sprachvergleichende Beschreibung des südostasiatischen Sprachraumes*. Tübingen: Max Niemeyer Verlag.

Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, chap. II, 35–54. Berlin: De Gruyter Mouton.

Haspelmath, Martin. 2012. How to compare major word-classes across languages. In *UCLA working papers in linguistics, theories of everything*, vol. 17, 109–130.

Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*, Understanding

Language. Hodder Education, 2nd edn.

Hasselmo, Nils. 1970. Code-switching and modes of speaking. In G. Gilbert (ed.), *Texas studies in bilingualism*. Berlin: DeGruyter.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. In M. Niepokuj, V. VanClay, V. Nikiforidou & D. Feder (eds.), *Proceedings of the XIth annual meeting of the Berkeley Linguistics Society*, vol. 11, 429–446.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4).

Heggie, Lorie & Francisco Ordóñez. 2005. Clitic ordering phenomena: The path to generalizations. In Lorie Heggie & Francisco Ordóñez (eds.), *Clitic and affix combinations*. Amsterdam: John Benjamins.

Held, Gerrit Jan. 1942. *Grammatica van het Waropensch (Nederlandsch Noord Nieuw-Guinea)*. Bandung: Nix.

Hengeveld, Kees. 1992. *Non-verbal predication: Theory, typology, diachrony*. Berlin: Mouton de Gruyter.

Henning, Jean C., Theodore A. Henning, Mina Sawaki, Okta Netty Mananian, Tomas Yoteni & Dorce Webori-Mamori. 1991. *Sane pai ve pir kavavo nana kavo Wondama/Perbendahara'an kata bahasa Wandamen/Wandamen vocabulary* (Series B, Publikasi Khusus Bahasa-Bahasa Daerah 8). Irian Jaya, Indonesia: Universitas Cenderawasih and Summer Institute of Linguistics.

Henning, Theodore. 2014. Wandamen poetic language obsolescence: Demise of a death wailing tradition. Unpublished ms; available for download on academia.edu.

van den Heuvel, Wilco. 2006. *Biak: Description of an Austronesian language of Papua*. Ph.D. thesis, Vrije Universiteit Amsterdam.

van den Heuvel, Wilco. 2007. Grammatical borrowing in Biak. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 329–342. Berlin: Walter de Gruyter.

Hewitt, B. George. 1995. *Georgian: A structural reference grammar.* Amsterdam: John Benjamins.

Holle, K.F. 1982. *Holle lists: Vocabularies in languages of Indonesia*, vol. 5/1 Irian Jaya: Austronesian languages; Papuan languages, Digul area (D 52). Pacific Linguistics.

Houghton, Paula. 2006. Ternary stress.

Hyde, Brett. 2011. Extrametricality and non-finality. In Marc van Oostendorp, Colin Ewen, Beth Hume & Keren Rice (eds.), *A companion to phonology*, 1027–1051. Oxford: Blackwell.

Hyde, Brett. 2012. Alignment constraints. *Natural Language and Linguistic Theory* 30. 789–836.

Hyman, Larry M. 1981. Tonal accent in Somali. *Studies in African Linguistics* 12(2). 169–203.

Hyman, Larry M. 1982. Globality and the accentual analysis of Luganda tone. *Journal of Linguistic Research* 2. 1–40.

Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23. 225–257.

Itô, Junko. 1986. *Syllable theory in prosodic phonology.* Ph.D. thesis, UMAss Amherst.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7. 217–260.

Itô, Junko & Armin Mester. 2007. Prosodic adjunction in Japanese compounds. In Yoichi Miyamoto & Masao Ochi (eds.), *Proceedings of Formal approaches to Japanese linguistics (FAJL) 4*, 97–111. Cambridge, MA: MITWPL.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1961. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: The MIT Press.

Jeanne, LaVerne Masayesva. 1982. Some phonological rules of Hopi. *International Journal of American Linguistics* 48(3). 245–70.

Jun, Jongho. 1995. *Perceptual and articulatory factors in place assimilation: An Optimality Theoretic approach.* Ph.D. thesis, UCLA, Los Angeles.

