DEVELOPING A TALKING DICTIONARY FOR A MORPHOLOGICALLY COMPLEX LANGUAGE: THE CASE OF LAZ

AUDREY LYNN RICHARDSON

BOĞAZİÇİ UNIVERSITY

2019

DEVELOPING A TALKING DICTIONARY FOR A MORPHOLOGICALLY COMPLEX LANGUAGE: THE CASE OF LAZ

Thesis submitted to

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Linguistics

by

Audrey Lynn Richardson

Boğaziçi University

2019

DECLARATION OF ORIGINALITY

I, Audrey Lynn Richardson, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature Audrey Richardson

Date 30/07/2019

ABSTRACT

Developing a Talking Dictionary for a Morphologically Complex Language:

The Case of Laz

In this thesis, I conduct a position class morphology (PCM) analysis (Inkelas, 1993) of the verbal morphology of Laz. I then apply this analysis in a discussion of computational methods in order to develop a talking dictionary (TD) which may also serve as a language learning resource and thus aid in language revitalization.

Following a brief overview of both the current state of language revitalization materials (including TDs) and the Laz language, I implement a PCM analysis of the Laz verbal template, in which I examine cases of morphological blocking among affixes and thus develop a hierarchical model which relates precedence to relative dominance. Following this morphological analysis, I discuss a potential computational application of the PCM model, identifying finite-state transducers (FSTs) to be the most promising avenue due to their compatibility with the morphological hierarchy. I propose a bidirectional FST model, utilizing both precedence-driven and dominance-driven ordering of states. The intersection of PCM and finite-state technology is exploited in that the states corresponding to verbal affixes follow a path of dominance (as specified by PCM) when translated from morphological features to Laz and a path of precedence when translated from Laz to morphological features. Ultimately, this approach represents valuable first steps in the application of morphological analysis to the sphere of language revitalization and highlights an important avenue for future study.

ÖZET

Morfolojik Olarak Kompleks Bir Dil İçin Konuşan Sözlük Geliştirme Süreci: Lazca

Bu tezde, Lazcanın sözlü morfolojisini, pozisyonel sınıflandırma morfolojisi (PSM) yöntemini kullanarak analiz etmekteyim. Daha sonrasında, bu analizi, Lazca için bir konuşan sözlük (KS) geliştirmek maksadıyla bilgisayarlı yöntemler üstüne yürüttüğüm tartışmaya uygulamaktayım. KS'nin, Lazca öğrenmek isteyenler için bir kaynak olacağını ve dilin yeniden canlandırılmasına katkı sağlayacağını düşünmekteyim.

Hem mevcut dil canlandırma materyallerinin (KS'ler de dahil) hem de Lazcanın mevcut durumuna değindikten sonra, Lazcanın sözel şablonu üstünde PSM analizi gerçekleştirmekteyim. Bunun için, takılar arasında morfolojik öbekleme vakalarını incelemekteyim ve böylece göreli hâkimiyetle bağlantılı hiyerarşik bir model geliştirmekteyim. Morfolojik analizin ardından, PSM modelinin sayısal uygulaması üstüne bir tartışma yürütmekteyim ve morfolojik hiyerarşinin yapısına uygunluğundan dolayı sonlu durum dönüştürücülerini (SDD) en umut vaat eden yaklaşım olarak ele almaktayım. Hem öncelik-odaklı hem de hâkimiyet-odaklı durumların sıralamasını kullanan iki yönlü bir SDD modeli önermekteyim. Burada, PSM analizi ve sonlu durum teknolojisinin kesişiminden faydalanmaktayım. Sözlü eklere karşılık gelen durumlar, morfolojik özelliklerden İngilizceye çevrildikleri zaman, bir hakimiyet yolu (PSM'de belirtildiği gibi) izlerlerken, Lazcadan morfolojik özelliklere çevrildiklerinde bir öncelik yolu izlerler. Sonuç olarak, bu yaklaşım, dil canlandırma alanında morfolojik analiz uygulanması adına önemli bir adımdır ve gelecekteki çalışmaların önünü açıcı niteliktedir.

ACKNOWLEDGEMENTS

At the risk of including too many *Lord of the Rings* references for a master's thesis in linguistics, I must admit that having reached this point I feel quite like Frodo, standing at the brink of Mount Doom and having just destroyed the One Ring: "It's gone. It's done."

I would not have been able to reach this point without the help of several important people, whom I would like to thank here.

Firstly, my husband Peter Godfrey has been the Samwise to my Frodo throughout this journey. He supported me with constant love and patience in moments of both progress and procrastination, and he helped carry the burden of my stress and hopelessness during the three years of this thesis-writing process. I couldn't have done this without him.

I would also like to thank my equally patient and wise linguistics professors at Boğaziçi, especially Aslı Göksel, Pavel Logačev, and Balkız Öztürk, who provided me with valuable tips and feedback, even when the writing process extended far beyond what any of us expected.

I would especially like to thank my Laz professor and informant, İsmail Avcı Bucaklişi, who took time out of his busy schedule to help me out and explain complicated aspects of his first language (Laz), primarily using his third language (English). Without his instruction and feedback, I never would have come to learn all I now know of the fascinating language and unique culture of the Laz people. His efforts to popularize Laz in academia and social media are inspiring, and I hope that this will not be the last of our linguistic collaboration.

I also thank my dear friend Onur Calap, who has always been a sympathetic ear for me to vent my numerous frustrations. I feel honored that he lent his translation prowess to both the abstract of this thesis and my frantic phone calls with various bureaucratic officials in the final days before my deadline.

Finally, I would like to thank my family and friends for always supporting me from afar. My parents, Terry and Anita Richardson, always believed in me and never failed to ask about my thesis progress. Thank you to all my friends who were understanding (albeit unimpressed) when I continually evaded social events for thesis-related reasons. And, last but not least, I owe quite a bit of my sanity to Rimjhim, my emotional support thesis dog and a very good girl.

This has been an incredible, painful process, and I've learned so much in these past few years—about Laz, about linguistics, and about myself. I will carry all these lessons with me as I continue to move forward, both in language and in life. *Ǩaite*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

1/2/3	first/second/third person
APPL	applicative
AUG	augmentative
AUX	auxiliary
CAUS	causative
CAUS$_{INTR}$	intransitive causative
CAUS$_{PERF}$	perfective causative
CAUS$_{TR}$	transitive causative
CL	class (i.e. position class)
COND	conditional
COP	copula
DAT	dative
ERG	ergative
FSM	finite-state machine
FST	finite-state transducer
IMPF	imperfective
LOC	locative
MLR	morpholexical rules
NOM	nominative
PCM	position class morphology
PFM	paradigm function morphology
PL	plural

POS	possessive
PRES	present
PRV	pre-root vowel
PST	past
PV	preverb
SG	singular
SIL	Summer Institute of Linguistics
SOV	subject-object-verb
TD	talking dictionary
TS	thematic suffix

CHAPTER 1

INTRODUCTION

1.1 Foreword

"The world has changed. I feel it in the water. I feel it in the earth. I smell it in the air. Much that once was is lost. For none now live who remember it." These famous opening lines from Peter Jackson's 2002 film adaptation of J.R.R. Tolkien's *The Lord of the Rings* are applicable to more than just an epic of fiction—in an ever-modernizing, media-centric world, the global shift to Western ideas and culture has had detrimental effects on minority cultures and languages. The UNESCO Atlas of the World's Languages in Danger estimates that over 43% of the world's approximately 6000 languages are currently endangered (Moseley, 2010), a shocking statistic that would undoubtedly have disturbed the philologist Tolkien himself.

In terms of endangered languages, the rise of technology may act as both a blessing and a curse. On one hand, the internet may serve as a conduit to facilitate the spread of mainstream cultures and languages. This can cause minority cultures to suffer, having the most profound impact on the younger generation. On the other hand, however, technology can be a valuable tool that may present the key to language preservation and revitalization.

1.2 The aim of this thesis

The aim of this thesis is to expand the genre of online lexical resources (specifically, talking dictionaries), in order to maximize practicality for language learners. This

proposed expansion will primarily be at the morphological level, as there is no current consensus on how to address morphological complexity in the talking dictionary genre.

The development of this new morphological model will take place through an examination of Laz (Lazuri), a Caucasian language spoken in the Black Sea region of Turkey. Laz, like many other languages in its family, exhibits a high degree of verbal inflection. It is thus a prime candidate for an examination of morphological complexity, through which a blueprint will be made that can be reproduced and applied to many of the world's languages.

In this thesis, I will bridge several subfields of linguistics, specifically morphology and—at a basic level—computational linguistics. Ultimately, the proposed model for representation of morphological complexity in an online resource will be maximally simplified in order to remain accessible to the largest possible audience.

The challenges facing would-be creators of lexical resources are at times staggering, especially for those who lack programming and web design skills. Although many talking dictionaries already exist and aim to provide members of minority language communities with resources and a valuable link to their linguistic roots in the form of the spoken word, the current status of the talking dictionary (TD) genre is that of an almost completely lexically-driven resource. This comes as no surprise when one considers the overarching 'dictionary' genre, which historically provides one-to-one mapping of form and meaning, giving little attention to syntax or morphology.

However, in the field of language revitalization, resources for developing materials for learning and/or cultural or academic preservation are often limited, and in the interest of functionality and usefulness it follows that any given resource should be designed in order to maximize its potential. Moreover, the nature of the TD should be

accessible to the community it has been designed for, and to a certain extent this accessibility means the encoding of morphological and/or syntactic information into the lexicon. Junker & Stewart (2008, p. 377) posit that "although resources like dictionaries [for indigenous languages] are being compiled, speakers often are at a loss in using them." Although this particular statement refers specifically to Native American languages, the argument may easily be extended to the status of minority languages worldwide.

Difficulties in using a dictionary may stem from a variety of sources, including orthographic variation and/or error, dialectal differences, morphological complexity, and others. The issue of morphological complexity is particularly relevant to the case of Laz as well as many other highly synthetic languages. A speaker attempting to read a text in the target language may wish to look up an unfamiliar word, and if that word happens to be morphologically complex the speaker may be unable to identify any meaning at all. Although parsing and conjugation are not the typical domain of a dictionary, it seems that circumstances in which language resources and materials are lacking (i.e. the case of many endangered and minority languages) that the TD genre may be modified to include more than just a list of simplex forms and empower learners to practically utilize the target language.

Through the following multi-dimensional examination of Laz, important initial steps will be taken to promote the expansion of the talking dictionary genre. Although the project discussed here has not yet been actualized (in terms of creating an operative resource), this thesis will identify important considerations in bridging the gap between form and function. The ultimate result of such considerations is an easier path to the creation of online lexical resources which will strengthen revitalization efforts by

providing maximal efficacy to language learners and members of minority language communities.

1.3 Outline of the thesis

Following a discussion of internet-based language revitalization and of the Laz language, a morphological analysis of Laz will be conducted in order to identify the ideal structural representation for complex verbs in an online lexical resource. The operative theory used in this analysis will be drawn from the tenets of position class morphology (PCM) as set forth in Inkelas (1989). This framework was chosen for its simplicity and applicability—its one-to-one structural mapping may easily be applied in a computational context, and the PCM theory boasts the flexibility to account for the morphology of traditionally challenging templatic languages, which exhibit otherwise inexplicable blocking phenomena. The resulting model from the PCM analysis will identify hierarchical relationships between various affixes and thus demonstrate that a given morphologically complex form (i.e. verbs, in the case of Laz) may be represented both in terms of precedence and dominance.

The thesis will conclude by discussing a practical implementation of the theoretical analysis, considering the efficacy for Laz as well as other languages. The primary computational tool selected for addressing morphological complexity will be finite-state transducer (FST) technology, which will be utilized in designing both a parsing tool and a generating tool for complex forms.

CHAPTER 2

THE TALKING DICTIONARY AS A REVITALIZATION TOOL

2.1 Language endangerment and revitalization

Of the approximately 6,000 languages spoken in the world today, 43% of these can be classified as endangered to some extent. The UNESCO *Atlas of the World's Languages in Danger* provides a breakdown of languages spoken into six categories, ranging from 'safe' (or 'data-deficient') to 'extinct since 1950', with endangerment broken down into 'vulnerable', 'definitely endangered', 'severely endangered', and 'critically endangered' statuses (Moseley, 2010). Causes of this endangerment vary, but these languages share the fate of being ousted from their communities by more dominant languages. Language death is a relatively common phenomenon, as languages may disappear or evolve to the point of unintelligibility (e.g. Latin and Ancient Greek), but never before in history have languages disappeared at their present rate. Globalization and other modern trends have tremendously accelerated the process of language extinction, and currently the twenty most spoken languages in the world are spoken by half of the world's population (Austin & Sallabank, 2011).

When determining what constitutes an endangered language, several criteria should be considered. UNESCO outlines nine different factors to be considered when determining the degree of classification of a given language. These are:

"intergenerational language transmission, absolute number of speakers, proportion of speakers existing within the total (global) population, language use within existing contexts and domains, response to language use in new domains and media, availability of materials for language education and literacy, government and institutional language

policies, community attitudes toward their language, [and] amount and quality of documentation" (Brenzinger, 2010). A small but vigorous language community may be less at risk than a language with thousands of speakers but lacking positive language attitudes and not spoken by children.

Operating under the common assumption that diversity is a strength rather than a weakness, many revitalization attempts are underway to preserve many of the world's endangered languages. While some of these are headed by linguists and consist of language documentation for archival preservation, many projects are headed by community members themselves, who wish to preserve their heritage language, often a key element of cultural identity as well as a tie to history. Methods of revitalization can vary widely and range from various pedagogical websites or lexical resources to more pop culture-centered approaches (e.g. the Navajo dubbing of *Star Wars Episode IV: A New Hope*). These different approaches can in turn have different goals, such as providing learners with tools to facilitate their desire to learn a language in the case of pedagogical approaches, and in other cases trying to create a positive language attitude and thus motivate both learners and speakers to use the target language.

Several language revitalization attempts are well-known throughout the world, the most noteworthy of which may be the case of Hebrew, a formerly extinct language which only existed in written form from around 400 CE until the late nineteenth century, when the spoken form was revived in Palestine over the course of only thirty years through both Hebrew schools and families who agreed to speak the language at home. Today there are five million native Hebrew speakers living in Israel, and the language has been promoted to official status in Israel, along with Arabic (Harshav, 2009). The unique religious significance of Hebrew (and, by extension, of Zionism) to those who

successfully revived it was undoubtedly a driving force behind the movement and the key to its success.

Several other noteworthy revitalization efforts have taken place, most famously those of the Maori language in New Zealand and the Irish (Gaelic) language in Ireland. As to the former, European colonization had negatively impacted Maori language and culture, as English was promoted and often even mandated in society and education. In the 1970s, however, a successful campaign by young Maoris saw the reintroduction of the language into classroom curricula, to great success (Hinton, 2001). In the case of Irish, the language began to fall out of use due to several reasons, including marginalization and famine, and even traditionally Irish-speaking areas called "Gaeltachtaí" saw a decrease in the number of speakers. Revitalization efforts have generally consisted of Irish classroom instruction but were largely unsuccessful until the past decade, which has seen the growth of Irish-medium education and a new Irish-speaking urban demographic (Carnie, 1996; O'Rourke & Walsh, 2015).

Although the attempts mentioned here have been relatively successful ones, as a whole the task of language revitalization is a formidable one, with much potential for error. Often, a multi-pronged approach is necessary in order to target different aspects of revitalization: a revitalization of language attitude may consist of assigning cultural capital to a language in some way, be it artistically or by facilitating its use on social media; a revitalization of fluency, on the other hand, might entail the provision of educational materials or the creation of "language nests" (monolingual home environments which nurture fluency in the target language), among other possibilities. The following section will detail one of the most basic and widespread revitalization

tools, and the chapter will close with an examination of Pazar Laz and how this tool may be applied to the language.

2.2 The talking dictionary

One common tool often employed in language revitalization is the talking dictionary (henceforth TD), an internet-based lexical resource in which users can both read and hear word forms, with the optional inclusion of pictures and other media. TDs are simplistic by design, as they often provide a community-oriented counterpart to the technical grammars arising from linguistic research. While a TD cannot (and should not) provide explicit grammatical information, it is sometimes coupled with other revitalization tools in order to allow users to achieve some level of competence.

One of the most valuable aspects of a TD is its simplicity and user-friendly interface. A TD is often a work in progress, with community members working independently or in collaboration with linguists to create a basic list which can expand to the extent the community desires. Entries can be created without sound files, which can later be added (along with any other desired media). Additionally, a TD is not only limited to words but can also include phrases, sentences, or even narratives. Users can search for target items in either the target language or the majority language, allowing both high-level speakers and those with little to no knowledge of the target language to make use of the dictionary.

A TD in itself is not an inherent agent of change in language attitude, but most importantly it may serve to make an endangered language accessible. If the speech community agrees to the creation of this internet resource, users all over the world can access a given TD, including not only the local community but also speakers living

abroad or even language-learning enthusiasts with no prior connection to the language. Furthermore, TDs allow elder generations to pass down their voices and their language to future generations, thus preserving a small part of their language for all time.

Although there are some common formats and software used to create TDs, there is little standardization and the quality and scope of existing TDs can thus vary widely. Whatever the layout of the site, however, the nature of the material (i.e. lexical items) insures that most TDs have similar features. Several representative examples of TDs come from the Living Tongues Institute for Endangered Languages, a nonprofit organization based in Oregon, USA, and founded by linguists Gregory Anderson and David Harrison. This organization works in conjunction with native communities to create TDs and currently its website has links to 64 TDs, each of which has an identical setup: a basic, black and white website with a search bar, and a speaker symbol next to each lexical entry with a corresponding recording.

Several more stylistic TDs have been created for other languages, such as that of Kanien'keha (Mohawk). This dictionary features additional elements such as a forum for learner to practice and/or ask questions about the language, as well as a language information tab which describes the Mohawk alphabet and provides basic ethnographic data for the language along with some sample texts. Furthermore, this TD also allows users to submit new words to the dictionary, allowing for maximum community involvement in the TD project. As opposed to the simplex, monochromatic format employed by Living Tongues, this TD (created by the Kanien'keha Endangered Language Initiative) offers a more colorful and arguable more aesthetically appealing presentation.

The TD is an excellent starting point for language revitalization in that its creation does not require fluency on the part of a linguist or even a community member. It is a work in progress which can start small and allows for any gaps in knowledge to later be filled in, and the technological fluency required for the creation of a TD has become relatively basic thanks to lexical software such as FLEx and Miromaa,[1] which can store dictionary entries and even export them in desired formats. Despite this ease of creation, however, the TD in its present state remains too simple to successfully represent some complex languages, as will be shown through an examination of the Pazar dialect of Laz in the following section.

2.3 Pazar Laz

Laz, a South Caucasian language from the Black Sea region of Turkey, has been classified by SIL (Summer Institute of Linguistics) as a 6b threatened language, which means that "intergenerational transmission is in the process of being broken, but the child-bearing generation can still use the language so it is possible that revitalization efforts could restore transmission of the language in the home" (Eberhard, Simons, & Fennig, 2019). Figure 1 (taken from SIL's Ethnologue) shows the location of Laz within the language cloud of human languages. Note that as the EGIDS level[2] increases, the status of a language becomes more endangered, and that any language at or above the 6b level is classified as endangered to some degree.

[1] Developed by SIL and ACRA (Arwarbukarl Cultural Research Association, based in Australia), respectively

[2] EGIDS here refers to the Expanded Graded Intergenerational Disruption Scale, a tool for assessing the endangerment of a given language.

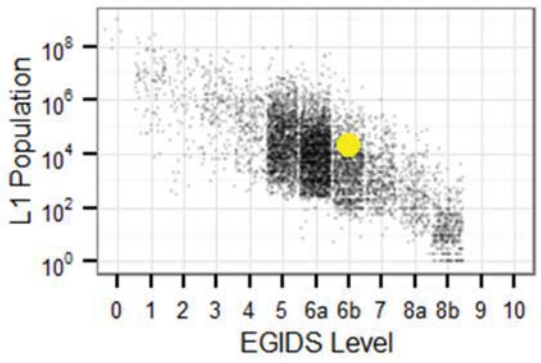


Fig. 1 Position of Laz in world language cloud

The general language attitude of native speakers of endangered Caucasian languages ranges from positive to apathetic. Although speakers seem to be aware that their languages are in decline (Haznedar, 2018), few have taken up the mantle of working to preserve these languages. Additionally, in the case of Laz, emigration from the Black Sea region to other more fiscally-promising areas of the country (such as Istanbul) is common, and so speakers are thus removed from their communities and no longer speak Laz in daily life. The difficulty of intelligibility between Laz dialects presents yet another obstacle, and so the pressure to use the significantly more standardized Turkish language is increased when Laz from different regions communicate with one another. In this way the language is being lost to the mainstream and over the course of time is slowly being overtaken by Turkish.

Revitalization attempts have not been nonexistent, however: in the past few decades, Laz revitalization has seen several new developments, one of the most significant of which was the development of a Latin-based alphabet by Iskender Tzitashi in the Soviet Union in 1929. The final, modern version of the alphabet was developed in

Germany in 1984 by Fahri Lazoğlu, and is still in use today. This alphabet has met with some criticism, as it is based on the Turkish alphabet (with additional diacritics), as opposed to the Georgian alphabet, which is the standard orthography for all other South Caucasian languages (Kavaklı, 2015). Despite the criticism, however, the role of the alphabet was crucial in facilitating the transition from completely spoken language to a written one as well, and thus has made possible many of the revitalization efforts in the years since, such as the body of publications that has emerged from the Laz Institute since its inception in 2013.

Recently, the Laz Institute published an in-depth report on the current state of Laz in Turkey (Haznedar, 2018). This report details a wide variety of factors that have an effect on the language use of individuals from the Laz community (e.g. geographical, socio-economic, age, profession, etc.) and was primarily conducted through the use of surveys. The survey found that although Laz is still the preference of many older participants (bilingual Laz-Turkish speakers), blue collar workers, and denizens of rural settings, the younger generation (specifically individuals under 25 years of age) as well as students and those living in urban settings express an overwhelming preference for Turkish. Regarding language attitudes, however, the majority of participants expressed positive attitudes towards Laz and a knowledge that the language is endangered.

Some notable revitalization efforts directed at Laz include media (and social media), music, and educational materials. The first Laz magazine was published in 1993 and was called *Ogni* (meaning “did you hear/understand”). Laz internet presence has also expanded recently, and to date there are several organizations with websites and Facebook pages which promote Laz language, music, and culture. At an educational level, Laz courses were opened at Istanbul’s Boğaziçi University in 2011 and have been

ongoing, and primary school level educational material has been developed by Laz native speaker and activist İsmail Avcı Bucaklişi which is available on the Turkish Ministry of Education's official website (Haznedar, 2018; CIDLES, 2013).

Several dictionaries have been developed for Laz in both printed and internet-based formats, but a TD is not among these. The ability to hear forms pronounced would undoubtedly be beneficial to native speakers of Turkish, English, or any other non-Caucasian language who would be understandably intimidated by the consonant clusters which characterize many Laz (and other Caucasian) words. This benefit, combined with the relative ease of creating a TD, would seem to make Laz an ideal candidate for this type of resource, and to a certain extent this is the case. However, past the realm of nouns and adjectives, structurally-complex and often highly irregular Laz verbs create difficulties for the development of a lexical resource, and no TD to date has attempted to deal with these complexities (the nature of which will be discussed in Chapter 3).

To summarize, the role of a TD is to provide a lexical resource for learners of (primarily endangered) languages that allow these learners to receive multimedia input and facilitate the use and therefore continuity of the language. Despite TDs' emphasis on simplicity, it seems that valuable adaptations could be made to account for various linguistic complexities to the end goal of facilitating language use. The task remains, then, to incorporate complexity into simplicity and find creative solutions to representing difficult forms in a way that is both intuitive and user-friendly. If this approach can be applied to Laz and at least partially account for complex forms, the resulting adaptations may be implemented in the TDs of other languages with similar phenomena, and in this way the most basic of revitalization resources will have a more holistic role in bringing learners to a higher level of linguistic competence.

CHAPTER 3

LAZURI—AN OVERVIEW

The Laz people are an ethnic minority group native to the Black Sea region of northeastern Turkey and southwestern Georgia, with the vast majority of the population located in the former. How vast this majority actually is, however, is unclear, as estimates of the Laz population in Turkey vary widely. An estimate by the Summer Institute of Linguistics (SIL) places the ethnic population at 92,000 and the speaker population at 20,000 in Turkey, with an additional 2,000 speakers in Georgia. The reason for this lack of clarity is the census policies of Turkey, which do not include ethnic information for minority groups. The following chapter will provide a brief historical overview of the Laz population in Turkey, note the current status of the people and language, and provide a brief grammatical sketch of Pazar Laz as well as a literature review of relevant work on the language.

3.1 History of Laz people and language

Even though they were not counted in Turkish censuses, the Laz population has been present in the Black Sea region for over two millennia. Founded during the Roman Empire, the territory of Lazica (formerly Cholchis) stretched from the center of present-day Georgia northward to the Caucasus Mountains and westward to the Black Sea, with a small strip of land situated in what is today northeastern Turkey. Lazica was, during the Byzantine Empire, quickly converted to Christianity, but Persian advances on the strategically located southern Black Sea coast (i.e. present-day Turkey) also exposed the region to Islam between the sixth and tenth centuries. Later, in the mid-15th century,

Lazica was conquered by the Ottomans under the leadership of Mehmet II, during which time the Laz people primarily converted to Islam. Apart from a small Christian Laz population in Georgia, Islam continues to be the primary religion of the Laz in Turkey ("Laz", n.d.).

During the years of 1551 to 1925, the region became known as the Lazistan Sancağı (administrative region) of the Ottoman Empire, with Batumi (in present-day Georgia) serving as the capital until it was acquired by Russia in 1878, at which time the region was divided and the capital shifted to Rize (Turkey). During the era of the Lazistan Sancağı, the Laz enjoyed relative autonomy until the administrative status of the region was abolished by the Republic of Turkey in 1925, at which time the term "Lazistan" was officially banned (Minahan, 2016).

Despite the significant differences between Laz and other Caucasian people groups in terms of nationality and religion, the language still retains strong ties to other Kartvelian (i.e. South Caucasian) groups. A further distinguishing factor of the Laz people is their language. The Laz language (also known as Lazuri) belongs to the Kartvelian language family, which also includes Mingrelian, Svan, and most famously, Georgian. Of these languages, Laz and Mingrelian are most closely related, as they are divergent dialects of the now-extinct Zan language. Although Laz and Mingrelian share many similarities, however, they are no longer mutually intelligible and thus cannot be classified as dialects of one language. Furthermore, Mingrelian speakers are primarily found in Georgia, and the geopolitical separation from Laz has facilitated the growing gap between these two languages over the past five centuries (Vamling & Tchantouria, 2005).

3.2 Current status

Within the Laz language, there are several distinct dialects, corresponding to several localities in the Black Sea region of Turkey: in the city of Hopa, Xopuri; in Arhavi and Fındıklı, Vits'ur-Arkabuli; and in Pazar and Ardeşen, Atinuri. This third dialect is also known as Pazar Laz and will be the primary focus of the following study. As the language has no official status and standardization is still slowly occurring, dialects even less than a hundred miles apart can have significant differences, both phonologically and morphologically. The landscape of the region may also be a contributing factor to both the divergence of dialects as well as the continued existence of Laz to present day: the southwest coast of the Black Sea is characterized by mountains (the foothills of the Caucasus Mountain chain) and dense forests. The Laz enjoyed relative autonomy throughout most of their existence, but Turkish-medium education policies along with expansion of agricultural production in the area has ensured that virtually all Laz are bilingual (or in many cases monolingual) speakers of Turkish (Kutscher, 2008).

Understandably, the number of speakers of Laz is significantly less than that of the population of the overall ethnic group due to this shift towards Turkish, the only official language of Turkey, especially on the part of the younger generation. As mentioned earlier, SIL estimates the total number of Laz speakers to be 22,000 (Lewis & Simons, 2013), but most estimates cite a higher number than this, such as a 2006 estimate by Turkish research and consultancy company KONDA, which placed the number of native speakers at 0.12% of the total population (totaling 82,356 people), with another 0.09% (61,767 people) using the language in daily life (KONDA, 2006). Whether 22,000 or 82,000, however, Laz speakers form a small linguistic minority in Turkey, and as technology and other factors have an ever-increasing impact on the

world, the resulting trend towards national or even global standards has been detrimental to the continuation of the Laz language. Laz has thus been classified by SIL's Ethnologue as a 6b (threatened) language. Revitalization efforts have gained momentum in the past several years, as have been described in Chapter 2.

3.3 Grammatical sketch

Before turning to a morphological analysis of Pazar Laz, a brief grammatical sketch of the dialect is in order. We will now briefly consider its key phonological, syntactic, and morphological characteristics.

3.3.1 Phonology

Like many other Caucasian languages, Laz possesses a relatively simple five-vowel system[3] but a rich inventory of consonants, which demonstrate a large number of combinatorial possibilities. These phonemes are represented in Table 1, condensed from Öztürk & Pöchtrager (2011).

Table 1. Laz Consonants

	Labial	Alveolar	Post-alveolar	Velar	Labio-velar	Glottal
stops (+/- voice)	[b]/[p^h]	[d]/[t^h]		[g]/[k^h]	[gw]/[k^{h}w]	
ejective stops	[p']	[t']		[k']	[k'w]	
affricates (+/- voice)		[dz]/[ts^h]	[dʒ]/[tʃ]			
ejective affricates		[ts']	[tʃ']			
fricatives (+/- voice)	[f]	[z]/[s]	[ʒ]/[ʃ]	[ɣ]/[x]	[ɣw]/[xw]	
nasals	[m]	[n]				
liquids		[l], [r]				
approximants			[j]		[w]	[h]

[3] Front vowels: [i] and [u]; mid vowels: [e] and [o]; back vowel: [a]

In morphologically simplex words, the combinatorial possibilities of consonants can, depending on the analysis, be posited to consist of up to four members within a single consonant cluster,[4] such as the case of the word *mskva*, or 'beautiful', the combinations of which are laid out in Öztürk & Pöchtrager (2011). Morphologically complex forms, however, can present even more difficult combinations for analysis, as in the case of *p't'k'vi*, or 'I said'.[5]

Aside from the various Laz phonemes and their combinatorial possibilities, several other noteworthy phenomena characterize the language. Firstly, virtually all uninflected Laz lexemes end with vowels, and words borrowed into the language (many of which are from Turkish) are thus modified, as we can see from Laz *dersi,* 'lesson', adapted from the Turkish form *ders.* Another phenomenon, the allophonic variation of several affixes, extends into the realm of morphophonology, as demonstrated by (1) and (2) below.

(1) Purely phonological:[6]

B-ğar-um	Mp'oli-s p-skid-ur
1-cry-TS	Istanbul-LOC 1-live-TS
'I cry'	'I live in Istanbul'

(2) Morphophonological:

Xe do p'ici v-i-bon-am	Layç'i m-a-şkur-in-e-n
hand and face 1-PRV-wash-TS	dog 1-PRV-fear-CAUS-TS-3SG
'I wash my hands and face'	'I am afraid of dogs'

[4] Öztürk & Pöchtrager (2011) analyze four-consonant clusters as consisting of three consonants, grouping the third and fourth consonants together as labiovelars

[5] Examples drawn from Öztürk & Pöchtrager (2011)

[6] Abbreviations: 1 (1st person), TS (thematic suffix), LOC (locative), PRV (pre-root vowel), CAUS (causative), 3SG (3rd person singular)

As we see in (1), the 1SG prefixes in the inflected forms *bğarum* and *pskidur* (allophones *b-* and *p-*, respectively) demonstrate agreement in voicing with the initial phoneme of the immediately following verb root (i.e. voiced when occurring before the root *ğar* and devoiced before *skid*). Additionally, in cases of voicelessness the feature of +/- ejective determines whether the allophone will manifest as (*p-* or *p'-*).[7] In (2), however, the distinction between *v-* and *m-* cannot be captured by phonology alone, and represents allomorphic rather than allophonic variation.[8]

3.3.2 Syntax

Syntactically, Laz is an SOV language and has been the focus of a significant number of syntax-driven studies, including morphosyntactic analyses. Let us begin with a brief examination of phrase and sentence-level phenomena before honing in on morphology.

3.3.2.1 Copula

Laz allows for several types of non-verbal sentences (nominal, pronominal, adjectival, and existential) formed with the use of copula:

(3)	Ma	vorsi	v-ore
	1.SG	good	1-COP.PRS
	'I am good/fine'		

(4)	Si	hak	ort'-i
	2.SG	here	COP.PST
	'You were here'		

(5)	Coğori	skani	on
	dog	2.SG.POSS	COP.3.PRS
	'The dog is yours'		

(6)	Xopa-s	Laz-epe	or-an
	Hopa-LOC	Laz-PL	COP-PL
	'There are Laz (people) in Hopa'		

[7] For non-ejective and ejective, respectively.
[8] Rather, the nature of the verbal argument dictates which allophone will appear: for agent-like arguments or arguments of intransitive verbs, a verb will choose the *v-* allophone

As shown above, in (3) we see an adjectival non-verbal sentence, in (4) an adverbial, in (5) a nominal, and in (6) an existential. The copula in each of these contexts may be conjugated for person as well as tense.

3.3.2.2 Question formation

Question formation in Laz includes polar questions, which are formed with the clitic *i*:

(7) Andğa tağdili on-i?
Today holiday COP.3.PRES-Q
'Is today a holiday?'

Word order is not affected in questions—rather, in polar questions, whether affirmative or negative,[9] the question-marking clitic *-i* attaches to final word of the clause (i.e. the predicate or the copula).

Questions can also be formed by using wh-words, which occur in the same positions as the nouns and/or adverbs they replace. Additionally, it is possible to have two wh-words in a single question:

(8) Biç'i-k mu-peri pont'uli mo-y-dum-s?
Boy-ERG what-kind pants PV2-PRV-wear-3.SG
'What kind of pants is the boy wearing?'

(9) Mi-k mu-peri pont'uli mo-y-dum-s?
Who-ERG what-kind pants PV2-PRV-wear-3.SG
'Who is wearing what kind of pants?'

[9] In negative polar questions, the negative marker *var* appears immediately preceding the predicate/copula.

3.3.2.3 Complex sentences

Laz also employs various techniques to form a variety of complex sentence patterns, including coordination (at word, phrasal, and clausal levels), embedded clauses, relative clauses, and adverbial clauses. Detailed explanations of these structures along with many examples can be found in Öztürk & Pöchtrager (2011).

3.3.3 Morphology

Moving on from sentence and phrase level syntax, the primary arena in which Laz (along with other Caucasian languages) demonstrates complex phenomena is in the realm of morphology. Laz possesses a rich verbal morphology which includes double person marking on transitive (and ditransitive) predicates:

(10) Ma Ali b-dzir-em-Ø
1.SG Ali-NOM 1-see-TS-1.PRS
'I see Ali'

(11) Ali-k ma m-dzir-em-s
Ali-ERG 1.SG 1-see-TS-3.SG
'Ali sees me'

(12) Ma si himus g-o-tzir-i
1.SG 2.SG 3SG.DAT 2-APPL-show-1.PST
'I showed you to him/her'

In these examples we see that person markers occur both before and after the root. In (14), the person markers only appear to reflect one of the arguments (i.e. the first person subject), whereas in (15) we see that both the third person subject and first person

object are marked on the verb. The explanation for this is a person marking hierarchy, which is exemplified in a ditransitive verb such as *-tsir-* 'show' in (16), as only two of the three arguments may be marked on the verb. This hierarchy will be examined in greater detail in the following analysis.

Continuing in this morphological vein, it becomes clear that in addition to allowing complex clusters of consonants, in some words (particularly verbs) we can see complex clusters of morphemes. A single verb root can take up to 15 different types of affix (although all fifteen never manifest simultaneously). A diagram of verbal slots posited by Öztürk & Pöchtrager (2011) is reproduced in Table 2.

Table 2. Laz Verbal Complex

1	pre-verb (affirmative)	[PV1]	affirmative particles
2	pre-verb (spatial)	[PV2]	denote manner/direction
3	person marker	[1], [2], [3]	first of two obligatory person markers
4	valency marker	[PRV] or [APL]	pre-root vowels related to valency of argument
5	ROOT		obligatory root
6	augmentative	[AUG]	augmentative stem formant
7	causative (intransitive); impersonal passive	[CAUS] or [IP]	causative or impersonal passive marker {-in}
8	causative (transitive)	[CAUS]	causative suffix for transitives {-ap}
9	causative (perfect)	[EM]	causative suffix for perfect construction {-ap}
10	imperfective thematic suffix	[TS]	imperfective, based on theta role: {-am}, {-um}, {-ur}, {-er}
11	copula	[COP]	auxiliary stem formant {-t'}
12	subjunctive	[SUB]	subjunctive marker {-a}
13	person marker/tense	[1], [2], [3]	second of two person markers +/- past tense
14	conditional	[COND]	conditional marker {-k'o}
15	plurality	[PL]	plurality marker {-t} (may be fused with third person)
16	auxiliary	[AUX]	auxiliaries {(e)re, donu, ert'u, dort'u}

In Table 2, the 15 verbal affixes are listed in the order in which they appear on a given verb, although not all slots may be filled simultaneously. However, heavily inflected forms are also possible, as in the following examples in (13):

(13)

a) p’-ç’ar-um
   1-write-TS
   ‘I write/I am writing’

b) p’-ç’ar-a-t-ere
   1-write-SUBJ-PL-AUX
   ‘We will write it’

c) m-i-ç’ar-ap-ur-t’-u
   1-PRV-write-CAUS$_{PRF}$-TS-IMPF-PST
   ‘I had written it before’

d) m-i-ç’ar-ap-ur-t’-a-s-ere
   1-PRV-write-CAUS$_{PRF}$-TS-IMPF-SUBJ-3-AUX
   ‘I will have written it’

These examples illustrate a variety of possible verbal conjugations with various person marking, tense, and aspect (in this case, all inflected forms containing the root *-ç’ar-* ‘write’). Extensive inflection is not mandatory, such as in the case of simple predicates such as *p’ç’arum* ‘I write/I am writing’, but is possible in more semantically complex forms such as the future perfect *miç’arapurt’asere* ‘I will have written it’.

Chapter 5 will delve deeper into the verbal affix categories and their combinatorial possibilities. Concerning substantives, however, the puzzle contains fewer

pieces, as inflectional suffixes are restricted to plurality and case, as demonstrated in (14) and (15), drawn from Öztürk & Pöchtrager (2011):

(14) Xordza-lepe-k bere-pe-s d-u-bon-u
woman-PL-ERG child-PL-DAT PV-APL-wash-PST
'The women washed the children.'

(15) Mskva-k ini mjalva ş-um-s
beautiful-ERG cold milk drink-TS-3
'The beautiful one is drinking cold milk.'

In (14), we observe both plurality and case markers on both nouns in the sentence, with subject and object receiving ergative and dative case, respectively. (15) demonstrates inflection of adjectives—the subject of the sentence is a derived nominal and is therefore able to receive a case marker, whereas the pure adjective *ini* ('cold') remains uninflected.

Having illustrated some of the important features of Laz, we now move on to a brief review of existing literature on the language. A far more comprehensive overview of the language can be found in Öztürk & Pöchtrager (2011).

3.4 Previous work on Laz

Although Laz entered the radar of academia in the nineteenth century with a descriptive text by Rosen (1844), it has not been the subject of intensive study until recently. Other Laz grammars include those of Anderson (1963), Kojima & Bucaklişi (2003), Lacroix

(2009),[10] and most recently a grammar of the Pazar dialect of Laz by Öztürk & Pöchtrager (2011). The latter is the result of field research facilitated by native informant İsmail Avcı Bucaklişi, co-author of an earlier grammar (Kojima & Bucaklişi, 2003). Other descriptive works include Tuite (1988), Holisky (1991), Kutscher et al. (1995), and Kutscher (2001, 2008).

In the past two decades a relatively large volume of literature has emerged from Boğaziçi University, stemming from the logistical advantage of the university and its interest in Caucasian languages and linguistics.[11] Research has been conducted on Laz syntax in multiple research projects funded by Boğaziçi University (Öztürk, 2012, 2007) and several MA theses have been written on various aspects of Laz: on its case system, Gürpınar (2000); on finiteness and complementation patterns, Emgin (2009); a sociological study on Laz identity, Taşkın (2011); on the AGREE model, Demirok (2013); and on directional prefixes, Eren (2016). Furthermore, the Lazca Sözlük Projesi ('Laz Dictionary Project'), a trilingual (i.e. Laz, Turkish, and English) online lexical database, has been developed (Kontovas, 2015).

Much of the research to date on Laz has been of a morphosyntactic nature. Aside from the creation of a grammar, little weight has been given to academic research that can serve as an aid in the revitalization of Laz. This is a shortcoming that needs to be addressed, not only in the case of Laz but in that of many other endangered or threatened languages as well. The following data analysis will incorporate theoretical elements but will remain functional in nature, keeping in mind the ultimate goal of the project.

[10] A description of the Ark'abi dialect

[11] In addition to the analytical literature mentioned here, note that other work on Laz has been published, including multiple language learning textbooks and a 2018 report on the current state of Laz. These have already been mentioned in the previous chapter.

CHAPTER 4

METHODOLOGY

Having introduced the Laz language and the importance of talking dictionaries in language revitalization, we now turn to view Laz through a more analytical lens, identifying and examining its structure and determining how morphological phenomena may be accounted for in a web-based resource. This chapter will detail the selection of an appropriate model to conduct an analysis of Laz morphology, ultimately going on to provide background information and a brief overview of the model.

4.1 Selecting a framework for morphological analysis

In selecting a theoretical framework to employ in the analysis of Laz morphological phenomena, the practical goal (i.e. a talking dictionary) provided the primary criteria for the choice of lens through which the data will be examined. Thus, several well-known frameworks such as Distributed Morphology (Halle & Marantz, 1993) were dismissed due to their highly technical natures and resulting inaccessibility to language learners and existing lexical technology.

Similarly, much previous analysis of Laz, such as is exemplified in Demirok (2013), is highly syntax-driven and draws heavily from the tenets of generative grammar (Chomsky, 1965). As such, these analyses account for the various morphological phenomena of Laz through the application of phrase-structure, subcategorization, and transformational rules, which account for various syntactic phenomena such as movement, binding, etc. Ultimately, these syntactic features introduced an additional layer of complexity and would require a correspondingly complex computational tool.

Thus, a more purely morphological approach was prioritized in an effort to be more compatible with a wider variety of computational tools.

Much research on Laz to date has been, fittingly, of a morphosyntactic nature,[12] as Laz is a highly inflectional language—many semantic variables may be expressed as affixes on a verbal root, and various processes summarized in Öztürk & Pöchtrager (2011) can be observed to occur within a single stem. Other Caucasian languages with similar features have been analyzed through still other approaches, such as the study of Georgian by Gurevich (2006) by means of the framework of constructional morphology. The constructional method provides building blocks for forms by representing various aspects, such as semantic, syntactic, and morphological information, in "generalized form-meaning pairings" (Gurevich, 2006). The inclusion of these multiple elements is highly applicable to Laz as well, as it is closely related to Georgian and likewise possesses a rich morphology.

Despite its merits, constructional morphology was not selected for the following analysis. Although it has the necessary scope to address phenomena of a wide typological variety of languages, another framework was identified which can more easily interact with a potential lexical software, cementing its role in this analysis. For this reason, the lens selected for the following examination of Laz was that of position class morphology (PCM), as the one-to-one structural mapping allows for an examination of the verbal template which may easily be applied computationally and boasts the flexibility to account for the morphology of traditionally challenging templatic languages, which exhibit strict linear ordering that is independent of syntax,

[12] See literature review in Chapter 3

semantics, and other factors. Furthermore, a previous analysis has already demonstrated Laz to be compatible with position class morphology (Atlamaz, 2013). The following section will provide a brief overview of this framework, including its origins and major tenets.

4.2 An overview of position class morphology

Although the notion of position classes or "slot-filler" morphology dates back to the 1950s (e.g. Hockett, 1958), the tenets of the position class morphology (PCM) framework to be used in the following analysis were first introduced in a study of position classes in Nimboran by Inkelas (1989), drawing from the framework of lexical morphology as described in Kiparsky (1982). As posited by Inkelas (1989), PCM offers insight into morphological complexity by identifying both linear and hierarchical relationships among affixes. Position classes have been incorporated into other morphological frameworks as well, most notably paradigm function morphology (PFM), popularized by Stump (2001, p. 139), who states that "the position classes into which the inflectional affixes can be sorted often reveal the organization and interaction of rule blocks[13] with complete transparency."

Although PFM, which seeks to represent morphological structure with an ordered series of realization rules mapping features to forms, could also be used to analyze highly inflectional languages like Laz, its more rigid style of formalization would require a higher level of technical knowledge in order to both conduct an analysis and implement the results computationally. The following partial PFM analysis of

[13] 'Rule block' in PFM refers to a grouping of realization rules which dictate surface form.