Kager, René. 1994. Ternary rhythm in alignment theory. Research Institute for Language and Speech, Utrecht University.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch.* Dordrecht: Foris.

Kager, René. 2001. Rhythmic directionality by positional licensing. HILP-5, University of Potsdam, 11 January, 2001. ROA-514.

Kager, René. 2004. Rhythm, locality and non-gradient alignment. Paper presented at the *Third North-American Phonology Conference* (NAPhC3), Concordia University, 22 May 2004.

Kager, René. 2005. Rhythmic licensing theory: An extended typology. In *Proceedings of the third international conference on phonology*, 5–31. Seoul: The Phonology-Morphology Circle of Korea.

Kager, René. 2007. Feet and metrical stress. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, chap. 8, 195–227. Cambridge University Press.

Kager, René. 2012. Stress in windows: Language typology and factorial typology. *Lingua* 122. 1454–1493.

Kamholz, David. 2009. More Moor. Paper presented at the *11th International Conference on Austronesian Linguistics* (11ICAL), Aussois, France, 25 June 2009.

Kamma, Freerk Christiaans. n.d. Enkele aanwijzingen voor het Wandammess. Het Utrechts Archief, nummer toegang 6737, inventarisnummer 1102-2.

Kang, Yoonjung. 2002. Perceptual similarity in loanword adaptation: English post-vocalic word-final stops in Korean. *Phonology* 20. 219–274.

Kang, Yoonjung. 2011. *The Blackwell companion to phonology*, vol. IV: Phonological Interfaces, chap. 95: Loanword Phonology, 2258–2282. Wiley-Blackwell.

Karubuy, Theopilus. 2011. Verbal morphology in Wamesa language with reference to Windesi. Undergraduate thesis, Fakultas Sastra, Universitas Negeri Papua.

Kayne, Richard S. 1994. *The antisymetry of syntax*, Linguistic Inquiry Monograph 25. Cambridge, MA: The MIT Press.

Kenstowicz, Michael. 2003. The role of perception in loanword phonology. *Studies in African Linguistics* 32. 95–112.

Kenstowicz, Michael. 2005. The phonetics and phonology of Korean loanword adaptation. In S.-J. Rhee (ed.), *Proceedings of the first European conference on Korean linguistics*, 17–32. Seoul: Hankook Publishing Co.

Kenstowicz, Michael. 2007. Salience and similarity in loanword adaptation: a case study from Fijian. *Language Sciences* 29. 316–340.

Kingston, John. 2002. Keeping and losing contrasts. Paper presented at the 28th Annual Meeting of the Berkeley Linguistics Society.

Kinkade, M. Dale. 1983. Salish evidence against the universality of 'noun' and 'verb'. *Lingua* 60. 25–40.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In In-Seok Yang (ed.), *Linguistics in the morning calm*, 1–91. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology* 2(1). 85 – 138.

Kiparsky, Paul. 2006. The Amphichronic Program vs. Evolutionary Phonology. *Theoretical Linguistics* 32. 217–236.

Kitto, Catherine & Paul De Lacy. 1999. Correspondence and epenthetic quality. In Smallwood C Kitto, C. (ed.), *Proceedings of ALFA VI*. Toronto: University of Toronto Press.

Klamer, Marian. 1998. *A grammar of Kambera*. New York: Mouton de Gruyter.

Klamer, Marian. 2002a. Ten years of synchronic Austronesian linguistics (1991–2002).

Lingua 112. 933–965.

Klamer, Marian. 2002b. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41(2). 363–383.

Kluge, Angela. 2014. *A grammar of Papuan Malay*. Ph.D. thesis, Universiteit Leiden.

Kondo, Riena. 2001. Guahibo stress: Both trochaic and iambic. *International Journal of American Linguistics* 67(2). 136–166.

Kuno, Susumu. 1976. Subject, theme, and the speaker's empathy – a reexamination of relativization phenomena. In C. Li (ed.), *Subject and topic*, 417–444. New York: Academic Press.

Labov, William. 1994. *Principles of linguistic change, internal factors*. Oxford: Blackwell.

LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36(2). 223–258.

Ladefoged, Peter. 1992. The many interfaces between phonetics and phonology. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Phonologica: Proceedings of the 6th international phonology meeting*, 165–179. Cambridge: Cambridge University Press.

Ladefoged, Peter & Keith Johnson. 2011. *A course in phonetics*. Wadsworth Cengage Learning, 6th edn.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford; Cambridge, MA: Blackwell.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*, vol. 71, Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

Lapointe, Steven. 1990. EDGE features in GPSG. *Proceedings of the Chicago Linguistic Society* 26(1). 221–235.

Lapointe, Steven. 1992. Life on the EDGE: Arguments in favor of an autolexical account of edge inflections. *Proceedings of the Chicago Linguistic Society* 28(1). 318–332.

Lass, Roger. 1984. *Phonology*. Cambridge: Cambridge University Press.

Lavoie, Lisa M. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland Publishing.

Laycock, Don. 1978. A little Mor. In SA Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings*, 285–316.

Lee, Sunghwa. 2008. Disyllabicity in Nuuchahnulth. In Susie Jones (ed.), *Proceedings of the annual conference of the Canadian Linguistic Association*, vol. 15.

Levinson, Stephen C. 1996a. Frames of reference and Molyneux's question: Crosslinguistic evidence. In Paul Bloom, Mary A. Peterson, Lynne Nadel & Mary F. Garrett (eds.), *Language and space*, 109–169. MIT Press.

Levinson, Stephen C. 1996b. Language and space. *Annual Review of Anthropology* 25. 353–382.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2013. *Ethnologue: Languages of the world*. Dallas: SIL International, 17th edn. Online version: http://www.ethnologue.com/.

Liberman, Mark. 2002. Simple models for emergence of a shared vocabulary. Paper presented at Laboratory Phonology VIII. New Haven, CT.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.

Lombardi, Linda. 2003. Markedness and the typology of epenthetic vowels. Unpublished manuscript, University of Maryland (ROA).

Longacre, Robert E. 1957. Proto-Mixtecan. *International Journal of American Linguistics* 23(4). Indiana University Research Center in Anthropology, Folklore and Linguistics. Bloomington: Indiana University.

López Ixcoy, Calendaria Dominga. 1997. *Ri ukemiik ri K'ichee' chii': Gramática K'ichee'*.

Guatemala: Centro Educativo y Cultural Maya.

Mackey, W. 1970. Interference, integration and the synchronic fallacy. In J. Alatis (ed.), *Georgetown University Roundtable on Languages and Linguistics*, vol. 23, 195–227. Washington, DC: Georgetown University Press.

Martínez-Paricio, Violeta. 2012. Superfeet as recursion. In Nathan Arnett & Ryan Bennett (eds.), *Proceedings of the 30th west coast conference on formal linguistics*, 259–269. Somerville, MA: Cascadilla Proceedings Project.

Martínez-Paricio, Violeta. 2013. *An exploration of minimal and maximal metrical feet.* Ph.D. thesis, University of Tromsø.

Martínez-Paricio, Violeta & René Kager. 2013. Non-intervention constraints and the binary-to-ternary rhythmic continuum. Paper presented at the *21st Manchester Phonology Meeting*, University of Manchester, UK, 23 May 2013.

Mayer, Mercer. 1967. *A boy, a dog, and a frog.* New York: The Dial Press.

Mayer, Mercer. 1969. *Frog, where are you?* New York: The Dial Press.

Mayer, Mercer & Marianna Mayer. 1971. *A boy, a dog, a frog, and a friend.* New York: The Dial Press.

McCarthy, John J. 1991. Synchronic rule inversion. In *Proceedings of the 17th annual meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society.

McCarthy, John J. 1995. Extensions of faithfulness: Rotuman revisited. Available at: http://works.bepress.com/john$_{jm}$ccarthy/59.