Swahili verbs (Stump, 2007) illustrates the typical rule blocking formalization present in PFM:

(1) Swahili Rule Schema (Stump, 2007, p. 166)

A.	*Morpholexical rules for slot I*	
A1.	$MLR_{I,[AGR(ob):[PERS:2, NUM:sg, GEN:m/wa]]}([_V\ x])$	$=_{def}\ [_V \text{ ku } [_V\ x]]$
A2.	$MLR_{I,[AGR(ob):[PERS:3, NUM:sg, GEN:m/wa]]}([_V\ x])$	$=_{def}\ [_V \text{ m } [_V\ x]]$
A3.	$MLR_{I,[AGR(ob):[PERS:2, NUM:pl, GEN:m/wa]]}([_V\ x])$	$=_{def}\ [_V \text{ wa } [_V\ x]]$
B.	*Morpholexical rules for slot II* (in which μ = su or ob)	
B1.	$MLR_{II,[REL(\mu):[NUM:sg, GEN:m/wa]]}([_V\ x])$	$=_{def}\ [_V \text{ ye } [_V\ x]]$
B2.	$MLR_{II,[REL(\mu):[NUM:sg, GEN:ki/vi]]}([_V\ x])$	$=_{def}\ [_V \text{ cho } [_V\ x]]$
B3.	$MLR_{II,[REL(\mu):[NUM:pl, GEN:m/wa]]}([_V\ x])$	$=_{def}\ [_V \text{ o } [_V\ x]]$
B4.	$MLR_{II,[REL(\mu):[NUM:pl, GEN:ki/vi]]}([_V\ x])$	$=_{def}\ [_V \text{ vyo } [_V\ x]]$
C.	*Morpholexical rules for slot III* (in which μ = su or ob)	
C1.	$MLR_{III,[TNS:definite]}([_V\ x])$	$=_{def}\ [_V \text{ na } [_V\ x]]$
C2.	$MLR_{III,[TNS:future]}([_V\ x])$	$=_{def}\ [_V \text{ ta } [_V\ x]]$
C3.	$MLR_{III,[TNS:future, REL(\mu):[NUM:\alpha, GEN:\beta]]}([_V\ x])$	$=_{def}\ [_V \text{ taka } [_V\ x]]$
C4.	$MLR_{III,[TNS:past]}([_V\ x])$	$=_{def}\ [_V \text{ li } [_V\ x]]$
C5.	$MLR_{III,[POL:neg, TNS:past]}([_V\ x])$	$=_{def}\ [_V \text{ ku } [_V\ x]]$
C6.	$MLR_{III,[POL:neg, REL(\mu):[NUM:\alpha, GEN:\beta]]}([_V\ x])$	$=_{def}\ [_V \text{ si } [_V\ x]]$

In the rule schema in (1), morpholexical rules (MLR) are laid out in order of precedence, and rule-ordering occurs within each slot. This type of system could prove promising for dealing with morphological complexity computational way, but, as previously noted, the level of technical knowledge required to perform a PFM analysis would add further difficulty to the process of creating a TD, especially for native speakers or language activists who do not possess a strong foundation of linguistic knowledge. The relative simplicity of PCM was therefore the deciding factor in its selection for the analysis of these data.

In PCM, inflectional affixes are grouped based on their positions relative to the root as well as by their distribution (i.e. two affixes occurring in the same location which are in complementary distribution would belong to the same position class). Resulting

representations could then be applied in conjunction with other frameworks, such as PFM (Stump, 2001) or distributed morphology as in Atlamaz (2013).

Although originally position classes were ranked according to linear precedence alone, Inkelas (1989) introduced a hierarchical structure which relates linear order to relative dominance, a system which draws heavily from the idea of stratum ordering in lexical morphology. A heavily suffixing language may then be represented as in the Figure 2, reproduced from Inkelas (1993):[14]

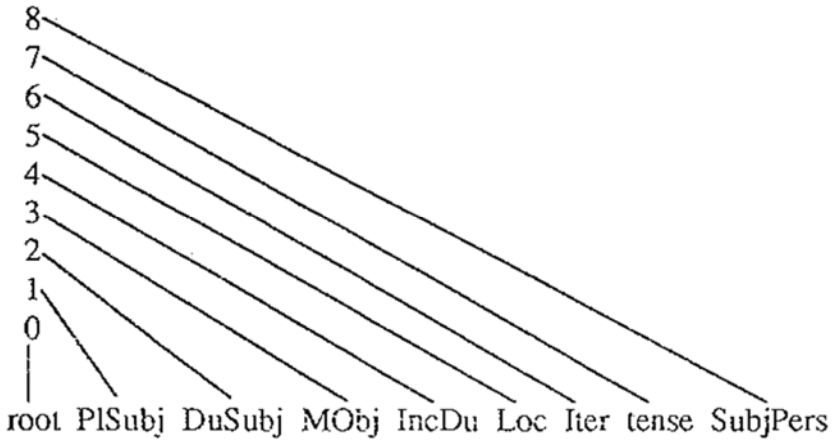


Fig. 2 Precedence and dominance-based hierarchy (Inkelas, 1993, p. 582)

The creation of this type of hierarchy was necessitated by morphological (as opposed to semantic) blocking phenomena in Nimboran that were otherwise inexplicable. The crucial evidence necessitating a two-dimensional representation comes from the behavior of particles in terms of blocking. An example of this behavior is reproduced in (2):

[14] Note that it is not mandatory for the root to appear initially in the verbal template

(2) Blocking Behavior of Particles (Inkelas, 1993, p. 593)

(63)a.	Blocks position 4	*-maN*[+ *A*]-
b.	Blocks positions 2 and 4	*-damaN*[+ *A*]-
c.	Blocks position 2	*-dáN*[+ *A*]-
d.	Blocks positions 5 and 6	*-N*-
e.	Blocks positions 2, 5 and 6	*-náN*[+ *A*]-
f.	Blocks positions 2 and 5	*-ta*[+ *A*]-

In (2), the blocking effects of the various particles (in the righthand column) behave unpredictably. For instance, in B, E, and F above, the blocking effects of the respective particles do not follow a linear pattern. Furthermore, the behavior of the position 4 particle, the inclusive dual subject (IncDuSubj) marker *-maN*, is shown to pattern only with its preceding affixes in classes 2 and 3, rather than the following ones. This behavior cannot be explained by the one-dimensional model reproduced in Figure 3:

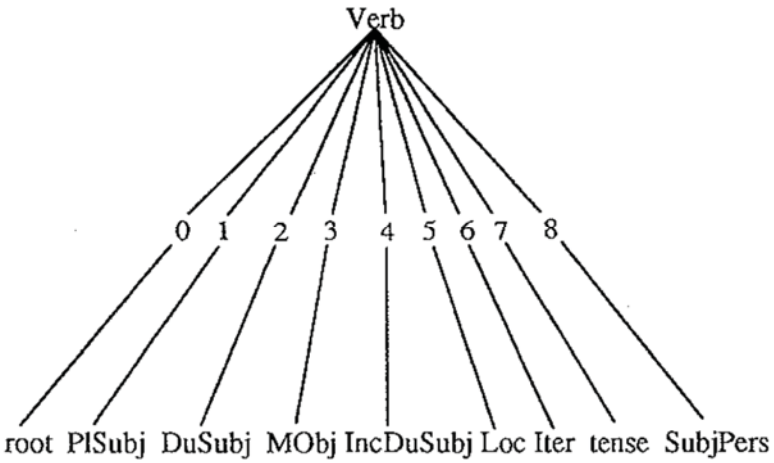


Fig. 3 Precedence-based model (Inkelas, 1993, p. 582)

In order to determine relative dominance, the key lies in the interactions between morphemes which can be ascertained from the phenomena of blocking. Blocking may be 'featural blocking' of a semantic nature (e.g. a past tense marker blocking a present tense one) or more importantly it may be 'morphological blocking', which is not semantically motivated.

Morphological blocking may be symmetrical, in which case two morphemes within the same position class are in complementary distribution (despite their semantic compatibility)—in Nimboran, one instance of symmetrical blocking is that the dual subject marker may not occur with the plural object marker (both of which are in position class 2). Another type of blocking is asymmetrical blocking, in which case a morpheme from one position class blocks one or more morphemes from other position classes. Nimboran also exhibits asymmetrical blocking, such as the durative (Dur) marker from position 3 blocking the position 2 dual subject (DuSubj) marker:

(3) Asymmetrical Blocking in Nimboran (Inkelas, 1993, p. 586)

a. Du. Subj:	ŋgedóu-ke-t-u draw-DuSubj-Pres-1	→	ŋgedóuketú 'We two draw (here)'	p. 186
b. Pl. Subj:	ŋgedói-⟨i⟩-t-u draw.pl-⟨PlSubj⟩-Pres-1	→	ŋgedóitiu 'We (many) draw (here)'	p. 186
c. Du/Pl. Subj, Dur:	ŋgedói-⟨i⟩-tam[+A]-t-u draw.pl-⟨PlSubj⟩-Dur-Pres-1	→	ŋgedóitiemtɨ́ 'We are drawing (here)'	p. 234

Once the relative dominance has been established, the verbal template may be expressed using a lexical hierarchy in which precedence is represented on the horizontal

axis and dominance on the vertical. Figure 4 shows the lexical hierarchy of Nimboran verbs as established by Inkelas (1993):

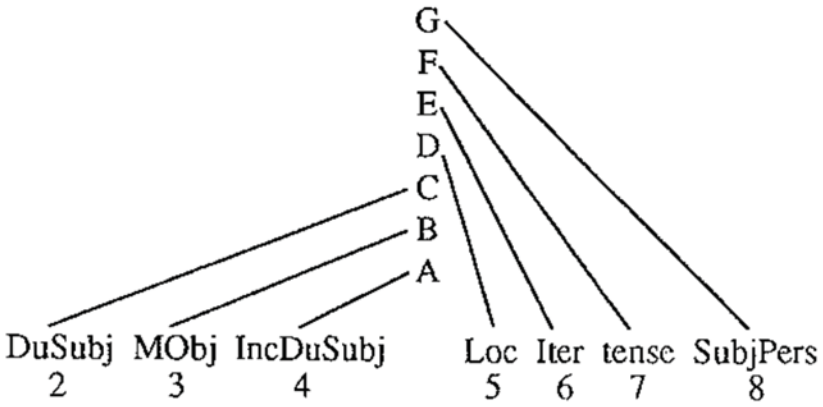
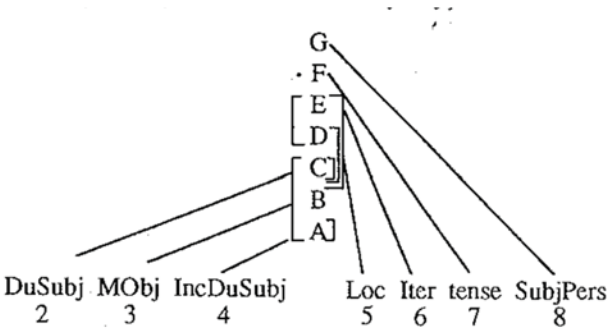


Fig. 4 Nimboran affixal hierarchy in verbs (Inkelas, 1993, p. 597)

Figure 4 shows a variety of Nimboran verbal affixes appearing in linear order (with the verb root occupying the first position) and position classes 4 and 8 being the most and least dominant, respectively, in terms of their ability to block other morphemes. As we can see, the blocking phenomenon mentioned above (i.e. that position 4 is immune to effects of blocking from rightward positions) can be explained through this model. Although positions 2 and 5 are not adjacent in terms of precedence, they are adjacent in terms of dominance (levels C and D, respectively). It should be noted that lines linking precedence to dominance must never cross, as such a structure would be considered ill-formed.

Not only does this model solve the problem of position classes 2 and 5 in Nimboran, but it may also explain multiple other cases of particle blocking, as shown in Figure 5:

a)



b)

Positions (levels) blocked	Examples
a. Blocks A	$[maN[+\mathbf{A}]\ [\ \ ]_{A}]_{B}$
b. Blocks A–C	$[damaN[+A]\ [\ \ ]_{A}]_{D}$
c. Blocks C	$[dáN[+A]\ [\ \ ]_{C}]_{D}$
d. Blocks D–E	$[N\ [\ \ ]_{D}]_{F}$
e. Blocks C–E	$[náN[+A]\ [\ \ ]_{C}]_{F}$
f. Blocks C–D	$[ta[+A]\ [\ \ ]_{C}]_{E}$

Fig. 5 Mass blocking in Nimboran (Inkelas, 1993, p. 601)

As we may observe in the representation in Figure 5, in all known instances of particle blocking in Nimboran (designated by brackets) which involve multiple position classes there is vertical adjacency, regardless of whether or not there is also linear adjacency. The flexibility to deal with a variety of blocking phenomena in order to capture the largest amount of generalizations possible is one of the most convincing reasons to choose PCM as a lens through which to examine a highly inflectional language such as Laz. Although Laz and Nimboran do not demonstrate identical phenomena, the flexibility afforded by PCM in dealing with a wide variety of concatenative blocking phenomena greatly increases its appeal as an analytical tool.

In the following analysis, data from Laz will be examined in the light of PCM and various hierarchical relationships among affixes (primarily in verbal paradigms) will be analyzed in an attempt to shed light on some of the more complex forms in the language. The hierarchical structure of Laz position classes will be identified as well as any patterns of blocking, and these findings will be utilized in the creation of a technological resource.

The following data were gathered from a variety of sources, including the grammar of Laz by Öztürk & Pöchtrager (2011) as well as notes taken in beginner and intermediate Laz courses at Boğaziçi University during the 2015-16 and 2016-17 academic years. Additional data came from interactions with a native speaker of Pazar Laz, language activist and educator İsmail Avcı Bucaklişi, who provided original data as well as grammaticality judgments.

CHAPTER 5

ANALYSIS

In constructing a lexical resource for Laz, firstly the organization and storage of the lexical items must be determined, and this in turn necessitates an examination of complex Laz roots in an effort to capture the morphological phenomena within the confines of a theoretical framework. As the verb is the part of speech in Laz which demonstrates the maximal amount of inflection, this chapter will focus solely on Laz verbal morphology. The conclusions of this analysis will lay the groundwork for the resulting selection of a computational model for the development of a lexical resource in Chapter 6.

5.1 Laz verbal morphology

Looking to the Laz verbal template set forth in Öztürk & Pöchtrager (2011), the situation is somewhat more complex than that of nouns, adjectives, and adverbs, which demonstrate minimal inflection, as mentioned in Chapter 3. The 16 listed inflectional verbal morphemes (including the root) may never all be realized simultaneously but nevertheless possess rich combinatorial possibilities. In order to gain a general understanding of the affixal slots in an inflected Laz verb, we may first consider the preexisting template, reproduced below in Table 3 (although we will go on to posit significant changes to this model in the following analysis).

Table 3. Laz Verbal Template (Öztürk & Pöchtrager, 2011)

1	pre-verb (affirmative)	[PV1]	affirmative particles
2	pre-verb (spatial)	[PV2]	denote manner/direction
3	person marker	[1], [2], [3]	first of two obligatory person markers
4	valency marker	[PRV], [APL]	pre-root vowels related to valency of argument
5	ROOT		obligatory root
6	augmentative	[AUG]	augmentative stem formant
7	causative (intransitive); impersonal passive	[CAUS], [IP]	causative or impersonal passive marker {-in}
8	causative (transitive)	[CAUS]	causative suffix for transitives {-ap}
9	causative (perfect)	[EM]	causative suffix for perfect construction {-ap}
10	thematic suffix	[TS]	based on theta role: {-am}, {-um}, {-ur}, {-er}
11	copula	[COP]	auxiliary stem formant {-t'}
12	subjunctive	[SUBJ]	subjunctive marker {-a}
13	person marker/tense	[1], [2], [3]	second of two person markers +/- past tense
14	conditional	[COND]	conditional marker {-k'o}
15	plurality	[PL]	plurality marker {-t} (may be fused with third person)
16	auxiliary	[AUX]	auxiliaries {(e)re, donu, ert'u, dort'u}

In some cases, the specific combination of several affixes expresses a feature that no one affix could accomplish on its own. For example, future tense in Laz is created by joining slots 12 (subjunctive) and 16 (auxiliary), in addition to the root and person markers in slots 3, 5, and 13. The present perfect construction is achieved via a combination of slots 4, 9, and 10, again with the addition of the root and person markers. The following examples demonstrate some of the combinatorial possibilities of Laz verbal morphemes:

(1) ncir-an
sleep-3.PL
Cl: 5-13
'They are sleeping'

(2) g-dzir-i
2-see-1.PST
Cl: 3-5-13
‘I saw you’

(3) p’-ç’kom-a-t-ere
1-eat-SUBJ-PL-AUX
Cl: 3-5-12-15-16
‘We will eat’

(4) v-i-ncir-a-t
1-PRV-sleep-SUBJ-PL
Cl: 3-4-5-12-15
‘Let us sleep’

(5) g-i-dzir-ap-u-n
2-PRV-see-$CAUS_{PERF}$-TS-3
Cl: 3-4-5-9-10-13
‘You have seen him’

(6) m-i-ç’kom-ap-u-t’-u-k’o
1-PRV-eat-CAUS-TS-COP-3.PST-COND
Cl: 3-4-5-9-10-11-13-14
‘If you had made me eat’

(7) o-m-o-ç’kom-in-ap-a-re
AFF-1-PRV-eat-$CAUS_{INTR}$-$CAUS_{TRANS}$-SUBJ-AUX
Cl: 1-3-4-5-7-8-12-16
‘She will certainly make me make you eat’

(8) g-o-tzir-i
2-PRV-show-1
Cl: 3-4-5-13
‘I showed you to him/her’

In the above examples, we see that every conjugated verb contains obligatory person marking, which in some cases—such as (2), (5), (6), and (8)—is realized on either side of the root. Some other morphological constructions present in these examples are future tense (3), optative (4), perfective aspect (5), conditional mood (6), and double causativization in (7). These examples show only a small set of combinatorial possibilities—the analysis in the following section will go on to examine the relationship between affixes in much greater detail.

5.2 PCM approach

Having briefly reviewed the various affixes of Laz, we may now turn to our analysis of the verbal morphology through the lens of the position class morphology (PCM) approach. PCM seeks to analyze blocking of a morphological, rather than a semantic nature—that is, blocking which is explainable in terms of meaning is not of interest to PCM, which rather seeks to examine cases in which the reason for blocking is not obvious. The following PCM analysis will be centered on Laz verbal morphology, as affixes on nouns and adjectives do not exhibit morphological blocking.

5.2.1 Semantic blocking

Before moving on to the PCM analysis of morphological blocking, let us examine cases of semantic morpheme blocking in the Laz verbal template. There are two main cases in which blocking of a semantic nature occur: reflexivity and valency.

5.2.1.1 Reflexivity

Firstly, we may consider the case of reflexivity. Laz restricts the coindexation of a grammatical subject and object on a single verb, with reflexivity being marked through alternative strategies, a phenomenon that is present in many languages, including English:[15]

English		*Laz*	
(9)	*I saw me	(10)	*Ma (ma) m-dzir-i/b-dzir-i 1.SG (1.SG) 1-SEE-1.PST 'I saw me'
(11)	I saw myself	(12)	Ti çkimi b-dzir-i head 1.POSS [myself] 1-SEE-1.PST 'I saw myself' (e.g. in the mirror) Lit. 'I saw my head'
		(13)	Ma (yali-s) v-i-dzir-i 1.SG (mirror-loc) 1-PRV-SEE-1.PST 'I saw myself (in the mirror)'

In the above examples we see that in the case of a reflexive, the first person subject marker and first person object marker are unable to cooccur in a non-applicativized transitive verb. Although the examples (10), (12), and (13) illustrate double person marking (i.e. person markers are realized both before and after the root), Laz prevents redundancy through blocking coindexation of subject and object in a predicate which lacks an applicative (class 4) prefix. Therefore, Laz must make use of

[15] Although in English there is no inherent reason that the subject and object of a verb may not be coindexed, in Laz this serves to eliminate potential ambiguity (see section 5.2.2.7 on Laz verbal agreement).

the mechanisms seen in (12) and (13) to express reflexivity via the introduction of a new argument.

Person markers will be discussed in greater detail in the following section in relation to morphological blocking, where I argue that these markers, which occur in slots 3 and 13, are actually circumfixal in nature and subject to an argument-based hierarchy.[16] As such, no circumfix exists which may represent a coindexed subject and object, and thus the affixes which may fill slot 3 (the prefixal person-marker position) exhibit semantically-motivated reciprocal blocking in order to prevent potential ambiguity in person marking.

5.2.1.2 Valency markers

The class 4 pre-root vowel (*a-*, *i-*, *o-*, *u-*) is obligatory in a variety of constructions, and it has been the topic of much syntactic study, including Demirok (2014) and Öztürk (to appear), who detail the distribution of this morpheme and its syntactic representation. (14) through (17) demonstrate several of the many constructions requiring valency markers:

(14)	Ma	si	himus	g-o-tzir-i
	1.SG	2.SG	3.SG.DAT	2-VAL-show-1.PST
	'I showed you to him/her'			

(15)	Xe	do	p'ici	v-i-bon-am
	Hand	and	mouth	1-VAL-wash-TS
	'I wash my hands and mouth'			

[16] This hierarchy, discussed in 5.2.2.7, is posited in Öztürk & Pöchtrager (2011) and is based on personhood and thematic role of the arguments of a verb.

(16) Xe do p'ici u-bon-am-s
Hand and mouth VAL-wash-TS-3.SG
'(S)he washes his/her hands and mouth'

(17) Ma ham layç'-epe-şe m-a-sk'ur-in-e-n
1.SG these dog-PL-ABL 1-VAL-fear-$CAUS_{INTR}$-TS-3.SG
'I am afraid of these dogs'

Demirok (2014) identified a hierarchy among these four pre-root vowels (i.e. *a->o->i-/u-*) and shows that they are in competition for a single slot on the verbal template. He further argues that the pre-root vowel *a-* is the result of the combination of the lower applicative vowels *i-* and/or *u-* (which denote non-third and third person, respectively), a hypothesis which is further supported by Öztürk (to appear), who posits that there are no low applicatives present in Laz and that combinatorial possibilities exist between the high applicative markers, further elaborating on the hierarchy among the various valency markers:

a > *o* = *i* > *u/i* > *u/i* > *i*
$Applicative_{DativeSubj}$ Causative Passive/middle $Applicative_{Benef/Poss}$ $Applicative_{Recip}$ Reflexive

Fig. 6 Hierarchy of applicatives (Öztürk, to appear)

Clearly, although these morphemes demonstrate a somewhat complicated relationship and merit significant examination, they are semantically determined by the argument structure of the constructions in which they appear. As such, the environments in which they occur may be specified and classified as semantic blocking, removing them from the realm of interest to PCM.