McCarthy, John J. 2000. Harmonic serialism and parallelism. In M. Hirotani (ed.), *Proceedings of the North East Linguistic Society 30*, 501–524. Amherst, MA: GLSA Publications.

McCarthy, John J. 2002. *A thematic guide to Optimality Theory.* Cambridge: Cambridge University Press.

McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20. 75–138.

McCarthy, John J. 2007. Restraint of analysis. In S. Blaho, P. Bye & M. Krämer (eds.), *Freedom of analysis*, 203–231. Berlin and New York: Mouton de Gruyter.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25. 271–319.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. In G.E. Booij & J. van der Marle (eds.), *Yearbook of morphology*, 79–153. Dordrecht: Kluwer.

McCarthy, John J. & Alan Prince. 1993/2001. Prosodic morphology I: Constraint interaction and satisfaction. Tech. rep., ROA/RuCCS-TR-3.

McCarthy, John J. & Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In Mercè Gonzàlez (ed.), *Proceedings of the North East Linguistic Society*, vol. 24, 333–379. Amherst, MA: Graduate Linguistic Student Association.

McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. In J. Beckman, L. Dickey & S. Urbanczyk (eds.), *University of massachusetts occasional papers in linguistics 18: Papers in optimality theory*, 318–66. Cambridge, MA: Blackwell.

McCarthy, John J. & Alan Prince. 1999. Faithfulness and identity in prosodic morphology. In Rene Kager, Harry van der Hulst & Wim Zonneveld (eds.), *The prosody-morphology interface*, 218–309. Cambridge: Cambridge University Press.

McCarthy, John J. & Kathryn Pruitt. 2013. Sources of phonological structure. In Hans Broekhuis & Ralf Vogel (eds.), *Linguistic derivations and filtering: Minimalism and Optimality Theory*, Advances in Optimality Theory, chap. 5, 110–135. Equinox.

Mofu, Suriel. 2009. *Biak morphosyntax*. D.Phil. dissertation, Oxford University.

Murphy, R. P. 1974. Interference: Integration and the verbal repertoire. *Linguistics* 128. 59–67.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Foris Publications. Reprinted 2008 by Mouton de Gruyter.

NordNordWest/Wikipedia. 2010. Indonesia Maluku-Western New Guinea location map.svg. Available online http://commons.wikimedia.org/wiki/File:Indonesia_Maluku-

Western_New_Guinea_location_map.svg.

Ohala, John J. 1992. What's cognitive, what's not, in sound change. In G. Kellermann & M.D. Morrissey (eds.), *Diachrony within synchrony: language history and cognition*, 309–355. Frankfurt am Main: Peter Lang.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, chap. 9, 237–278. New York: Routledge.

O'Herin, Brian. 2001. Abaza applicatives. *Language* 77(3).

Ongkodharma, Nitya. 1985. Kepercayaan orang Wandamen. *Irian: The Bulletin of Irian Jaya* 13. 87–128.

Paradis, Carole. 1996. The inadequacy of filters and faithfulness in loanword adaptation. In J. Durand & B Laks (eds.), *Current trends in phonology: Models and methods*, 509–534. Salford, UK: University of Salford Publications.

Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.

Peperkamp, Sharon & Emmanuel Dupoux. 2002. A typological study of stress 'deafness'. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7*, 203–240. Berlin: Mouton de Gruyter.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In E. Reuland & A. ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge: MIT Press.

Poplack, Shana & Marjory Meechan. 1998. Introduction: How languages fit together in codemixing. *International Journal of Bilingualism* 2(2). 127–138.

Poplack, Shana & David Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22. 99–135.

Poplack, Shana, Susan Wheeler & Anneli Westwood. 1987. Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. In P. Lilius & M. Saari (eds.), *The Nordic languages and modern linguistics*, vol. 6, 33–56. Helsinki: University of Helsinki Press.

Price, David S. & Mark Donohue. 2007. Stress in Ambai and how it became that way. Paper presented at the *Workshop on the Languages of Papua* (WLP), Manokwari, Indonesia, 9 August 2007.