5.2.1.3 Valency and thematic suffixes

Another instance in which we observe semantic morpheme blocking is in the relationship between classes 4 and 10 (valency/applicative markers and thematic suffixes, respectively), which has been examined by Taylan & Öztürk (2014), among others. Laz verbs take one of five thematic suffixes when conjugated in certain constructions (e.g. present tense). These suffixes have been determined by Taylan & Öztürk (2014) to denote imperfectivity. Furthermore, they reveal information about a verb's argument and event structure as well as its lexical aspect (Öztürk & Pöchtrager, 2011, p. 90). The choice of thematic suffix is dependent on the argument structure of the verb as well as the semantic roles its arguments fulfill, as shown in Table 4:

Table 4. Overview of Thematic Suffixes and Their Functions

TS	Verb Type	Argument(s)	Nature of Verb	Use in Other Constructions
-um	transitive (ergative)	agent/patient	state change (e.g. "cut"), creation, and consumption	
-am	unergative	agentive activities or non-agentive verbs of emission (e.g. "swim", "shine")	No material change (e.g. "hold", "send")	Causatives
-u(r)	unaccusative	patientive subjects	statives, e.g. "live (in a place)", achievements (e.g. "die"), directed motions (e.g. "come")	Present perfect
-e(r)	unaccusative	patientive subjects	Processes (e.g. "lose weight"), psychological states (e.g. "love")	Passives
-em	unaccusative	experiencer	verbs of cognition (e.g. "see")	

The class 4 valency/applicative marker, which immediately precedes the root, has a very specific relationship with the various class 10 thematic suffixes. This prefix always occurs with unaccusative verbs taking *-e(r)*, may sometimes occur with *-em*, *-am*, and *-u(r)*,[17] and never occurs with *-um*. The following examples in (18) demonstrate these relationships:

(18)

a) *Ma b-dzir-er
1.SG 1-see-TS
'I see me'

b) Layç'i g-a-şkur-in-er
Dog 2-PRV-fear-CAUS-3
'You are afraid of dogs'

c) K'oçi b-dzir-em
man 1-see-TS
'I see the man'

d) Himu-k oxorca i-çin-em-s
He-ERG woman PRV-know-TS-3.SG
'He knows the woman'

e) Bere-k dits-am-s
Child-ERG laugh-TS-3.SG
'The child is laughing'

f) Himu-k layç'i-s n-o-dits-am-s
He-ERG dog-DAT PV-PRV-laugh-TS-3.SG
'He is laughing at the dog'

g) Ma Xopa-s p-skud-ur
1.SG Hopa-LOC 1-live-TS
'I live in Hopa'

h) Ma Xopa-s m-u-skud-ap-ur[18]
1.SG Hopa-LOC 1-PRV-live-$CAUS_{PERF}$-TS
'You have made me live in Hopa'

i) Kart'ali p'-ç'ar-um-t
Letter-NOM 1-write-TS-PL
'We write a letter'

j) *Cku himu-s v-u-nç'ar-um-t
1.PL 3.SG-DAT 1-PRV-write-TS-PL
'We are writing to him/her'

[17] Only in the case of present perfect conjugation
[18] Note that the informant found this expression to be semantically awkward and preferred the form *mosk'uledinam,* the inflected form of *osk'uledinu* ('to feed, nourish')

In (a), (c), (e), (g), and (i) we see that four of the five thematic suffixes can be present without a valency marker/pre-root vowel, with the exception of *-e(r)*. In the second column—(b), (d), (f), (h), and (j)—we see that only in the case of *-um* is a valency marker not allowed. Looking at the argument structure associated with each of these suffixes, we see that *-um* patterns with verbs of a strictly transitive (agentive-patientive) nature. The addition of arguments to this structure alters the semantics of the verb, and we can see this more clearly by examining examples (i) and (j) above.

In (i), *pç'arumt* ('we write') contains the thematic suffix *-um* because there is a clear agent ('we') and patient ('a letter'). In (j), however, the addition of an indirect object changes the relationship of the arguments from one of agent-patient to an unergative verb denoting agentive activity, causing it to select the thematic suffix *-am*. Whereas in (i) the subject (i.e. 'letter') is experiencing a change in state in being written, the action of writing to a third party patterns more closely with an agentive activity in which there is no material change of state. Therefore, the grammatical form of **vunç'arumt* in (j) is actually the following:

(19)

v-u-nç'ar-am-t
1-PRV-write-TS-PL
Cl:[19] 3-4-5-10-15
'We are writing to him/her'

As we have seen, each of the verbs exemplified in (18) and (19) may encode additional arguments or achieve other grammatical functions (e.g. causativization)

[19] 'Cl' refers to classes (numerical classes represent horizontal slots)

through applicativization, the need for which is indicated by the thematic suffix as well as the arguments of the verb (i.e. in the case of *-em*/*-am*/*-ur*, the necessity of the class 4 morpheme may be deduced from the theta roles of the arguments).[20] Only a certain subset of the class 10 thematic suffixes (i.e. *-um*) can block the class 4 prefix. Moreover, this case of blocking is clearly of a semantic nature and is therefore of no concern to PCM, which seeks to explain instances of blocking which may not be deduced from the meaning alone.

5.2.1.4 Aspectual blocking of thematic suffixes

As previously mentioned, TSs have been analyzed by Taylan & Öztürk (2014) as markers of imperfect aspect, often appearing in conjunction with the copula *-t'*. As such, some moods/aspects require a thematic suffix, while others prohibit it. Table 5 below illustrates various contexts in which TSs may (or may not) appear. Note that the semantic roles associated with the verbs, *oskudu* ('to live') and *odistu* ('to laugh'), necessitate different TSs, despite the fact that both are inherently intransitive.

As we can see in Table 5, TSs are realized in constructions with imperfect aspect (e.g. present/progressive and past imperfective), as well as in the present perfect construction, in which the TS *-u(r)* denotes a stative reading. The blocking of this morpheme in other constructions is semantically motivated, given that imperfective aspect is incompatible with certain other constructions, such as simple past and future tense.

[20] The various thematic suffixes and their associated argument structures are detailed in Öztürk and Taylan (2014).

Table 5. Realization of Thematic Suffixes

	oskudu (‘to live’)	*oditsu* (‘to laugh’)
Present tense	Xopa-s skud-ur Hopa-LOC live-TS Cl (verb): 4-10 ‘You live in Hopa’	dits-am laugh-TS Cl: 5-10 ‘You are laughing’
Imperfect	Xopa-s skud-u-t’-i Hopa-LOC live-TS-COP-2.PST Cl (verb): 5-10- You used to live in Hopa’	dits-am-t’-u laugh-TS-COP-3.PST 5-10-11 ‘(S)he used to laugh’
Present perfect	Xopa-s u-skud-ap-u-t’-u Hopa-LOC PRV-live-CAUS$_{PERF}$-TS-COP-3.PST Cl (verb): 4-5-7-10-11-13 ‘(S)he has lived in Xopa’	g-i-dits-ap-u-n 2-PRV-laugh-CAUS$_{PERF}$-TS-3 Cl: 3-4-5-9-10-13 ‘You have laughed’
Past tense	Xopa-s skud-i Hopa-LOC live-2.PST Cl (verb): 5-13 ‘You lived in Hopa’	b-dits-i 1-laugh-1.PST Cl: 3-5-13 ‘I laughed’
Future tense	Xopa-s skud-a-re Hopa-LOC live-SUBJ-AUX Cl (verb): 5-12-16 ‘You will live in Hopa’	dits-a-re laugh-SUBJ-AUX Cl: 5-12-16 ‘you will laugh’
Past of Future	Xopa-s skud-a-rt’u Hopa-LOC live-SUBJ-AUX Cl (verb): 5-12-16 ‘(S)he was going to live in Hopa’	u-dits-ap-ur-t’-a-s-ert’u PRV-laugh-CAUS-TS-COP-SUBJ-3-AUX Cl: 4-5-9-10-11-12-13-16 ‘(S)he was going to laugh’
Present (CAUS)	Xopa-s g-i-skud-ap-u-n Hopa-LOC 2-PRV-live-CAUS-TS-3.SG CL (verb): 3-4-5-7-10-13 ‘(S)he makes you live in Hopa’	v-o-dits-in-am 1-PRV-laugh-CAUS$_{INTR}$-TS Cl: 3-4-5-7-10 ‘I make him/her laugh’
Past (CAUS)	Xopa-s g-o-sk’uled-in-u Hopa-LOC 2-PRV-take care-CAUS$_{INTR}$-3.PST Cl (verb): 3-4-5-7-13 ‘(S)he made you live in Hopa’	v-o-dits-in-i 1-PRV-laugh-CAUS$_{INTR}$-1.PST Cl: 3-4-5-7-13 ‘I made him/her laugh’
Double causative	Hak m-o-sk’uled-in-ap-u Here 1-PRV-take care-CAUS$_{INTR}$-CAUS$_{TRANS}$-3.PST Cl (verb): 3-4-5-7-8-13 ‘(S)he made me make him/her live here’	m-o-dits-in-ap-u 1-PRV-laugh-CAUS$_{INTR}$-CAUS$_{TRANS}$-3.PST Cl: 3-4-5-7-13 ‘(S)he made me make him/her laugh’

5.2.1.5 Perfective causative and thematic suffixes

Continuing the discussion of thematic suffixes, we encounter another instance of semantic blocking when a TS cooccurs with a perfective causative suffix. Not all thematic suffixes are affected: this blocking does not occur with the TS *-u(r)*. The following examples in (20) demonstrate this phenomenon:

(20)

a)	ç'ar-um-s write-TS-3.SG Cl: 5-10-13 'he/she is writing'	b)	u-nç'ar-ap-u-n PRV-write-CAUS$_{PERF}$-TS-3.SG Cl: 4-5-9-10-13 'he/she has written'
c)	dits-am laugh-TS Cl: 5-10 'you are laughing'	d)	g-i-dits-ap-ur 2-PRV-laugh-CAUS$_{PERF}$-TS Cl: 3-4-5-9-10 'you have laughed'

If the perfective causative suffix *-ap* is present in a stem, the TS is always *-u(r)*, which, as stated earlier, is used in unaccusative verbs, which (in Laz) consist of statives, achievements, or directed motions. As stated in Öztürk & Pöchtrager (2011, p. 72), the perfective construction "grammatically marks the lack of temporal delimitation on the predicate." This lack of temporal information triggers a semantic shift and forces a stative reading of the verb, and the perfective construction thus selects the TS *-u(r)*. As this affix selection is clearly semantically motivated, it is therefore of no further interest to the following analysis of morphological blocking.

5.2.2 Morphological blocking

In addition to semantic blocking, crucially, Laz demonstrates several different instances of blocking which occur independently of the semantics of the verb. This section will examine these cases, which will all be classified as morphological blocking.

5.2.2.1 Valency and causativization

The first case of morphological blocking concerns the relationship between class 4 (the pre-root vowel, which can serve as a valency marker) and the causative markers in classes 7, 8, and 9. Although the former class is not contiguous with the latter classes in terms of precedence, there is a direct connection between them, as evidenced by the examples in (21):

(21)

a) m-i-nçir-s
   1-PRV-swim-3.SG
   Cl: 3-4-5-13
   'he/she is swimming for me (instead of me)'

b) m-o-mçvir-in-ap-u
   1-PRV-swim-$CAUS_{INTR}$-$CAUS_{TR}$-3.PST
   Cl: 3-4-5-7-8-13
   'he/she made me make him/her swim'

c) *p-çvir-in-ap-u
   1-swim-$CAUS_{INTR}$-$CAUS_{TR}$ -3.PST
   Cl: 3-5-7-8-13
   'he/she made me make him/her swim'

d) i-çaliş-in-u
PRV-work-CAUS$_{INTR}$-3.PST
Cl: 4-5-7-13
‘it was worked on’

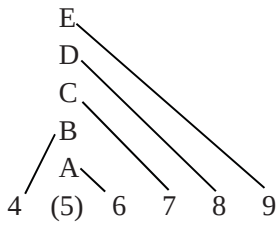
e) *çaliş-in-u
work-CAUS$_{INTR}$-3.PST
Cl: 5-7-13
‘it was worked on’

As these examples demonstrate, the class 4 prefix effectively licenses the causative suffixes in classes 7, 8, and 9. Without a pre-root vowel, no causative suffix may appear on the verb. However, the fact that a class 4 prefix may appear in the absence of causative suffixes demonstrates that class 4 is not dominated by classes 7, 8, and 9, and we thus will place the valency marker at a higher level of dominance than the causative markers.

In an initial iteration of Laz affixal hierarchy, we assign class 4 to dominance level B, creating contiguity with classes 7, 8, and 9 on the vertical axis. These classes demonstrate the licensing mentioned earlier in this section, and this quality of patterning together suggests a dominance-motivated relationship. As a result, the augmentative suffix in class 6, in order to not intervene vertically between class 4 and the causative suffixes, will fill position A on the vertical axis,[21] as illustrated in (22), a partial iteration of our emerging hierarchical model:

[21] Furthermore, the augmentative (AUG) affix attaches only to single-consonant roots and is therefore directly licensed by the root.

(22)



In (22), each of the numbers along the horizontal axis corresponds to a slot on the verbal template (with slot 5 being the root, which is not assigned to a dominance class). In keeping with PCM, the relative dominance of each affix is represented on the vertical axis, with A being the most dominant position. At this point, only classes 4 through 9 are illustrated, as they are contiguous and in close proximity to the root. The addition of more classes later will not alter these structural relationships because, as per the rules of PCM, lines connecting precedence with dominance may not cross. As we continue to analyze the various instances of blocking in Laz, we will expand this hierarchy to include the remainder of the position classes.

5.2.2.2 More on causative suffixes

Continuing with the discussion of causative suffixes, we now reach an examination of the relationship between classes 7, 8, and 9. Each of these morphemes indicate causation: intransitive, transitive, and perfect, respectively. The following examples in (23) show all possible combinations of these affixes (in the case of double causativization, more than one of these affixes may appear at a given time):

(23)

a) Causative occurring alone

(i) Class 7 alone	(ii) Class 8 alone	(iii) Class 9 alone
v-o-ncir-**in**-i	m-o-ç'k'om-**ap**-u	m-i-çk'om-**ap**-ur-t'-u
'I made him sleep'	'(S)he made me eat'	'I had eaten it before'

b) Double causatives (grammatical)

(iv) Classes 7 and 8	(v) Classes 8 and 9	(vi) Classes 8 and 9
m-o-ncir-**in**-**ap**-u	u-ç'k'om-**ap**-**ap**-u-n	u-bgar-**ap**-**ap**-u-n
'(S)he made me make him sleep'	'(S)he has made him eat it'	'(S)he has made him make it cry'

c) Double/triple causatives (ungrammatical)

(vii) Classes 7 and 9	(viii) Classes 7, 8, and 9
*v-o-ncir-**in**-**ap**-i	*u-bgar-**in**-**ap**-**ap**-u-n
'I have made him sleep'	'(S)he has made him make it cry'

As demonstrated in (23), only classes 7 and 9 are in complementary distribution among the causative suffixes, including in constructions that encode the semantics of all three suffixes, as in example (vi). Although there is no semantic reason that an intransitive causative suffix and a perfective causative suffix should not coexist, it appears that the *-in* morpheme of class 7 is blocked in perfective double causative verbs such as *ubgarapapun* (example vi).

At this point I argue for an alteration of the existing Laz verbal template, namely that the class 8 causative transitive marker does not occupy a fixed slot and may be realized either in slot 7 or slot 8 on the template. As a result, the perfective causative suffix can be posited to occupy slot 8, eliminating slot 9 from the domain of causative morphology. This new representation is formalized in (24):

(24)

a) Previous template of causative classification

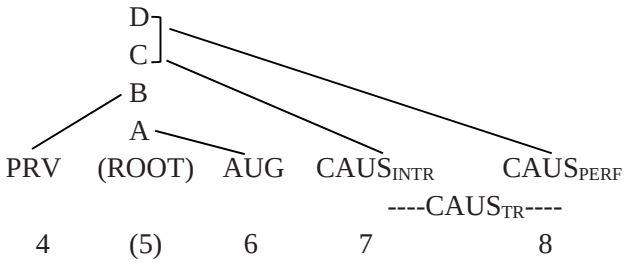
7	[CAUS.INTR]/[IP]	causative marker for intransitive verbs or impersonal passive marker
8	[CAUS.TR]	causative marker for transitive verbs
9	[CAUS.PRF]	causative marker for perfective verbs

b) New proposal for causative classification

7	[CAUS.INTR]/[IP]	causative marker for intransitive verbs or impersonal passive marker
8	[CAUS.PRF]	causative marker for perfective verbs
7-8	[CAUS.TR]	causative marker for transitive verbs

In shifting the model from (a) to (b) in (20), we may now alter our hierarchical structure to reflect these changes:

(25)



This change in representation is sufficient to explain the complementary distribution between the perfective and intransitive causative suffixes—the transitive causative marker will select class 8 when it is not occupied by a perfective causative. If class 8 is occupied, however, the transitive causative suffix will select class 7, where it may block the intransitive causative marker. The limitation of two causative markers per

verbal stem is further evidence that the three morphemes are competing for a total of two available positions, explaining the occurrence of formerly problematic forms such as *ubgarapapun.*

One additional modification appearing in this iteration of the hierarchical model is that of the bracket connecting classes C and D. This signifies the mass blocking effect of the pre-root vowel (PRV) in class B on the various causative suffixes. This representation was first employed by Inkelas (1989) to explain various mass blocking phenomena in Nimboran.

5.2.2.3 Blocking of the conditional marker

Following the class 11 copula, the class 12 subjunctive marker, and the class 13 person marker occupying, the conditional marker in class 14 is blocked by preceding suffixes in many cases:

(26)

a) p’-ç̌kom-i-ǩo-t
1-eat-1.PST-COND-PL
Cl: 3-5-13-14-15
‘If we ate it’/‘Let us eat it’

b) gor-um-t-na
want-TS-PL-COND
5-10-15-(14)
‘If you (pl.) want’

c) gor-um-s-na
want-TS-3.SG-COND
Cl: 5-10-13-14
‘If he/she wants’

d) m-i-ç̌om-ap-ur-t’-u-ǩo
1-PRV-eat-CAUS$_{PERF}$-TS-IMPF-3.PST-COND
Cl: 3-4-5-9-10-11-13-14
‘If I had eaten it’

e) ç̌om-a-na
eat-SUBJ-COND
Cl: 5-12-14
‘If you are going to eat’

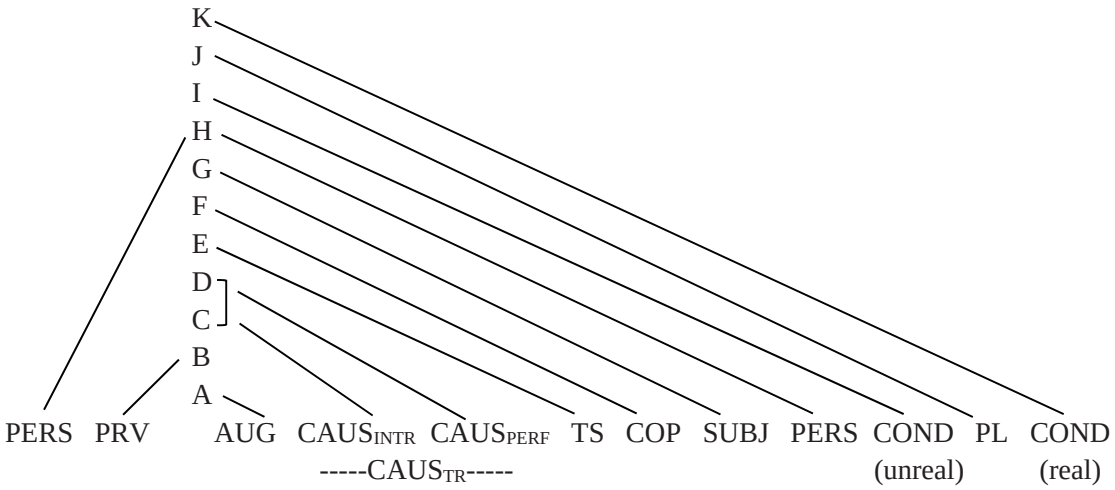
f) p’-ç̌om-a-t-na
1-eat-SUBJ-PL-COND
Cl: 3-5-12-15-(14)
‘If we are going to eat’

g) g-o-bgar-in-i-k’o
2-PRV-cry-CAUS$_{INTR}$-1.PST-COND
Cl: 3-4-5-7-12-14
‘If I made you cry’

In these examples, we observe that these two conditional suffixes differ in their distributions. While the unreal conditional *-ǩo* may occur only after the past tense marker—and the thematic suffix in the present perfective construction in (d)—the real conditional *-na* has a wider distribution and may follow the plural suffix as in (b), the person marker as in (c), and the subjunctive marker as in (e).

Based on the examples in (26), it is evident that the two conditional suffixes, *-ǩo* and *-na*, do not belong to the same position class, as the unreal or counterfactual conditional *-ǩo* occurs only after the past tense marker, while the real conditional *-na* occurs after the plural marker in class 15. The behavior in relation to the plural suffix is the clearest indicator that these two conditionals, although semantically related, are not variations of the same affix. As a result, there is no reason to group these two morphemes together within the template, and they may be more accurately represented as in (27):

(27)



In (27), we see that the unreal conditional *-ǩo* occupies the slot between the person and plural markers, which corresponds to class I on the dominance axis. The real conditional *-na* follows the plurality marker and is assigned to position class K. This new representation aligns with the behavior of each affix, as *-ǩo* may only appear immediately after the past tense marker, while *-na* has a wider distribution and may occur with a variety of tense markers, including SUBJ (to create future tense) and TS (to

create present tense, etc.). This is significant because all the suffixes preceding *-ǩo* (i.e. CAUS, TS, COP, and SUBJ) exhibit a mass blocking effect on the unreal conditional, which immediately follows them on the vertical axis.

One exception to this mass blocking is in verbs with the present perfect construction, such as *miçǩomapurt'uǩo* in (26d). Only in this construction may the unreal conditional appear with any other suffixes than the past tense marker, with the inflection CAUS$_{PERF}$+TS+(COP). The real conditional *-na* may also appear in present perfect forms:

(28)

g-i-çǩom-ap-u-n-na
2-APPL-eat-CAUS$_{PERF}$-TS-3.SG-COND$_{REAL}$
Cl: 3-4-5-7-9-13-15
'If you have eaten'

We will return to this issue in an upcoming section detailing the phenomenon of inversion.

5.2.2.4 Blocking among preverbs

At the leftmost position on the horizontal axis are two preverbal affixes which convey affirmation and spatial information, respectively. In total there are four affirmative prefixes and 32[22] spatial prefixes. Semantically, affirmative prefixes add a level of certainty to the occurrence of the verb, while the spatial prefixes provide information

[22] Nine of these are simplex forms, and the remaining twenty-three are complex, having fused with one out of a closed set of particles. These prefixes and their combinatorial properties are discussed at length in Eren (2015).

about the manner, direction, and location of the action. The manifestation of the spatial prefix is not subject to any effects of blocking, but the preceding affirmative preverb may only appear in certain contexts:

Table 6. Laz Preverbs

Preverb	Occurs With	Example
do-	non-spatial prefixes (or root)	do-t'ax-u PV_{AFF}-break-3SG.PST Cl: 1-5-13 '(s)he broke it'
ko-	spatial prefixes	ko-mo-xt'-u PV_{AFF}-PV_{SPAT}-come-3SG.PST Cl: 1-2-5-13 '(s)he did come' ko-dzir-i-t PV_{AFF}-see-2.PST-PL Cl: 1-5-13-15 'you (pl.) did see it'
menda-	limited number of roots	me-g-a-ndv-i PV_{AFF}-2-APPL-confide-1.PST Cl: 1-3-4-5-13 'I told you a secret'
o-	limited number of roots	o-ǰǩom-u PV_{AFF}-eat-3SG.PST Cl: 1-5-13 '(s)he did eat'

Firstly, the prefix *do-* is completely blocked by the presence of a spatial morpheme in the following slot, while *ko-* appears with a spatial prefix in most cases but a small number of verb roots select this prefix by itself. The other two affirmative preverbs, *menda-* and *o-*, are lexically determined but like *do-* are blocked by the spatial morpheme. This behavior reaffirms the tenets of PCM, as the outermost affix receives

the effect of blocking from its neighbor, which is by necessity at a higher level of dominance.

At this point we may finally produce a full hierarchical model in Figure 7, which represents all verbal affixes in Laz. The decision to place the preverbs in the lowest-ranked slots was a speculative one—semantically, these leftmost prefixes (PV_{AFF} and PV_{SPAT}) encode affirmative, directional, and locational information and may occur with any other constructions (e.g. tense, aspect, mood), while the rightmost auxiliary suffix encodes key temporal information and has a closer tie to other temporally and aspectually significant affixes. This assignment is speculative and open to alternate interpretation.

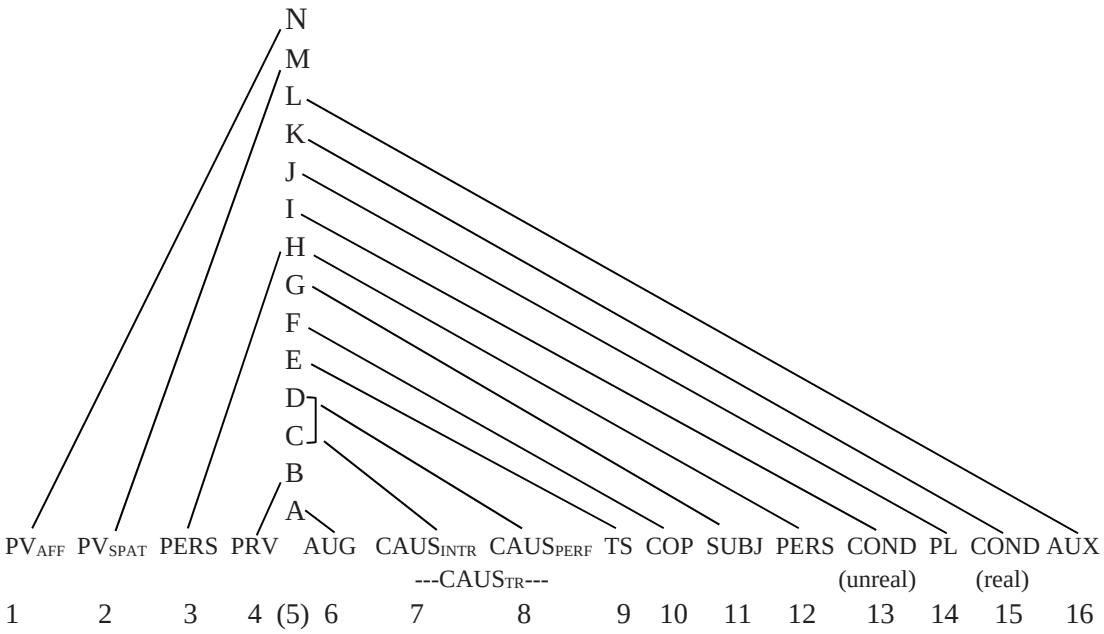


Fig. 7 Laz verbal hierarchy

As we can see in Figure 7, significant changes have taken place within the verbal template set forth at the beginning of the chapter, most noticeably the addition of the vertical element into the structure. Despite the changes, I still posit a total of 16 slots in the verbal template, as two causative affixes have been collapsed into the new precedence class 7, and conditional affixes have been separated into classes 13 and 15. Following the upcoming discussion of person markers, this hierarchy will be applied within a discussion of potential applications for an online lexical resource for Laz.

5.2.2.5 Person markers

The last point to consider is the relationship between the two person markers, which has been briefly touched upon in the discussion of thematic suffixes. Hierarchically, the person markers have thus been classified as an ambifix which has its head in slot 3. This relationship will be explored in more detail in this section.

In determining the hierarchical relationships present in Laz verbal morphology, the system of person marking provides what is possibly the most substantial challenge. Laz verbs have double person marking and can agree with subjects and objects (both direct and indirect) simultaneously,[23] as in (29):

(29)	Ali-k	ma	m-	dzir	-em	-s
	Ali-ERG	1SG	1	see	TS	3
	Cl:		3	5	9	12
	‘Ali sees me’					

[23] Note that in ditransitive constructions with more than two arguments, only two of these can be marked on a single verbal stem.

The following reproduced verbal template in (30) shows the location of the two person-marking slots:

(30)

1	2	**3**	4	5	6	7	8	9	10	11	**12**	13	14	15	16
PV1	PV2	**pers**	val	ROOT	aug	caus-intr/ caus-perf	caus-tr	TS	COP	subj	**pers**	cond (unreal)	PL	cond (real)	mod

In order to understand the relationship between slots 3 and 12, we may begin with an examination of nominally intransitive verbs[24] in order to identify the subject-marking morpheme:

(31)

	onciru (to sleep)	*oskidu* (to live)	*oğuru* (to die)
1SG.PRES	v-incir	p-skidur	bğurur
2SG.PRES	incir	skidur	ğurur
3SG.PRES	incir-s	skidu-n	ğurur-s
1SG.PST	v-inc-i	p-skid-i	b-ğur-i
2SG.PST	inc-i	skid-i	ğur-i
3SG.PST	inc-u	skid-u	ğur-u

As shown in (31), first person, both singular and plural, is marked with a labial prefix (*v-*/*p-*/*p̌-*/*b-*/*f-*), also known as the 'v-set' (Holisky, 1991). Neither second nor third person prefixes are overtly realized and thus can only be disambiguated through a

[24] Öztürk and Taylan (2014) have argued that intransitive verbs do not exist in Laz, because semantically every verb has an initiator and an undergoer.

difference in suffix (Ø or -s/-n, respectively).[25] I therefore reaffirm the proposal that the subject marker in Pazar Laz is circumfixal in nature and occupies both slots 3 and 12 on the verbal template.[26] These circumfixes are reproduced for clarity in (32), with the addition of past tense circumfixes, which vary slightly:

(32)

	Non-Past	Past
1	{v/p/p̌/b/f-…-Ø}	{v/p/p̌/b/f-…-i}
2	{Ø-…-Ø}	Ø-…-i
3	{Ø-…-s/-n}	Ø-…-(t)u

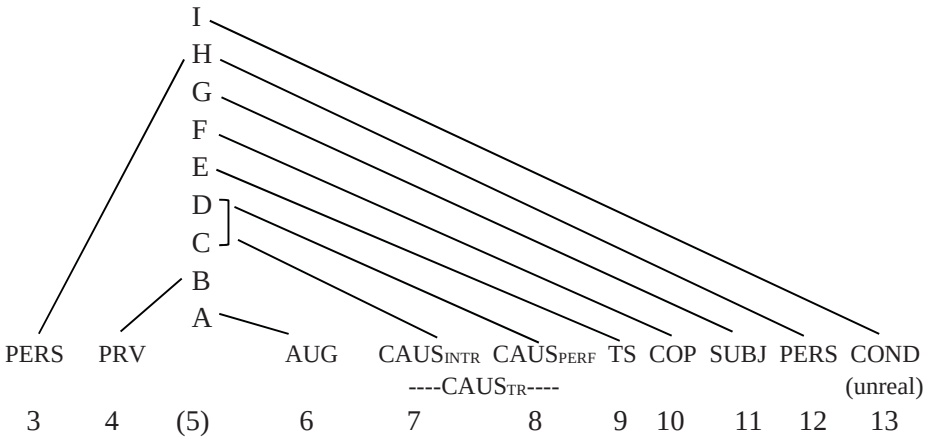
As we can see, the second person marker is the most minimally marked, and we will later observe that it is also exhibits the most regular behavior. Notice that there is double overlap in each conjugation, as first and second person "suffixes" are always identical (i.e. *-i* for past or -Ø for non-past) and second and third person "prefixes" are always identical (i.e. Ø-). For this reason, it is impossible to isolate any single form (other than a circumfix) to correspond to a person marker, as each item on its own is syncretic. These two sets of circumfixes (i.e. past and non-past) are sufficient to account for any intransitive verb.

Having posited the circumfixal nature of person markers, we may return to the consideration of its representation on the hierarchical model. In the previous section (5.2.2.6), a hierarchical model of all verbal slots was produced. Below, we see a

[25] These forms also contain thematic suffixes, which occur in slot 10 and will be discussed in a later chapter
[26] The content of slot 13 also includes tense marking

simplified version of this chart in (33), which includes only classes 3-13 and posits the person marker circumfix in position class H:

(33)



In a closer examination of transitive verbs, however, we encounter our first difficulty: the realization of an object can manifest in unexpected ways. (34) shows the verb *odziru* (“to see”) conjugated in present tense with all possible person markers:

(34)

	1SG (subj)	2SG (subj)	3SG (subj)
1SG (obj)	*	mdzirem	mdziren
2SG (obj)	gdzirem	*	gdziren
3SG (obj)	bdzirem	dzirem	dziren[27]

Based on a cursory observation of this data, it seems that the slot preceding the verb is filled by the object marker, with the subject marker signified by the information

[27] This form is only possible when the subject and object are not coreferential

following the verb. However, we see two main discrepancies here: firstly, there seems to be no visible subject marker to overtly differentiate between first and second person forms, and thus the grammatical subject of a form such as *mdzirem* ('you see me') or *gdzirem* ('I see you') may only be determined by somewhat of a process of elimination (since reflexives take alternate forms to transitive verbs and thus both 'I see me' and 'you see you' are barred in this form).[28]

A second obstacle arises in the examination of forms with a third person object, as we would not expect a visible prefix on the verb, given that objects are primarily marked prefixally and that there is no overt prefixal marking to signify third person. Nevertheless one exists in the form *bdzirem* ('I see him/her'). At first glance, this seems to be a further case of disambiguation, since the expected realization of a form meaning 'I see him/her' would not have an overt prefix (due to the third person objecthood) and thus yield the form *dzirem*, identical to the form meaning 'you see him/her'.

Having already posited that the subject marker is circumfixal in nature, an examination of the transitive forms demonstrates that the object marker has the potential to alter the circumfix in some way, although this alteration seems to only create a distinction in the prefixal rather than the invariable suffixal portion of the circumfix. The pattern of object marking, therefore, appears to be primarily prefixal in nature—as will be exemplified in (35)—but this explanation alone is insufficient to explain the existence of a form such as *bdzirem*, which contains a prefixal subject marker appropriated to the position of the object marker. In order to understand why this is the case we must

[28] A first person object marker would signify a second person subject, and vice versa. Reflexives are marked by using the noun *ti* ("head") as the recipient of the reflexive action (and the verb is thus conjugated as one having a third person subject). Alternatively, the pre-root vowel *i-* may be used to signify verbal reflexivization (although this morpheme may fulfill other roles as well).

examine the hierarchical relationships among person markers in Laz and connect the argument to PCM.

As previously posited in Öztürk & Pöchtrager (2011), the existence of hierarchical relationships among verbal morphemes in Laz seems to hold true, particularly that of the realization of various person markers. Without such distinctions it would be impossible to account for the distinctions between the verbal forms from (34), reproduced in (35):

(35)

a)	Ma	si	g-	dzir	-em	-Ø
	1SG	2SG	2	SEE	TS	1
	Cl:		3	5	9	12
	‘I see you’					
b)	Ma	Ali	b-	dzir	-em	-Ø
	1SG	Ali	1	SEE	TS	1
	Cl:		3	5	9	12
	‘I see Ali’					
c)	Si	ma	m-	dzir	-em	-Ø
	2SG	1SG	1	SEE	TS	2
	Cl:		3	5	9	12
	‘You see me’					
d)	Si	Ali	Ø-	dzir	-em	-Ø
	2SG	Ali	3	SEE	TS	2
	Cl:		3	5	9	12
	‘You see Ali’					

In (35a) we observe a case of object agreement, but in (35b) this is not the case, as we have already established that the labial prefix *b-* is linked to the first person subject marker, not the third person object marker. The reason for this variation can only be explained by the existence of a hierarchy between subject and object. Öztürk & Pöchtrager (2011) illustrate that, for transitive verbs, a three-level hierarchy exists both between both person markers and theta roles, reproduced in (36):[29]

(36)

1) P: 1st and 2nd person
2) A: 1st person
3) P: 3rd person=A: 2nd and 3rd person[30]

Firstly, there is a general object > subject hierarchy (with the exception of third person objects), when the object of the verb is a theme or patient-like argument of a transitive verb and the object fulfills the theta role of agent. This is apparent from the examples in (35), as the only case in which we do not observe an object marker is that of *bdzirem* (i.e. example 35b), in which the first person object ranks higher than the third person subject. As a rule, first person dominates third person, no matter the theta role, while second person markers follow a more predictable pattern in that the object (i.e. P) is at the highest level of dominance and the subject (i.e. A) is at the lowest level of dominance. The prefixal person markers are listed in (37):

[29] Morphological rules for Georgian person marking (a similar system to Laz) can be found in Anderson (1982).
[30] P generally references a grammatical object which fulfills the theta role of patient or theme, while A references grammatical subjects which act as agents

(37)

1	m-
2	k/g-
3	Ø-

The ambiguity in the lowest level of the hierarchy (i.e. P3=A2/3) can be explained by the lack of an overt object prefix for second or third person. Although the second person subject-marking circumfix is identifiable in a form such as *dzirem* ("you see him/her" from example D), the lack of realized prefix presents two possibilities: firstly, the Ø- prefix could be linked to the third person object role. Alternatively, it could be a parallel structure to *v-*, the first person subject marker used in a first person subject/third person object scenario. As there is no subject-marking prefix in the case of second person, it is impossible to ascertain whether the unrealized prefix in *dzirem* corresponds to second or third person.

Expanding upon the hierarchical relationships linked to transitivity in Laz, we may introduce another argument, that which is marked by dative case and represents a non-core argument (possessing the theta role of recipient, goal, etc.). Ditransitive verbs may provide further insight into relationships between various arguments in Laz. The following examples from Öztürk & Pöchtrager (2011) in (38) show various ditransitive structures and bring further complication into the hierarchical discussion:

(38)

a) Ma si himus g-o-tzir-i
1.SG 2.SG 3.SG.DAT 2-APPL-show-1.PST
Cl (verb): 3-4-5-12
'I showed you to him/her'

b) Himuk ma si g-o-tzir-u
3.SG.ERG 1.SG 2.SG 2-APPL-show-3.PST
Cl (verb): 3-4-5-12
‘(S)he showed me to you’

c) Ma him himus v-o-tzir-i
1.SG 3.SG 3.SG.DAT 1-APPL-show-1.PST
Cl (verb): 3-4-5-12
‘I showed him/her to him/her’[31]

d) Himuk him himus o-tzir-u
3.SG.ERG 3.SG 3.SG.DAT APPL-show-3SBJ.PST
Cl (verb): 4-5-12
‘(S)he showed him/her to him/her’

From these examples, we can see that the inclusion of a third argument creates some unexpected forms. Looking at examples (39a) and (39b), we observe that both verb forms begin with the second person object marker *g-*, although the second person *si* does not fulfill the same theta role in both sentences (i.e. acting as the patient/theme in 39a and the recipient/goal in 39b). Obviously, then, the hierarchy system present in transitive verbs may be expanded to include a third theta role for the case of ditransitive verbs, as verbs in Laz may only contain maximally two person markers. Reproduced in (39) is the prefixal hierarchy posited in Öztürk & Pöchtrager (2011):

(39) D1/2[32]
P1/2
A1
D3=P3=A2/3

[31] Note that in (c) and (d) the third person subjects and objects cannot be coreferential
[32] D is defined as a dative-marked argument fulfilling the theta role of recipient or goal

With the exception of the third person form (which aligns with the transitive hierarchy and remains at the lowest level of priority), we can see that the dative-marked arguments (i.e. recipient and goal) rank at the highest level, as evidenced by (41b). Laz seems to prioritize the representation of more semantically-removed elements in verb forms, and this hierarchy can work in conjunction with the various affixes identified in this discussion to shape the final form of a given verb.

However, this hierarchy can cause ambiguities in certain ditransitive forms, such as that of *gotziru*:

(40) g-o-tzir-u
2-APPL-show-3.PST
Cl: 3-4-5-12

Based on the hierarchy, the only conclusion we may definitively draw from this utterance is that the agent in this case is third person singular, as shown by the final *-u*. The potential meanings, then, are threefold: '(s)he showed him/her to you'; '(s)he showed you to him/her'; or '(s)he showed me to you'. Further disambiguation is required in these cases by pronouns in the sentence as well as contextual clues.

To conclude the discussion of person markers, let us review the claims made in this section and connect them to PCM. Acknowledging that the consideration of theta roles in the establishment of a hierarchy is unavoidable, we may use this information to single out the specific morphemes that represent each person marker, whether subject or object (including non-core arguments). Those morphemes have been reproduced in (41):

(41)

	Subject (agent) + Non-Past	Subject (agent) + Past	Object (theme)	Object (recipient/goal)
1	{vi/p/p̌/b-…-Ø}	{vi/p/p̌/b-…-i}	m-	m-
2	{Ø-…-Ø}	{Ø-…-i}	k/g-	k/g-
3	{Ø-…-s/-n}	{Ø-…-(t)u}[33]	Ø-	Ø-

In order to combine these affixes within the PCM hierarchy, I propose that person markers in Laz must conform to a subhierarchy within the larger model, as these demonstrate the same blocking phenomena on a smaller scale. Although the prefixal elements of the person marking circumfixes are in complementary distribution, the semantic relationships alone cannot account for the blocking within the position class. By expanding the PCM hierarchy from the previous section to include the subhierarchy within person markers, we may capture all dominance relationships at the word and circumfixal level in Figure 8:

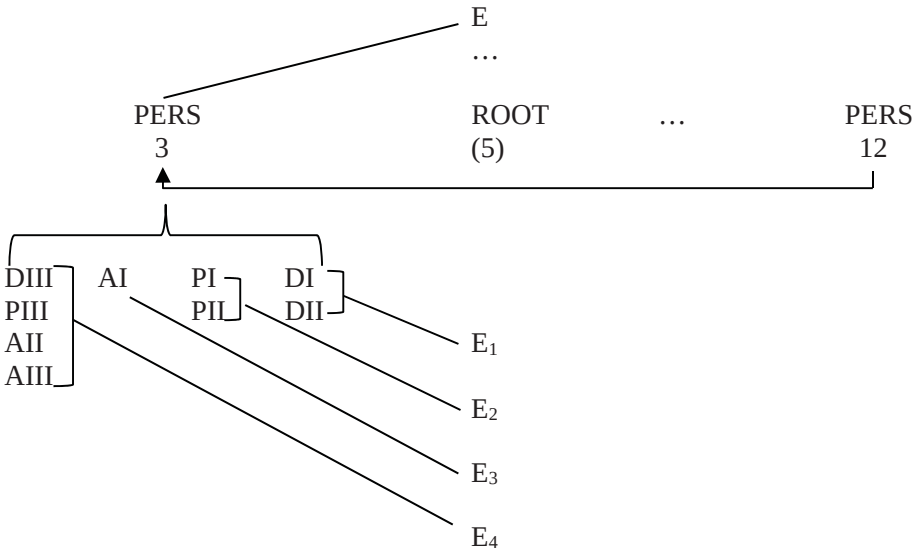


Fig. 8 Sub-hierarchy of person markers

[33] A {t} will surface between a vowel-final root and the past tense suffix *-u*

In this simplified representation, only person markers are shown, which, as previously posited, are assigned to dominance class E (in the vertical hierarchy). Furthermore, the slot 12 postverbal position attaches to the slot 3 prefixal position in order to create the overall circumfixal person marker. Although the information in slot 12 is not determined hierarchically (rather by the argument selected by the verb), the listed forms under slot 3 in Figure 8 represent all possible prefixes and their relative dominance.[34]

Within this class E subhierarchy, the dominance scale is "flipped" in that the highest level (E_1) represents the most dominant position (given that it is represented in the closest proximity to the verbal stem). As this is a subhierarchy occurring within the realm of dominance class E, the dominance positions for prefixal person markers are identified by class E with subscripts, with class E_1 being the most dominant and therefore able to block the realization of any of the following prefixes in alignment with the person marking hierarchy from (40). For example, the presence of a person marker from subclass E_1 (i.e. a dative-marked first or second person argument) would block the lower subclasses from being realized.

Based on the possibilities identified in this model, we may now make a comprehensive list of the Laz circumfixes in both non-past and past paradigms. The blocking effects in the subhierarchy posited in Figure 8 create a total of 34 possible person marker circumfixes:[35]

[34] In this model, D corresponds to dative (i.e. recipient/goal), P corresponds to patient/theme, and A corresponds to agent. D and P therefore pattern with object markers while A patterns with subject markers, as shown in the previous table.

[35] Note that intransitive verbs are subject to semantics rather than the blocking effects of the subhierarchy.