Price, David S. & Mark Donohue. 2009. *Report on the Ansus survey, West Yapen Island, Papua, Indonesia.* SIL International.

Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19–100.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. In Karen Deaton, Manuela Noske & Michael Ziolkowski (eds.), *CLS26-II: Papers from the parasession on the syllable in phonetics and phonology*, 355–398. Chicago: Chicago Linguistics Society.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar.* Malden, MA, and Oxford, UK: Blackwell. [Revision of 1993 technical report, Rutgers University Center for Cognitive Science. Available on Rutgers Optimality Archive, ROA-537].

Proctor, Michael. 2009. *Gestural characterization of a phonological class: The liquids.* Ph.D. thesis, Yale University.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *Comprehensive grammar of the English language.* London/New York: Longman.

Ramar, Origenes, Nitya Ongkodharma & Rachel Flaming. 1983. *Vekavavo nana kavo Wondama/Percakapan-percakapan dalam bahasa Wandamen/Wandamen conversations* (Series A, Publikasi Khusus Bahasa-Bahasa Daerah 6). Irian Jaya, Indonesia: Universitas Cenderawasih and Summer Institute of Linguistics.

Redden, James E. 1979. *A descriptive grammar of Ewondo*, vol. 4, Occasional Papers on Linguistics. Carbondale, Illinois: Southern Illinois University at Carbondale, Department of Linguistics.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53–94.

Remijsen, Bert. 2001. *Word-prosodic systems of Raja Ampat languages.* Utrecht: LOT.

Revithiadou, Anthoula. 1999. *Headmost accent wins: Head dominance and ideal prosodic form in lexical accent systems.* Ph.D. thesis, Universiteit Leiden.

Richards, Norvin. 2009. Nouns, verbs, and hidden structure in Tagalog. *Theoretical Linguistics* 35(1). 139–152.

Rose, Yvan & Katherine Demuth. 2006. Vowel epenthesis in loanword adaptation: representational and phonetic considerations. *Lingua* 116. 1112–1139.

Ross, Malcolm. 1995. Some current issues in Austronesian linguistics. In Darrell T. Tryon (ed.), *Comparative Austronesian dictionary*, 45– 120. Berlin: Mouton de Gruyter.

Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, vol. 518, Pacific Linguistics, 17–62. Canberra: Australian National University.

Ross, Malcolm. 2003. The grammaticalisation of directional verbs in Oceanic languages. In Isabelle Bril & Francoise Ozanne-Rivuerre (eds.), *Serial verbs and compound verbs in Oceanic languages.*

Ross, Malcolm. 2007. Talking about space: Terms of location and direction. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society.*, vol. 2: The physical world, chap. 8, 221–284. Canberra: ANU E Press.

Saggers, Naomi. 1979. *A sketch grammar of Wandamen.* Master's thesis.

Saltarelli, Mario, with Miren Azkarate, David Farwell, Jon Ortiz de Urbina & Lourdes Oñederra. 1988. *Basque*, Croom Helm Descriptive Grammars. London: Croom Helm.

Saragih, Chrisma Fernando. 2012. *The practical use of person reference in Papuan Malay.*

Master's thesis, Radboud University Nijmegen.

Sawaki, Yusuf. in prep. *Forthcoming Wooi dissertation*. Ph.D. thesis, The Australian National University.

Scarborough, Rebecca. 2004. *Coarticulation and the structure of the lexicon*. Ph.D. thesis, UCLA.

Schachter, Paul & Fé T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press. Reprinted in 1983.

Selkirk, Elisabeth. 1981. On the nature of phonological representation. In Terry Myers, John Laver & John Anderson (eds.), *The cognitive representation of speech*, 379–388. Amsterdam: North Holland Publishing Co.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, Elisabeth. 1995. The prosodic structure of function words. In Jill Beckman et al (ed.), *Papers in Optimality Theory*. Amherst, MA: GLSA.