Table 7. Person Marker Circumfixes

a)

Non-Past	A1	A2	A3
D1	---	{m-…-Ø}	{m-…-s/-n}
D2	{k/g-…-Ø}	----	{k/g-…-s/-n}
D3	{p/ p̌/b-…-Ø}	{Ø-…-Ø}	{Ø-…-s/-n}
P1	---	{m-…-Ø}	{m-…-s/-n}
P2	{k/g-…-Ø}	---	{k/g-…-s/-n}
P3	{p/p̌/b-…-Ø}	{Ø-…-Ø}	{Ø-…-s/-n}
Ø (intransitive)	{vi/p/p̌/b-…-Ø}	{Ø-…-Ø}	{Ø-…-s/-n}

b)

Past	A1	A2	A3
D1	---	{m-…-i}	{m-…-(t)u}
D2	{k/g-…-i}	----	{k/g-…-(t)u}
D3	{p/p̌/b-…-i}	{Ø-…-i}	{Ø-…-(t)u}
P1	---	{m-…-i}	{m-…-(t)u}
P2	{k/g-…-i}	---	{k/g-…-(t)u}
P3	{p/p̌/b-…-i}	{Ø-…-i}	{Ø-…-(t)u}
Ø (intransitive)	{vi/p/p̌/b-…-i}	{Ø-…-i}	{Ø-…-(t)u}

However, the discussion of Laz person markers remains unfinished. There is a set of verbs which pattern differently by demonstrating ‘inversion’ of subject and object. In section 5.3, I will examine these verbs more closely and argue that they do not present any significant problems to a PCM analysis.

5.3 The inversion phenomenon

To this point, the only theta role that has been exemplified in the subject position is that of agent. However, a unique difficulty presents itself in the examination of verbs which have a subject that satisfies the theta role of experiencer. Consider the following sentences in (42):

(42)

a) Ma oroperi-çkimi m-a-orop-e-n
1SG lover-1SG.GEN 1-APPL-love-TS-3.SG
Cl: 3-4-5-9-12
'I love my significant other'

b) Si ma g-a-orop-er-Ø
2SG 1SG 2-APPL-love-TS-1.SG
Cl: 3-4-5-9-12
'You love me'

c) Ali-s si g-a-orop-e-n
Ali-DAT 2SG 2-APPL-love-TS-3.SG
Cl: 3-4-5-9-12
'Ali loves you'

In these examples we observe a phenomenon in which person markers seem to be reversed. This has been labelled by some (e.g. Harris, 1977) as 'inversion', and in Laz may be found in psychological verbs, verbs of ability, involuntary or physiological actions, and the present perfect construction.[36] Note that the hierarchy of person markers still applies in these cases, as shown in (42) above, and furthermore that in each example, a pre-root vowel (e.g. *a-* in the above examples) is obligatory.

At this point we must give special attention a subset of Laz verbs—namely, those verbs which have datively marked subjects. Demirok (2013) shows that several dative-marked 'notional' subjects in Laz only pass some syntactic subjecthood tests. While they may successfully bind anaphors and serve as gaps in control constructions, they do not

[36] Other analyses of Laz and other Caucasian languages propose alternative explanations for 'inversion', such as the multi-level rules in Anderson (1982).

always trigger verb final agreement,[37] instead defaulting to third person default agreement as demonstrated in (43):

(43) Ma si m-a-cer-u
1.SG 2.SG 1-VAL-believe-3.PST
'I believed you'

Let us now examine in more detail the set of verbs which display this behavior—verbs taking experiencer arguments, verbs of ability or involuntary/physiological actions, and verbs in the present perfect construction. In (44) we see each of these types exemplified:

(44)

a) Ma layç'-epe-şe m-a-şkur-in-e-n
1SG dog-PL-ABL 1-APPL-fear-CAUS-TS-3SG
Cl (verb): 3-4-5-7-9-12
'I am afraid of dogs'

b) Ma Lazuri do-m-a-gur-e-n
1.SG Laz PV-1-APPL-learn-TS-3SG
Cl (verb): 1-3-4-5-12
'I can learn Laz'

c) Ma skani gyari m-a-çkom-u
1SG 2.POSS food 1-APPL-eat-3SG.PST
Cl (verb): 3-4-5-12
'I (accidentally) ate your food'

[37] But may trigger agreement in the case of focus/emphasis, as detailed in Öztürk (2019)

d)	Ma	si	ce-m-i-ç-am-ap-ur-Ø
	1SG	2SG	PV-1-APPL-beat-AUG-CAUS.PERF-TS-2.SG
	Cl (verb):		2-3-4-5-6-8-9-(12)
	‘I have beaten you’		

In each of the examples in (44), we see that although the grammatical subject of the utterance is *ma* (i.e. the first person singular pronoun), the person markers seem to indicate that the first person argument fills the role of an object rather than that of a subject, in accordance with the previously-discussed person marker hierarchy. Furthermore, each of these examples contains an applicative marker in slot 4. Note that in (40b) and (40c), the inversion phenomenon (including the applicative vowel) is the only Furthermore, each grammatical subject exemplifying inversion is inherently dative-marked, which is not visible on pronouns but may be observed in (45):

(45)	Bere-s	ham	layç’-epe-şe	a-şkur-in-e-n
	Child-DAT	this/these	dog-PL-ABL	APPL-fear-CAUS-TS-3.SG
	Cl (verb):			3-4-5-7-9-12
	‘The child fears these dogs’			

Several analyses have been proposed to explain inversion, which is present across Caucasian languages. Taylan & Öztürk (2014) identify a transitive syntax and implicit initiator in the case of stative verbs, including those taking experiencer subjects such as *oşkurinu* (‘to fear’) in (41) as well as experiential perfect constructions such as in (40d), and this argument may extend to abilitative and involuntary/psychological forms as well.

Furthermore, in the case of verbs taking experiencer arguments, Landau (2010) elaborates on the special case of experiencers, asserting that the cognitive aspect of experiencers is linguistically unique (compared to other NPs) and that an experiencer may serve as a mental location in which a state holds true. This may provide some structural insight into the formation of these verbs in Laz, explaining why the applicative prefix *a-* is present in many stative experiencer verbs.[38] This argument holds for verbs of ability and involuntary actions as well, as each of these situations refers to a state holding true within an animate, cognizant experiencer.

The issue of inversion bears only minimal weight on the PCM analysis. Although syntactically this phenomenon merits the significant examination it has received, in our PCM model it is sufficient to simply note that this process occurs and include these "inverted" circumfixes in the inventory of affixes, given that they are semantically determined by certain constructions in order to denote a mental location and/or stative reading. Practically, however, the selection of person markers is dependent on the grammatical subject and object of the verb, and thus the root of a verb which bears an experiencer subject or an abilitative or perfective reading could be assigned incorrect (i.e. uninverted) person markers. This may be accounted for computationally, which will be detailed in Chapter 6.

5.4 A note on syntactic and morphological hierarchies

Before moving on to a discussion of the implementation of this analytical model, I wish to note an important distinction which must be made between syntactic and

[38] Demirok (2014) discussed these affixes and their syntactic derivations in more detail, claiming that the prefix [a-] is a result of multiple higher applicatives being realized simultaneously.

morphological hierarchies, particularly as they relate to Laz. I would be remiss if I did not emphasize the importance and relevance of syntactic (and morphosyntactic) analyses of Laz to the analysis in this chapter, as these constitute many of the most substantial and significant findings to date on the structure of the language.

As mentioned in the Chapter 4 as well as throughout the analysis, several key syntactic works have outlined and unearthed some important linguistic phenomena of the Laz language, including but not limited to the morphosyntactic analysis of agreement and case by Demirok (2013), the discussion of transitivity by Taylan & Öztürk (2014), and the examination of various applicatives in Laz by Öztürk (to appear).

The previous PCM analysis of Laz draws from these syntactic analyses in certain regards. For example, in considering the various subhierarchies discussed here (i.e. of valency markers and person markers, respectively), an examination of existing morphosyntactic analysis was critical in classifying these instances of blocking as either semantic or morphological. In the case of valency markers, the syntactic analysis of Demirok (2014) reveals that the prefix [*a*-] is actually the result of the cooccurrence of multiple higher applicatives. In cases such as these, the findings of syntactically-driven analyses may be incorporated into purely morphological analyses such as that of PCM, given that they may shed light on instances of blocking, both semantic and morphological.

However, a crucial difference between syntactic analyses and the PCM one here is that the hierarchies present in each are of a fundamentally different nature. Hierarchies are encoded in syntactic representations, as smaller elements are embedded within larger constituents to form a larger unit (e.g. NPs and VPs). Within these hierarchies, however, the phenomena occurring beneath the surface level are the primary focus of study, as

these processes derive the surface-level form and provide insight into how these forms may align with, or in some cases shape, our understanding of universal grammar. Figure 9, a syntactic representation from Demirok (2014), illustrates the hierarchical structure of applicatives occurring within or above the VP:

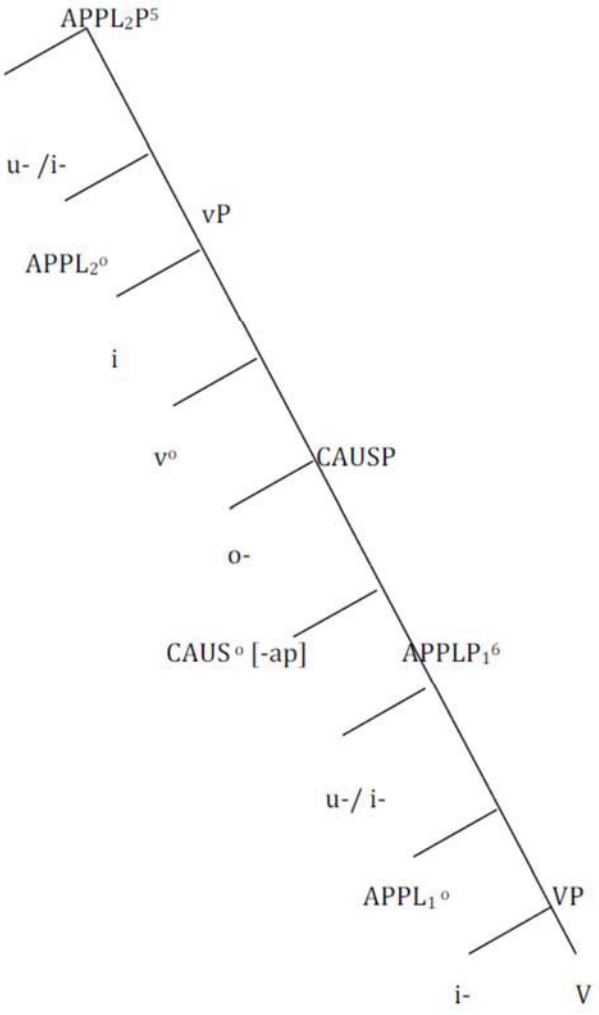


Fig. 9 Syntactic hierarchy of applicatives (Demirok, 2014)

The representation in Figure 9 shows the location of the various valency markers within the verbal structure and supports the interpretation of the morpheme [*a*-] to show multiple applicative exponence of [*u*-] and/or [*i*-] at various levels.

PCM, on the other hand, takes an admittedly more myopic view of structure, as the underlying structures and transformations are given no examination. Rather, linear surface position and instances of complementary distribution and asymmetrical blocking within the surface form constitute the sole motivations behind assigning a given morpheme to a particular level of dominance. PCM in this sense merely seeks to explain ordering and blocking phenomena based on empirical evidence from a given language.

Clearly, a PCM analysis has many shortcomings and lacks the depth and scope of many other models, and as such is it often passed over in favor of more far-reaching or in-depth models. However, I argue that this perceived 'weakness' of PCM may also be its strength. In focusing on the "what" rather than the "why" or the "how", the bridge from form to meaning is much more direct and accessible. In Chapter 6 I will discuss the ways in which a PCM analysis may be connected to a computational model in the attempt to create an internet-based learning resource for Laz (and other endangered languages).

5.5 Final remarks

Through examining the Laz verbal template in detail, we have established a two-dimensional model which combines linear precedence with dominance rankings and reflects morphological relationships between all verbal affixes, including a subhierarchy which illustrates the attachment of person markers. Figure 10 demonstrates the final version of this hierarchy:

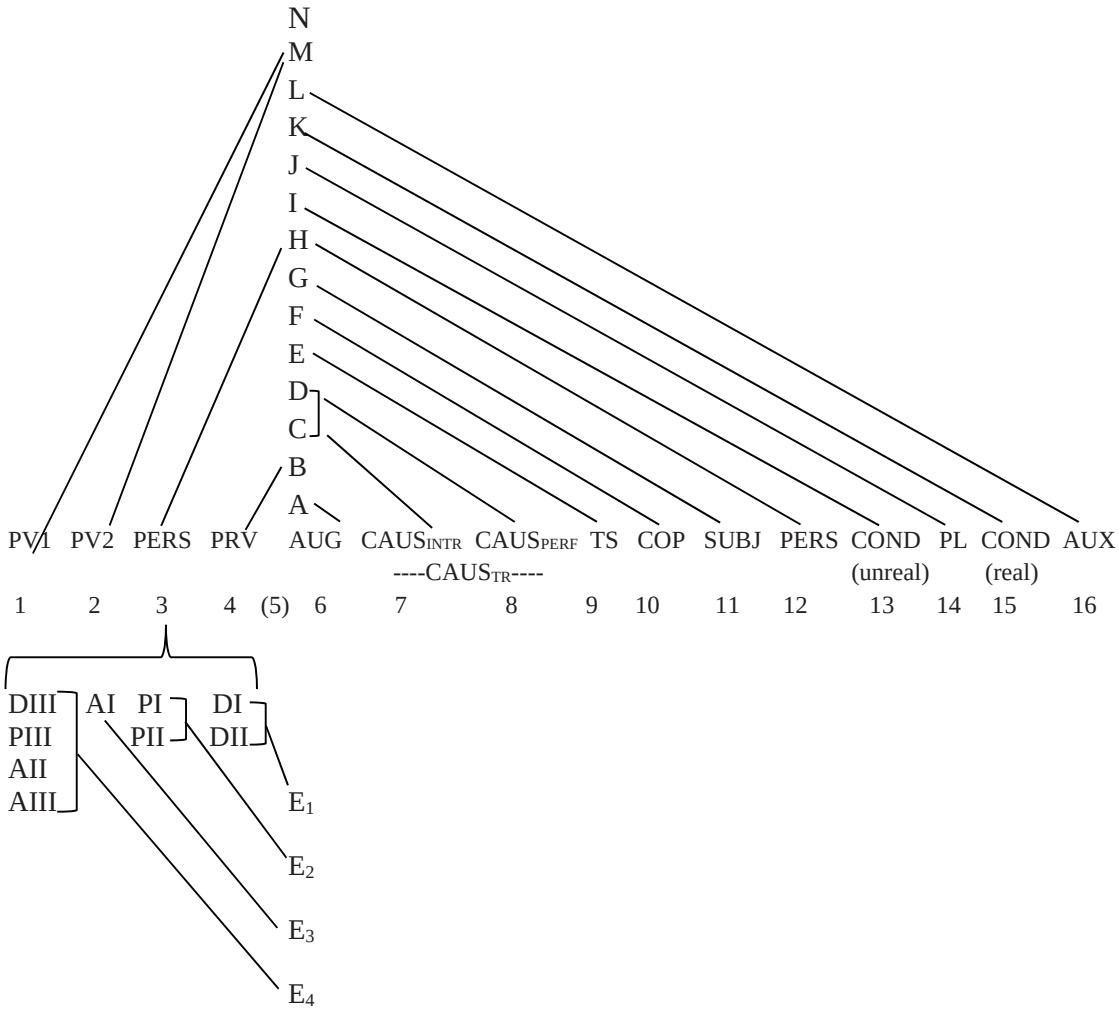


Fig. 10 PCM hierarchy of Laz verbal morphology

Having captured the complex relationship between Laz verbal affixes, we may now move on to the implementation phase of this project. In Chapter 6, the application of this model to a potential online lexical resource will be explored in order to bridge the gap between form and function and facilitate the creation of practical and meaningful tools for the fields of language preservation and revitalization.

CHAPTER 6

APPLICATION

Having discussed the morphological structure of Laz and how it may be represented within the framework of position class morphology (PCM), we now turn to determining how morphological phenomena may be accounted for in a web-based resource and considering how PCM might be incorporated into this representation.

This chapter will discuss the incorporation of the theoretical analysis into a functional learning resource. First, the challenges of creating lexicons for morphologically rich languages will be discussed, as well as ethical considerations for the creation of such a resource. Next, several existing computational approaches will be detailed and their applicability to the task at hand will be determined. Finally, a plan for the implementation of an approach for Laz will be laid out.

6.1 Challenges in creating a revitalization-driven lexicon for morphologically complex languages

Specifically, the problems Laz dictionary learners may face include orthography and morphological complexity. Firstly, let us examine the case of the former. Laz primarily utilizes the Latin alphabet,[39] but includes several additional letters as well as diacritics, as shown in Table 8:

[39] Note that varieties of Laz spoken in Georgia employ the Georgian script, but these varieties represent a small minority of Laz speakers and are not the focus of the present study

Table 8. Laz Orthographic Symbols

Stops	p	p̌ (or p’)	b	t	ť (or t’)	d	k	ǩ (or k’)	g	q
Affricates	ʒ (or ts)	ǯ (or ts’)	ž (or z’)	ç	ç̌ (or ç’)	c				
Fricatives	f	v	s	z	ş	j	x	ğ	h	
Nasals/Liquids/Glides	m	n	l	r	y					
Vowels	a	e	i	o	u					

As shown in Table 8, several letters have multiple manifestations, due in part to the lack of complete standardization in Laz and in part to the lack of widespread Laz keyboard shortcuts. A casual computer user may have difficulty typing with non-Latin characters. Furthermore, learners of the language often have trouble differentiating between forms such as glottalized and aspirated consonants (p̌, ť, ǩ, ǯ, ç̌ and p, t, k, ʒ, ç, respectively) or between voiceless uvular and velar stops (k/ǩ and q, respectively), so spelling errors are relatively frequent.

These orthographical challenges are minimal compared to those faced by less phonetic languages, and the types of adjustments that may be made to a lexicon can be found in Junker & Stewart (2008), in which the authors develop a search engine for the Eastern James Bay Cree language of Quebec, Canada, implementing what is known as a “relaxed” dictionary search. Algorithms were developed to account for misspellings and alternations by taking into account the common misspellings resulting from vowel length, antiquated spellings, etc. These algorithms were based on a customized version of Levenshtein Distance, identifying substitutions, insertions, and deletions of previously specified letters. This type of algorithm can easily be applied to a Laz lexicon and allow a user to find a form such as *ǩaťu* (‘cat’) even after entering the misspelled

katu. In the case of a minimal pair such as *qvali* (‘cheese’) and *xvali* (‘mucus’), both forms would be returned in a search regardless of which form had been entered.

The problem of morphological complexity is less easily solved. The highly affixing nature of Laz verbs enables a single root to have thousands of possible conjugations. This issue is not unique to Laz, however, and may be even more pronounced in the case of polysynthetic languages. It is obviously neither practical nor feasible to store every form of every verb in a lexicon, but as of yet there is no widely-accepted model for the optimal treatment of inflected forms in a lexicon, especially within the TD genre.

6.2 Ethical considerations for language revitalization tools

One of the most challenging aspects of this thesis—and any language revitalization work—is identifying the intersection of theory and practicality. In considering the goals of many language documentation and revitalization efforts, an ideal scenario would place power in the hands of language activists and/or in the target language community, some of whom may have minimal training in the field of linguistics. However, drawing upon a given theoretical framework for analysis often necessitates knowledge that is only attainable from years of education, and this may create a barrier to practicality for the target community.

How then should a linguist proceed in the attempt to integrate form with function, or, more precisely, to make use of a somewhat complex theoretical model in order to represent various phenomena which will in turn be incorporated into a community-driven, media-based lexical resource? The creation of one such resource for one such language may be feasible, but the attempt to replicate this process for multiple

resources across multiple languages will likely fall beyond the reach of more casual or untrained linguists.

It is therefore of the utmost importance that the analytical framework and ultimate production of this lexical resource be as simple as possible, yet as far-reaching as to be able to accommodate as many phenomena of natural languages as possible in order to provide accessibility to others wishing to create revitalization tools for other endangered and/or minority languages. Having noted this point, we may now proceed to an examination of several computational approaches to complexity in the search for the most universally-applicable model. Throughout this process, we will keep in mind the accessibility of these tools to would-be language activists in minority language communities.

6.3 Morphology-based computational approaches

Two relevant systems currently in use for dealing with morphological structure include algorithms which match affixes to paradigms (Junker & Stewart, 2008) and finite-state representations of morphology[40] (Snoek et al, 2014). These will now each be explained in more detail.

6.3.1 Paradigm matching algorithms

In this approach, a complex form is identified in a verb paradigm database (which in the case of Cree contains all observed verb forms from years of fieldwork), and algorithms isolate the affixes from the verb stem, returning the canonical form found in the

[40] The application of finite states to linguistics was introduced by Chomsky (1958) and advanced in the computational linguistics work of Koskenniemi (1983)

dictionary. Paradigm matching can provide results quickly and accurately and does not require a prohibitive computational load and is useful in translation from the target language to the source language.

However, the substantial database entailed in the paradigm matching approach is prohibitive for many lesser-described languages, in that extensive time and resources are required to assemble and annotate a lexical corpus containing thousands of word forms. This lengthy and labor-intensive process may not be an option for language activists or documentary linguists who wish to create a learning-driven lexical resource. It is thus not an appropriate choice for Laz or for a universal approach but may be valuable for more well-documented languages and/or those which have been the subject of extensive fieldwork.

6.3.2 Finite-state machine technology

This approach, also described as the "two-level" (i.e. input and output) approach, has enjoyed both success and popularity and has been applied to languages with rich morphologies such as Finnish (Koskenniemi, 1983) and Turkish (Oflazer, 1994), among others. A finite state machine (FSM) is a mechanism which, by means of automata, navigates through received input from left to right and ultimately judges acceptability of that input based on a series of previously specified states[41] which includes a set of final states. If, after undergoing the computation process, the automaton does not reach one of the final states from the specified set, the input is rejected as ungrammatical.

[41] These states should not be confused with stages in word formation—finite state automata only function based on the actual manifestation of the input (i.e. the final state) and are not concerned with the path to word formation

Finite state automata can recognize all regular languages (i.e. Type-3 grammars, as per Chomsky, 1963). This contains some, but not all, of the characteristics of natural language and an FSM could therefore easily reject a given natural language input as ungrammatical due to a variety of the phenomena present in natural languages. An FST (finite-state transducer), however, takes the FSM one step further in that rather than judging grammaticality of input, it can map form to function (i.e. morphemes to their features), allowing it to produce an output. A finite state transducer (FST) would therefore be useful in the analysis as well as generation of word forms.