Siltzer, Peter J. & Helja Heikkinen. 1984. *Index of Irian Jaya languages*, vol. 12, IRIAN Bulletin of Irian Jaya. Program Kerjasama Universitas Cenderwasih dan Summer Institute of Linguistics.

Silverman, Daniel. 2006. The diachrony of labiality in Trique, and the functional relevance of gradience and variation. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory phonology 8: Varieties of phonological competence*, 133–52. Mouton de Gruyter: Mouton de Gruyter.

Silzer, Peter J. 1983. *Ambai: An Austronesian language of Irian Jaya, Indonesia*. Ph.D. thesis, Australian National University.

Slump, F. 1924-38. Grammatica en woordenlijst van de Seroei-Laut taal. Het Utrechts Archief, nummer toegang 2274, inventarisnummer 1102-1.

Smith, Norval SH. 1977. Vowel epithesis in the Surinam creoles. *Amsterdam Creole Studies*

1. 1–54.

Stassen, Leon. 2000. AND-languages and WITH-languages. *Linguistic Typology* 4(1). 1–54.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-Map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.), *The nature of the word*, 151–179. MIT Press.

Stevens, Alan M. & Ed A. Schmidgall-Tellings. 2004. *A comprehensive English-Indonesian dictionary*. Athens, OH: Ohio University Press.

Strawson, Peter. 1964. Identifying reference and truth values. *Theoria* 30. 96–118.

Sweet, Henry. 1875/6. Word, logic, and grammar. *Transactions of the Philological Society for 1875/76* 470–83.

Tadmor, Uri. 2007. Grammatical borrowing in Indonesian. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 301–328. Berlin: Walter de Gruyter.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, chap. III, 55–75. Berlin: De Gruyter Mouton.

Thomas, Erik R & Tyler Kendall. 2007. NORM: The vowel normalization and plotting suite. Online resource: http://ncslaap.lib.ncsu.edu/tools/norm/norm1.php.

Uffmann, Christian. 2006. Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua* 116. 1079–1111.

Vennemann, Theo. 1972. Rule inversion. *Lingua* 29(3,4). 209–242.

Vergnaud, Jean-Roger. 1974. *French relative clauses*. Ph.D. thesis, MIT.

de Vries, J. W. 1988. Dutch loanwords in Indonesian. *International Journal of the Sociology of Language* 73. 121–136.

Wanner, Dieter. 1977. On the order of clitics in Italian. *Lingua* 43(2-3). 101–128.

Weber, David John. 1989. *A grammar of Huallaga (Huánaco) Quechua*. Berkeley: University of California Press.

Wilson, Colin. 2000. *Targeted constraints: An approach to contextual neutralization in Optimality Theory*. Ph.D. thesis, Johns Hopkins University, Baltimore, MD.

Wissing, Daan & Wim Zonneveld. 1996. Final devoicing as a robust phenomenon in second language acquisition: Tswana, English and Afrikaans. *South African Journal of Linguistics, Supplement* 34. 3 – 24.

Woodbury, Anthony C. 2002. The word in Cup'ik. In *Word: A cross-linguistic typology*, chap. 3, 79–99. Cambridge: Cambridge University Press.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 34–57. New York: Cambridge University Press.

Yallop, Colin. 1977. *Alyawarra: An aboriginal language of central Australia*. Canberra: Australian Institute of Aboriginal Studies.

Yu, Alan C. L. 2007. *A natural history of infixation*. Oxford: Oxford University Press.

van Zanten, Ellen, Ruben Stoel & Bert Remijsen. 2010. Stress types in Austronesian languages. In Rob Goedemans & Harry van der Hulst (eds.), *Stress patterns of the world: Data*. Amsterdam: John Benjamins.

Zuraw, Kie. 1996. *Floating phonotactics: Variability in infixation and reduplication of Tagalog loanwords*. Master's thesis, UCLA.

Zuraw, Kie. 2000. *Patterned exceptions in phonology*. Ph.D. thesis, University of California, Los Angeles, Los Angeles.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28. 417–472.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59. 502–13.