FST technology is not without its limitations, however. Due to its strict left-to-right processing of words, non-concatenative languages such as Arabic have been historically problematic, although not ignored—see Attia et al. (2011) and Beesley (1996). Furthermore, bidirectionality may present an obstacle, as moving between the states from, for example, Laz to English, would not be the direct reversal of a path from English to Laz due to morpheme ordering—the mapping of a root and features to a Laz form would not progress in the same order as those of an English form due to the varying morphologies of the two languages. This issue is compounded when another language (in the case of Laz, Turkish) is added to the mix, as yet another morphological system comes into play.

Even with its shortcomings, however, FSTs may provide the simplest and most straightforward computational approach to incorporate into a basic learning resource. We will shortly return to examine these transducers in more detail.

6.3.3 Alternative methods

Other methods of morphological processing commonly used in machine translation include neural networks, such as that of Google Translate. However, these approaches are statistical in nature and require an extensive bilingual corpus. The large scale and computational load of technology coupled with the lack of linguistic data makes it prohibitive to minority language for which sufficient resources are unavailable. Due to these shortcomings, these large-scale alternative methods will not be considered in the discussion of an appropriate approach to be used with Laz and/or other endangered languages.

6.4 Identifying a suitable approach for Laz

Having examined several of the major (and minor) approaches to computationally addressing morphological complexity, we must now identify the optimal technology to be used in the representation of Laz in an online lexical resource. Before doing so, however, let us first clearly lay out the goals and scope of this lexical resource project.

6.4.1 Goals and scope of the talking dictionary

As a resource which has the ultimate goal of facilitating language revitalization, the goal of such a talking dictionary must be to benefit the language community firstly, most specifically the language learner, and must keep in mind the ethical considerations discussed earlier. Although a resource aimed at other audiences (e.g. linguists) has its own benefits, the primary goal of such a dictionary must keep in mind the needs of those for whom it was designed.

Continuing along this vein, we must pose the following speculative question: who is the average Laz language learner? To briefly restate the demographic information in a previous chapter, Laz is natively spoken primarily by the denizens of the Black Sea region of Turkey. It would then follow that the majority of language learners would be speakers of Turkish, possibly with ethnic or emotional ties to the Laz people. Outside of this group are those in academia studying Laz for research or academically-motivated purposes, including students in a primary or secondary school setting or academicians or tertiary students at a higher level. Additionally, many participants in the Haznedar (2018) report self-assess their ability to read and write in Laz to be significantly lower than their ability to speak and understand, and thus users of a resource from this group may seek to improve upon any existing literacy.

Given this hypothesized group of language learners, it is important to identify characteristics shared by the majority of members of the group in order to formulate an approach tailored to their needs. There is no single trait that will necessarily be present in every learner, apart from their goal of advancing his or her knowledge of Laz. The skill level, education level, and knowledge of grammar and/or linguistics may vary widely between individuals in the group. Therefore, in terms of grammatical metalanguage and linguistic terminology, the simplest approach and presentation which assumes minimal linguistic knowledge on the part of the user, will be the most effective in that it will be usable by the largest subset of learners. Although there may also be much to be gained by including technical or non-learner-oriented information, this should not be the primary focus of the modified talking dictionary.

Moving on from the anticipated audience for the talking dictionary, we now consider the scope of the approach. Although the talking dictionary in this project will

be created exclusively for the Laz language, the approach chosen should be applicable to the widest possible variety of languages in that it should be both easily replicable by members of other language communities as well as able to address the widest possible degree of morphological variation, including but not limited to the traditionally problematic polysynthetic and non-concatenative morphologies.

An ideal talking dictionary, then, would provide learners with the ability to perform a bidirectional lexeme search between a given language and the target language, limited not only to morphologically simplex forms. It would also, by means of a simple and intuitive framework, equip learners with the basic building blocks of the language, both morphological and syntactic. In keeping with the archetypal talking dictionary, as many word entries as possible would have corresponding sound files allowing users to hear a given word pronounced by a native speaker. All these criteria combined would allow learners to independently build their own proficiency in the target language, phonetically, morphologically, and syntactically.

Of the approaches identified in 6.3, finite state technology is best equipped to handle the challenges presented by the Laz data as well as those of other languages. The following section will more closely examine how a finite state transducer could be applied in the context of a Laz dictionary. Alternative solutions will be discussed in the case of shortcomings on the part of the finite state technology.

6.4.2 Finite state machines and transducers: Application to Laz

The finite state machine, as described earlier, is a relatively simple construction. Unlike the pushdown automaton and Turing machine, the FSM, as its name suggests, is built under the assumption that there are a finite number of states that the machine can assume

when moving from the initial to the final states. The memory of an FSM at any given point is limited to the information contained in the previous state: for example, an FSM that has just passed from a hypothetical state X to state Y to state Z will not retain the information of being at X once it reaches Z. The machine will merely proceed through the states in a left-to-right fashion until it is either blocked, causing it to reject the input, or reaches a final state. It is therefore critical to predetermine the states and the connections between them in order to prevent the FSM from proceeding down an ungrammatical path.

While an FSM is effective as a recognizer, it does not produce any other output than an acceptance or rejection of the provided input. As such, it is not particularly valuable to a language learning resource. A finite state transducer (FST), on the other hand, can recognize the components of a given input as it progresses from the initial to the final state, ultimately producing an output. An FST can be used both in analysis of an input, such as breaking down a complex form into its individual morphemes, and in generation, the reverse process in which a form is created in the target language based on the received input. Figure 11 example shows a simple FST for a morphologically complex English form.

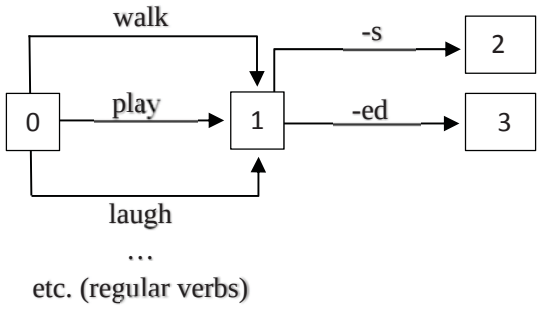


Fig. 11 Sample FST

In Figure 11, each form starts out in the empty initial state (here denoted as ‘0’). The transducer must then progress through a prespecified path which is shaped by the morphemes it encounters while analyzing a form. For example, in Figure 11, the arrival at state 2 prohibits the possibility of arriving at state 3, as there is no path between these states (and indeed no grammatical possibility of encountering a form such as “laughsed”).

In the following sections I will argue that two significantly different FSTs are necessary to capture translations from Laz (into glosses) and translations to Laz (from glosses). Whereas the generation of a glossed Laz form[42] may be captured in terms of precedence (i.e. left to right) by progressing through morphemes and providing their morphological information, generation of a Laz form (based on ordered morphological features collected via English or Turkish) should rely on the dominance hierarchy established in the previous chapter. I will now demonstrate how these models may be put into place.

6.4.2.1 From Laz (analysis)

Figures 12 and 13 illustrate a basic FST model for an intransitive Laz verb with an indicative aspect (for prefixes and suffixes, respectively).[43] [44]

[42] This resulting gloss may be expressed in either English or Turkish and further simplified for intelligibility

[43] Note that this model does not address allophones or phonotactic variation. This is a secondary issue to the FST designed here, as various automata may be implemented to address alternations resulting from spelling or allophonic variation.

[44] Separate diagrams have been created here for prefix + root stems and from root + suffix stems for aesthetic reasons, since the obligatory root serves as a dividing point and allows for a clearer representation. A functioning computational model would not make this division.

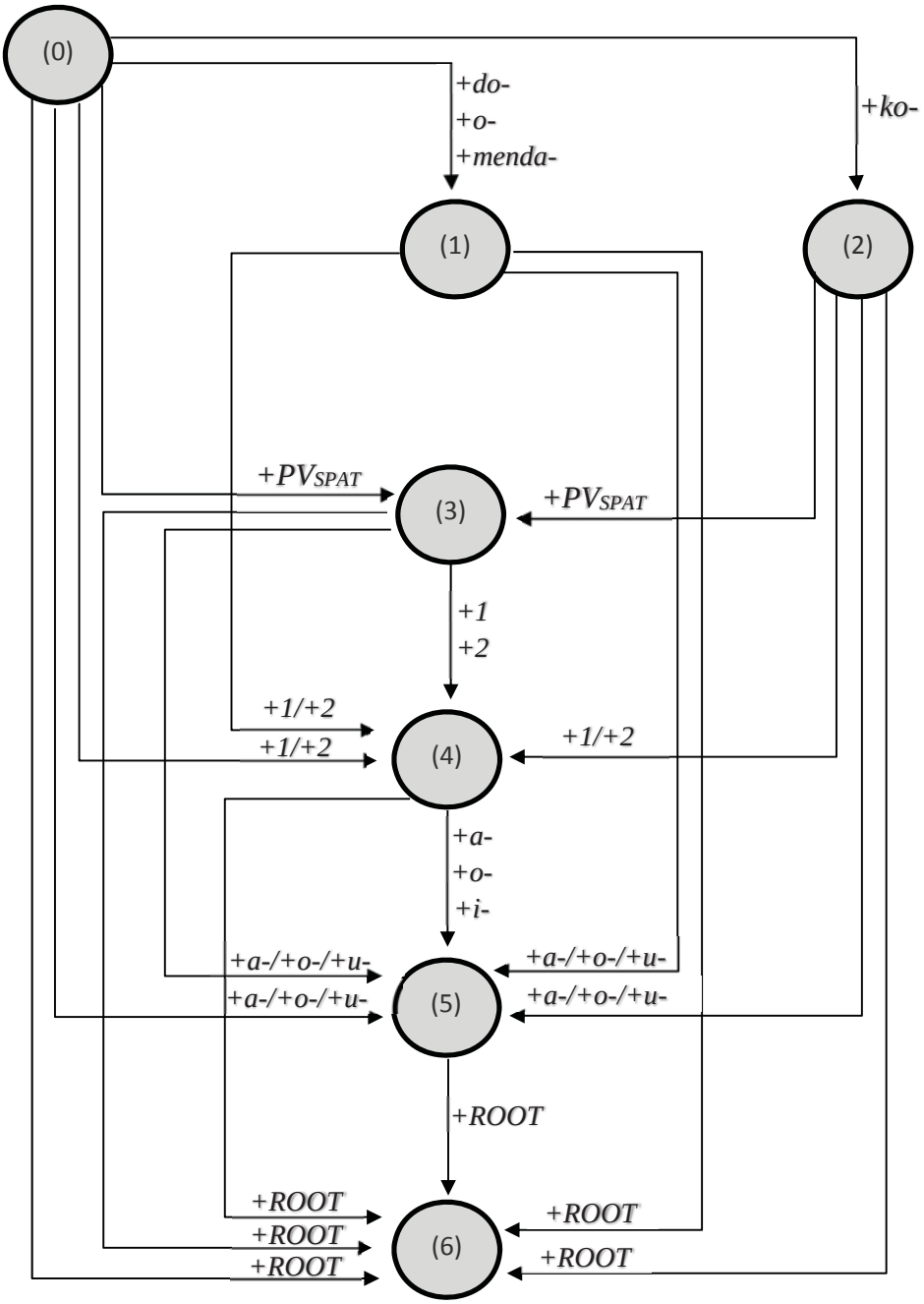


Fig. 12 Laz FST (prefixes)

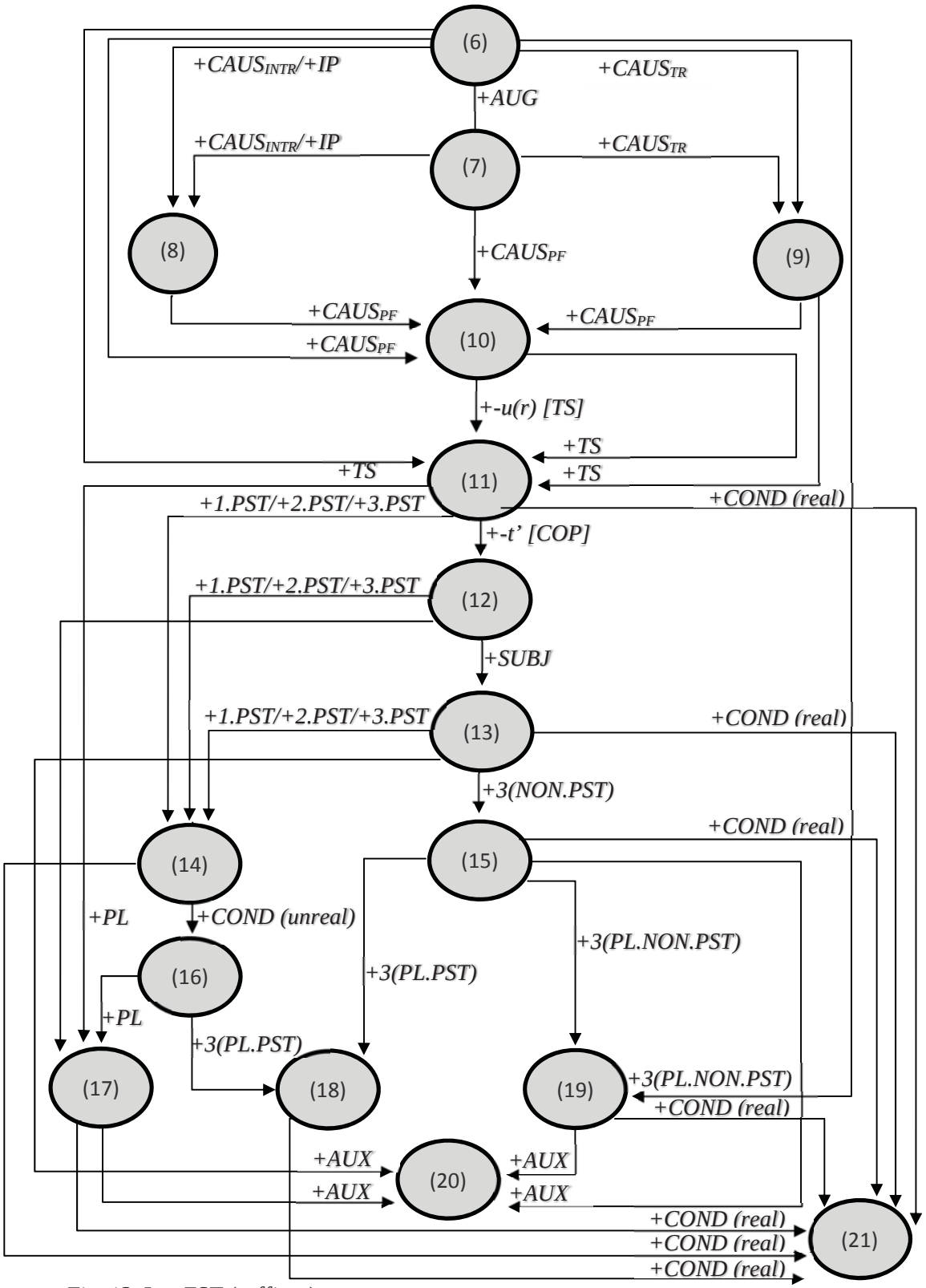


Fig. 13 Laz FST (suffixes)

In the first model, the parsing of any verb begins in the initial state, identified by "(0)" in the upper left-hand corner. As we observe the FST mechanism at work, the verb proceeds down a path to its final form through each additional affix processed between states. Again, note that this path has nothing to do with the rules of inflection and thus does not interact with rule ordering—it merely disassembles the input verb from left to right, and upon reaching the final state we will be able to analyze the verb by seeing its component parts (i.e. its gloss).

As an example of this process, if the transducer were to be applied to a form such as *dopskudatere* ('we will certainly live'), the component morphemes would be identified state by state along the following path in Figure 14 (taken from the models in Figures 12 and 13):

Input: do-p-skud-a-t-ere
PV_{AFF}-1-live-SUBJ-PL-AUX
'We will certainly live'

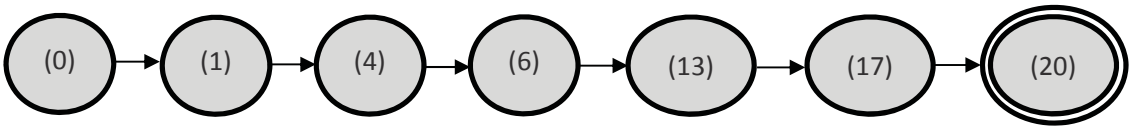


Fig. 14 Sample Laz-to-feature FST

In the representation in Figure 14, a new affix is added during the transition to each new state until the verb reaches its final state, denoted by a double circled node. During the parsing process, the transducer never visits states that are unrealized—as in PCM, only the surface forms are relevant for analysis.

In terms of generating Laz verbs (i.e. translating from Laz to a gloss which could then be realized in English and/or Turkish), the transducer would achieve relative success in that the morphological information in the output could be restated in a way that could be understood by the average language learner, although this would require an additional mechanism to translate between glosses and language. Although the process of determining the path through the stages would be labor-intensive and the paths themselves could be convoluted and labyrinthine, the result of such a process would be a fairly simple route to accessibility for thousands of forms. This Laz to English directionality could be useful for Laz speakers seeking a translation of a word form as well as language learners who have encountered an unknown written or spoken form.

6.4.2.2 To Laz (generation)

Conversely, however, the path from morphological features to Laz would necessitate a complete reworking of the computational path of the FST through the states, despite the deceiving simplicity of fact that the initial and final states would merely be reversed. However, retracing the previously identified path is not possible due to the fact that the ordering of incorporated morphological information becomes a key component in generating Laz forms (an issue which does not occur in parsing of existing Laz forms consisting of already ordered affixes).

To illustrate this, let us consider the previously analyzed example *dopskudatere* (‘we will certainly live’). The final state for this form is at the addition of the auxiliary suffix *-(e)re*. It is at this point that the first problem arises—with no clues from past states to dictate the correct form (i.e. *-re* or *-ere*), the FST will be unable to account for any dependencies and resultingly will be unable choose the correct path through the

states. It follows that the more complex the FST model, the higher the likelihood that situations such as these will arise.

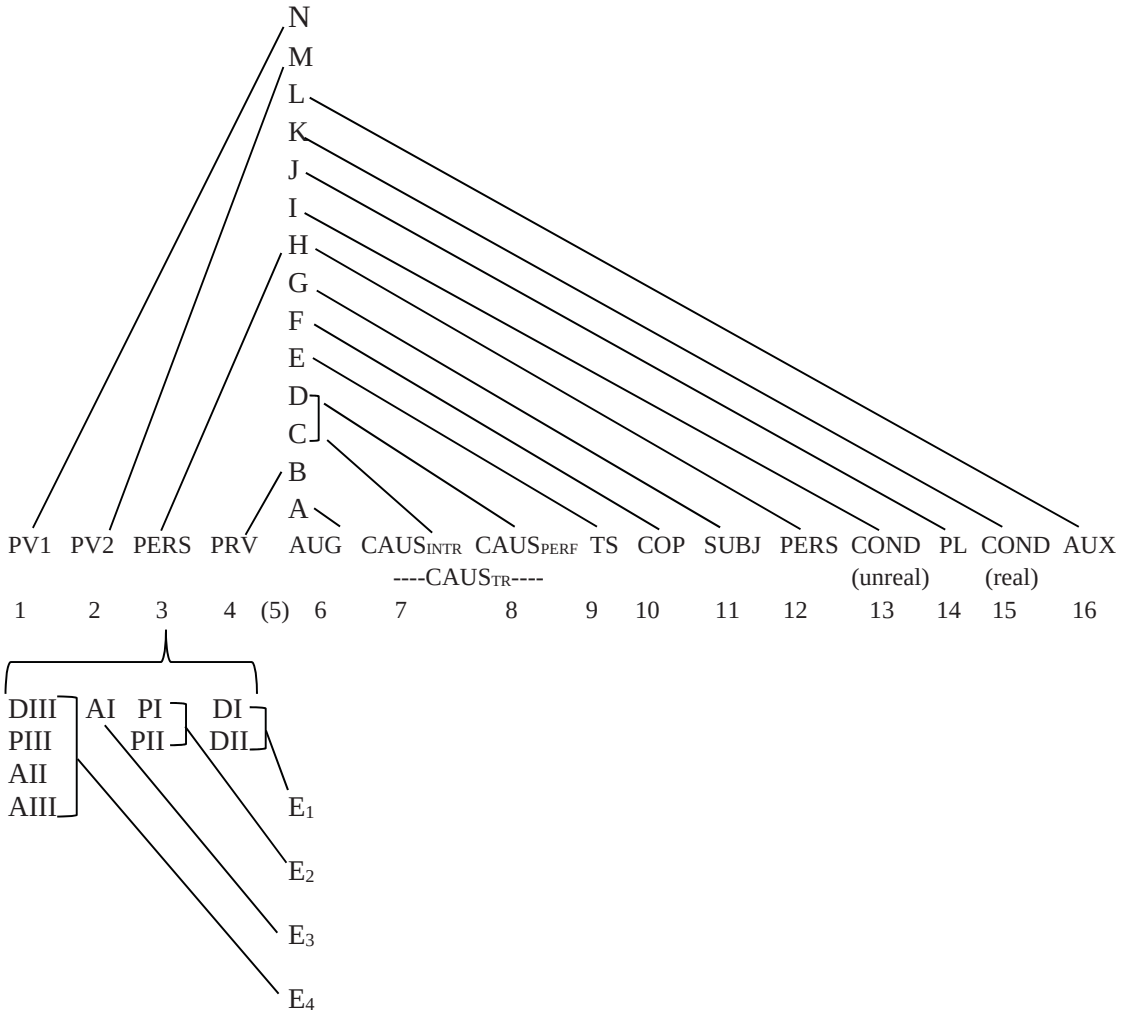
One alternative to retracing the steps of the existing FST would be to create a new transducer for the generation of Laz forms, which would require relevant morphological features to be specified in a certain order (rather than a direct translation from English or Turkish). For example, the information needed to produce *dopskudatere* would be the root "live" as well as the relevant information for the affixes: [+AFF], [+1], [+PL], and [+FUT]. However, we again encounter an issue similar to that of moving backward through the steps—as the FST is unable to predict the future states, it will not be able to assign the correct affirmative prefix without knowing whether a following spatial/directional prefix will be realized (as the affirmative prefix *do-* is most productive, while its counterpart *ko-* for the most part precedes spatial/directional prefixes).[45]

Given the challenges associated with a gloss to Laz FST, I propose that our findings from the previous analysis (i.e. the PCM hierarchy of Laz) are the key to cracking the puzzle of the generation of Laz forms for a lexical resource. Although in parsing (i.e. from Laz to a gloss), left-to-right directionality is viable, this is not the case with generation of Laz verbs. Rather, the input of morphological information must be ordered with the dominance of each morpheme in accordance with the PCM hierarchical model. In this way, problems resulting from the phenomenon of morphological blocking will be avoided, as more dominant affixes will automatically eliminate the option of incompatible, less dominant affixes.

[45] Note that *ko-* may also precede certain roots, and that the less productive affirmative prefixes *menda-* and *o-* have not been included here.

Let us once again examine our PCM hierarchical model and discuss how to apply the findings from this analysis to a finite-state approach:

(1)



Drawing upon the hierarchical information in this model, Table 9 will demonstrate the ideal ordering of featural input for the generation of a Laz verb:

Table 9. Laz Morpheme Order (Dominance-Based)

1	root
2	augmentative (AUG)
3	valency/applicative (PRV)
4	causative: intransitive/transitive ($CAUS_{INTR}$/$CAUS_{TR}$) impersonal passive (IP)
5	causative: perfective ($CAUS_{PERF}$)
6	thematic suffix (TS)
7	subjunctive (SUBJ)
8	copula (COP)
9	person markers (PM)
10	conditional: unreal (COND)
11	plural (PL)
12	conditional: real (COND)
13	auxiliary (AUX)
14	preverb: directional, spatial (PV_{SPAT})
15	preverb: affirmative (PV_{AFF})

Based on this model of dominance, we may posit a new FST that follows the hierarchical model in converting features into morphemes (i.e. generation) in Laz. In Figure 15 we see a partial model, illustrating the initial stages of such a transducer. In this transducer, only dominance classes A, B, C, and D are represented in order to demonstrate the workings of the model (with all future morphemes designated by '+OTHER' en route to unspecified states). The formulation of a finalized model would be equally as labor-intensive and the paths equally labyrinthine as those described in the Laz-to-gloss section, requiring intensive data collection, native speaker input, and computational troubleshooting, thus exceeding the bounds of this thesis. However, the

partial model here sufficiently illustrates the potential for interaction between PCM and FST technology.

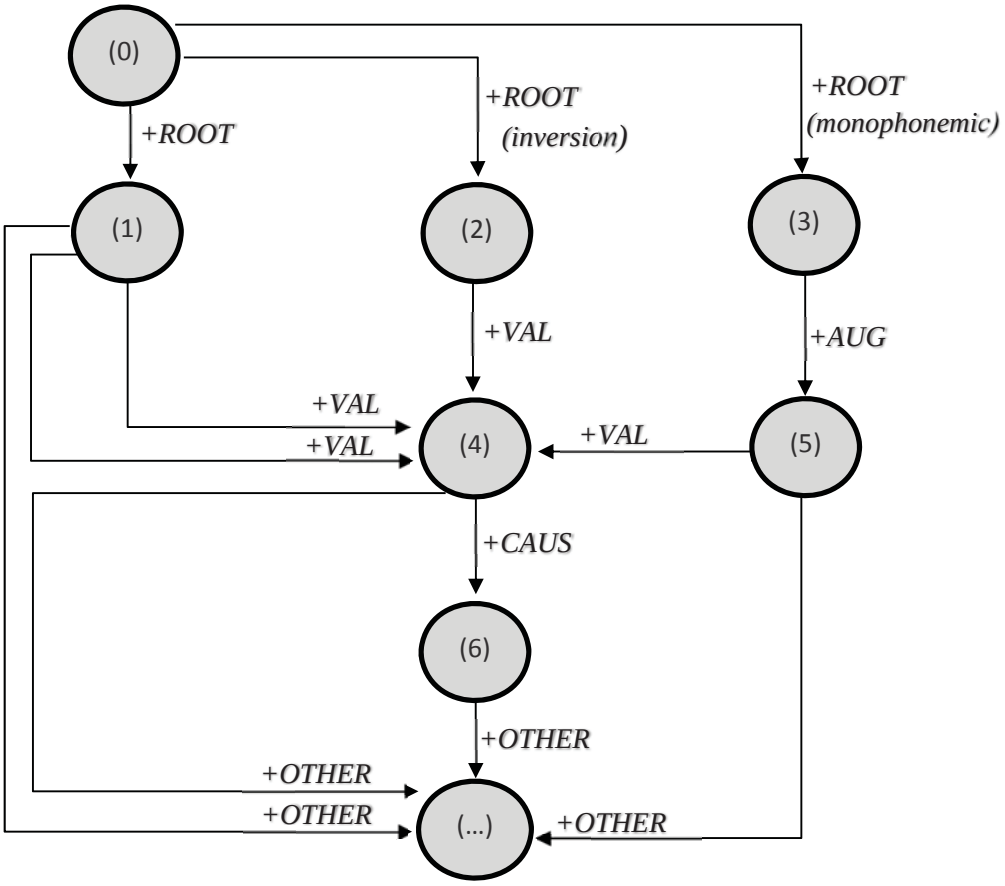


Fig. 15 Features-to-Laz FST

The ordering of the FST in Figure 15 is drawn directly from our previous PCM analysis but incorporates more detail within each position class. For example, the transducer in Figure 15 differentiates between three different types of roots, all of which must be specified in the database from which they will be drawn. All roots classified as taking experiencer subjects will be routed through a selection of valency marker, whereas this path is optional for agentive verbs. Another subcategory, single consonant

roots pass through state 3 to state 5, during which time they are assigned an augmentative stem formant.

In this way, the environment of all possible morphemes may be specified, creating a clearer path to a final result than would a precedence-based model—all morphological blocking is accounted for as the transducer moves through the dominance-ranked states in the order specified here. The selection of a particular feature will guide the path of the transducer to the appropriate state, thus streamlining the process of verb generation.

This dominance-based model shows superiority over a precedence-based model in other ways as well. Firstly, an obstacle to precedence-based models is that some important information for selecting less dominant affixes will not be available without first knowing details of more dominant ones. To exemplify this, we may consider dominance classes B and H (i.e. pre-root vowels and person markers, respectively). As the choice of person marker is partially dependent on the argument structure of the predicate, there is no way to know of this argument structure without first obtaining information about the nature of the predicate. For example, verbs denoting involuntary actions are known to demonstrate the “inversion” phenomenon, as in (2):

(2)

Ma	skani	gyari	m-a-çkom-u
1SG	2POSS	food	1-APPL-eat-3SG.PST

‘I (accidentally) ate your food’

In a precedence-based model, an accurate representation of a first person subject and third person object (in the past tense) for such a predicate would still be insufficient to assign a certain affix to this position, as multiple circumfixes would satisfy these criteria (e.g. *p̌-…-u* or *m-…-u*). A dominance-based model easily avoids this issue by implementing the relevant features before reaching any point of potential ambiguity.[46]

Obviously, such a tool is only as useful as its accessibility to users of the resource. As much metadata as possible should be encoded in the lexical database—for example, users of the resource should not have to specify the argument structure in the selection of a thematic suffix in the process of generating a verb form. Rather, the roots selecting a specific thematic suffix should be specified within the lexical entry itself, thus removing the burden of linguistic analysis from the user. This additional encoding of information could be useful in other areas as well, such as dealing with irregular verbs or those with multiple roots—the environment of non-standard forms could be specified and automatically selected during the generation process.

Before moving on to the consideration of future direction, let us briefly discuss the person marker circumfix, which was not included in the sample FST. Circumfixes have already been examined within the literature on FSTs, such as by Cohen-Sygal & Wintner (2006), who posit that FSRAs (finite-state registered automata)—an extension of finite-state automata which encode small amounts of memory for storing forms and rules—may more easily deal with non-concatenative morphology, including vowel

[46] Although a dominance-based model would derive a form in which affixes are in an incorrect order (i.e. reflecting dominance rather than precedence), an additional mechanism may be employed to correctly reorder the affixes.

harmony and reduplication. This research offers a promising direction for a broader future application of finite-state technology to a wider typology of languages.

6.4.3 Additional considerations

Additional modifications may be made to the afore-mentioned process as well, such as allowing the user to view a list of common conjugations for a given verb (e.g. showing various combinations of person or tense marking). A more comprehensive approach could entail using FST technology to generate all possible forms resulting from all possible paths. Each verb in the resulting database would have tags based on the affixes captured by the transducer, and the user could then search for forms by toggling on and off different features (e.g. $+1^{st}$ person, –past tense, etc.) However, the sheer number of possibilities generated may not be easily digestible for a casual learner who may be unable to sufficiently narrow down such a list.

A natural next step to the processes outlined above is the ability to move directly from Laz to English/Turkish as well as English/Turkish to Laz without the need for an intermediary gloss. However, this would necessitate a morphosyntactic model of English and Turkish, and while this falls within the realm of possibility it falls outside the scope of this thesis.

In addition to finite-state technology, I propose that additional resources be included in order to facilitate a learning experience for users that will promote morphological examination and awareness. These tools may vary from language to language. One example of a supplemental resource to be used in tandem with FSTs is a learning portal for morphological inflection, which could provide tutorials and thus serve as a learning tool to facilitate both language learning and efficient use of the verb

generation tool. This section could contain many pre-generated examples as well as voice recordings of full conjugated verbs and even sentences (adding the “talking” element to the talking dictionary). For a more basic analysis of forms, stemmers may be a promising avenue for a partial short-term solution to morphological complexity.

The process described here is by no means a panacea which can solve all problems of the representation of morphological complexity. However, it is basic enough that it can be adapted to describe many other languages as well. With time and resources, the development of a system which will allow bidirectional searching of morphologically complex forms will further the goal of revitalization by facilitating a smoother translation experience. However, with a more advanced system comes a more advanced required skill set and a larger workload. Unfortunately, it cannot be expected that minority communities will be able to provide or access these resources, and currently more short-term solutions such as the inclusion of verbal paradigms within TDs may be a more practical medium of preserving and revitalizing a language in a web-based resource.

6.5 Implementation

To this point, the analysis and application discussion have been of a theoretical and speculative nature—no precise guidelines have yet been laid out for the reproduction of the tool described in this thesis. In this section, I will briefly identify the steps that must be followed to create and utilize such a resource. Note that the actual development of the resource itself requires significant data collection as well as collaboration with a programmer and thus falls beyond the scope of this thesis.

6.5.1 Step one: Develop affixal templates

Firstly, any morphological complexity in the source language should be identified. Multiple parts of speech may exhibit various levels of morphological complexity—in Laz, the highly affixing nature of the verb cemented its role in the analysis. Once the morphological complexity has been identified, a template should be developed to assign the various affixes to position classes, firstly in terms of precedence. This may be done through a series of morphological tests, such as determining which affixes are in complementary distribution. In the case of Laz, a template may already exist in the literature, such as that in the grammar of Öztürk & Pöchtrager (2011). Keep in mind that existing templates may undergo modification during the application of a position class morphology analysis.

6.5.2 Step two: Apply PCM to establish a hierarchy

Once a template has been established and all affixes assigned to position classes in terms of precedence and distribution, an analysis of morphological blocking must be conducted to establish levels of dominance among affixes in keeping with the tenets of PCM. Any unexplained morphological phenomena should be examined at this point and accounted for within the PCM structure (such as the case of “inversion” in Laz verbs).

6.5.3 Step three: Create a lexicon

Before any morphological analysis may be conducted, lexical items must be gathered for the talking dictionary. These may be stored in a database such as Microsoft Excel, and should include metadata—e.g. part of speech, semantic information, etc. If a given root selects a particular affix for semantic reasons, such as thematic suffixes in Laz verbs,

this must be reflected in the lexical entry in the database. The lexicon may also contain examples of the word used in a sentence or appearing in a conjugated form.

Additionally, as the resource will ultimately be a talking dictionary, recordings of elicited lexical items from native speakers should be linked to each entry. However, as both the insertion of lexical items and the recording process may be lengthy (or access to native speakers may be limited, additional lexical entries and sound files may be added at any point in the process, even after the completion of the dictionary.

6.5.4 Step four: Create an FST for parsing (precedence-based)

Using the morphological template from step two, the first FST may now be produced by mapping all the possibilities of affixation from left to right and accurately capturing the relationship between states and identifying all final states. The implementation of this step will require the expertise of a programmer. Note that this step refers to the parsing of forms in the source language, providing morphological information that may be converted into the desired majority language of the user.

6.5.5. Step five: Create an FST for generation (dominance-based)

In the same way that the precedence-driven FST was created in step four, a dominance-driven FST may be created for the generation of forms in the source language. This will be a labor-intensive process, as it will entail a complete examination of the distribution of all morphemes within a template. As with step four, the process of implementing this step will require programming knowledge. The execution of this step will allow users to generate their own morphologically complex word forms from morphological features.

6.4.6 Step six: Create any supplemental resources

After both levels of transducers have been created, any additional supplementary teaching materials can be added for the benefit of users. These will be customized to the needs of the target language and may be static or interactive. Examples of these supplements include a how-to guide for inflecting forms, an explanation of tones (if any), and an interactive verb portal allowing users to explore various affixal positions and learn about the morphology of the target language. Aside from morphological supplements, the inclusion of other information, such as syntactic and cultural (e.g. recordings of stories, songs, etc.) would further facilitate the language learning experience.

CHAPTER 7

CONCLUSION

7.1 Contributions to the field of linguistics

In this thesis I have laid out a blueprint for an “expansion pack” to the genre of talking dictionaries. Although it is uncommon for lexical resources to contain a large amount of morphological information, I argue that the inclusion of a strategy for dealing with various types of complexity is necessary for the field of language preservation and revitalization. After all, the knowledge of a given verbal root or infinitive form is only as valuable as far as it is useful to the language learner. This issue is exacerbated by the fact that language learning resources for endangered languages may not be readily available or accessible to online learners.

I argue that, in cases of language endangerment, any resource is obligated to be maximized to its full potential, keeping in mind the goal of user friendliness and practicality. Furthermore, the process for language activists to create resources for their own languages should not be prohibitively difficult or opaque. The blueprint designed here, while it does involve a substantial level of education in linguistics as well as computer programming, has been purposely created to be a simplified process. Though it will require both training and collaboration, the process described in this thesis is ultimately within reach for those who wish to make a difference within their own language communities.

7.2 Summary of claims and findings

In this thesis, I have endeavored to bridge the gap between morphological complexity and revitalization-driven lexical resources (i.e. the talking dictionary). I have asserted that, in the interest of functionality, the talking dictionary genre should be expanded to account for morphological complexity in order to facilitate users who may not have access to the target language in a classroom setting.

In determining the necessary steps to create an augmented talking dictionary, I relied heavily on morphological analysis of Laz through the framework of position class morphology, which seeks to identify the hierarchical relationships between affixes in a stem. This framework was chosen because of its simplicity, its applicability to a wide variety of the world's languages, and the ease with which it coordinates with FST technology.

In a PCM analysis of Laz verbal morphology, I developed a hierarchical model to represent the Laz verbal template, and this model was able to capture all instances of morphological blocking. Some phenomena, such as that of subject-object inversion, required reinterpretation in order to conform to this analysis.

Having created a hierarchical model, I turned to the consideration of a technological framework that could support the principles of PCM. I found that finite state technology was able to adequately support the PCM hierarchical model. I thus created an FST model that illustrated the paths through various states in the affixation of a verbal stem, both in terms of precedence when parsing Laz forms and in terms of dominance when generating Laz forms. Finally, I outlined the procedure the create such a model in a way that may be replicated with other languages in the future.

7.3 Future research

Although a tentative bridge linking form and function has been proposed here, the issue of morphological complexity as it relates to language revitalization-driven lexical resources is far from being definitively resolved. I hope that this thesis will serve as a starting point for further research on potential applications to address this issue.

Additionally, the topic of syntax has not been detailed in this thesis, but it follows from the points here that if a resource is to be valuable for language learners, it must also be useful. Combining words at the phrasal, clausal, and sentence levels is most certainly a critical component of language learning. Although a talking dictionary and a grammar are two entirely different categories, the fusion of online language resources may be a key to language revitalization.

This thesis reflects one perspective on integrating morphology with computational linguistics and language revitalization. I am aware that there are many more morphological and computational models and tools which may have successfully been implemented in the case of Laz. In the pursuit of simplicity and universality, it is possible that a superior method has been overlooked. It is my hope that the field of language revitalization garners even more attention in the coming years, and that issues such as the one explored in this thesis be prioritized by linguists of various disciplines in an effort to preserve and revitalize the world's endangered languages.

REFERENCES

Anderson, S. (1982). Where's morphology? *Linguistic Inquiry, 13*(4), 571-612.

Anderson, R. D. (1963). *A grammar of Laz.* (PhD thesis). University of Texas at Austin, Austin, TX.

Atlamaz, Ü. (2011). Cyclic agreement and empty slots in Pazar Laz. *Annual Meeting of the Berkeley Linguistics Society, 37*(2), 18-31.

Attia, M., Pecina, P., Toral, A., Tounsi, L., & van Genabith, J. (2011). An open-source finite state morphological transducer for modern standard Arabic. In *Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing* (pp. 125-133). Stroudsburg, PA: Association for Computational Linguistics.

Austin, P. K., & Sallabank, J. (Eds.). (2011). *The Cambridge handbook of endangered languages*. New York, NY: Cambridge University Press.

Beesley, K. R. (1996). Arabic finite-state morphological analysis and generation. In *Proceedings of the 16th Conference on Computational Linguistics* (pp. 89-94). Stroudsburg, PA: Association for Computational Linguistics.

Bonami, O., & Boyé, G. (2005). French pronominal clitics and the design of Paradigm Function Morphology. In G. Booij, et al. (Eds.). *On-line Proceedings of the Fifth Mediterranean Morphology Meeting* (pp. 291-322). Bologna, Italy: University of Bologna.

Brenzinger, M. (Ed.). (2008). *Language diversity endangered.* Berlin, Germany: Mouton de Gruyter.

Bucaklişi, I. A. (2014). *Lazca ders kitabı*. Istanbul, Turkey: Lazika Yayın Kolektifi.

Bucaklişi, I. A. & Uzunhasanoglu, H. (1999). *Lazuri-Turkuli nenapuna/Lazca-Türkçe Sözlük.* Istanbul, Turkey: Akyüz Yayıncılık.

Bucaklişi, I. A., Uzunhasanoglu, H. & Aleksiva, I. (2007). *Didi Lazuri nenapuna/Büyük Lazca sözlük*. Istanbul: Chiviyazıları.

Caballero, G., & Inkelas, S. (2013). Word construction: Tracing an optimal path through the lexicon. *Morphology, 23*(2), 103-143.

Carnie, A. (1996). Modern Irish: A case study in language revival failure. *MIT Working Papers in Linguistics, 28,* 99-114.

Chomsky, N., & Miller, G. A. (1958). Finite state languages. *Information and control, 1*(2), 91-112.

Chomsky, N. (1963). Formal properties of grammars. In Luce, R. D., Bush, R. R., & Galanter, E. (Eds.). *Handbook of Mathematical Psychology* (pp. 323-418). New York, NY: John Wiley.

CIDLES. (2013). *Laz*. Retrieved from http://www.cidles.eu/events/conference-ele-2013/language-fair/programme-1/laz/

Cohen-Sygal, Y., & Wintner, S. (2006). Finite-state registered automata for non-concatenative morphology. *Computational Linguistics, 32*(1), 49-82.

Demirok, Ö. F. (2013). *AGREE as a unidirectional operation: Evidence from Laz* (Master's thesis). Boğaziçi University, Istanbul, Turkey.

Demirok, Ö. F. (2014). *Valency vowels in Laz*. Retrieved from http://demirok.scripts.mit.edu/web/wp-content/uploads/2018/01/valency_vowels_in_Laz.pdf

Demirok, Ö. F., & Öztürk, B. (2015). The logophoric complementizer in Laz. *Dilbilim Araştırmaları Dergisi, 26*(2), 45-69.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2019). *Ethnologue: Languages of the World*. Dallas, TX: SIL International. Retrieved from https://www.ethnologue.com/cloud/lzz

Emgin, B. (2009). *Finiteness and complementation patterns in Pazar Laz* (Master's thesis). Boğaziçi University, Istanbul, Turkey.

Eren, Ö. (2016). *Spatial prefixes of Pazar Laz: A nano-syntactic approach* (Master's thesis). Boğaziçi University, Istanbul, Turkey.

Gurevich, O.I. (2006). *Constructional morphology: The Georgian version.* (PhD thesis). University of California at Berkeley, Berkeley, California.

Gürpınar, T. (2000). *The dialect of Pazar Laz and its case system* (Master's thesis). Boğaziçi University, Istanbul, Turkey.

Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale and S.J. Keyser (Eds.), *The View From Building 20*, (pp. 111-176). Cambridge, MA: MIT Press.

Hammarström, H., & Borin, L. (2011). Unsupervised learning of morphology. *Computational Linguistics, 37*(2), 309-350.

Harshav, B. (2009). Flowers have no names: The revival of Hebrew as a living language after 2,000 years was no miracle. *Natural History*, 24-29.

Haznedar, B. (2018). *Türkiye'de Lazcanın mevcut durumu*. Istanbul, Turkey: Laz Enstitüsü.

Hinton, L., & Hale, K. (Eds.). (2001). *The green book of language revitalization in practice*. Leiden, The Netherlands: Brill.

Holisky, D. A. (1991). Laz. In Harris, A. C. (Ed.). *The Indigenous Languages of the Caucasus* (pp. 89-154). Delmar, NY: Caravan Books.

Inkelas, S. (1989). *Prosodic constituency in the lexicon* (PhD thesis). Stanford University, Stanford, CA.

Inkelas, S. (1993). Nimboran position class morphology. *Natural Language & Linguistic Theory*, *11*(4), 559-624.

Junker, M. O., & Stewart, T. (2008). Building search engines for Algonquian languages. In K. S. Hele & R. Darnell (Eds.), *Papers of the 39th Algonquian Conference* (pp. 378-411). London, Canada: The University of Western Ontario.

Kutscher, S. (2001). Lazuri nena: The language of the Laz. Retrieved from http://www.uni-koeln.de/philfak/ifl/asw/forschung/projekte/Lasisch/benningwd.pdf

Kanien'kéha: An open source endangered language initiative. (n.d.). Retrieved from https://kanienkeha.net/

Kavaklı, N. (2015). Novus ortus: The awakening of Laz language in Turkey. *İdil Sanat ve Dil Dergisi*, *4*(16), 133-146.

Klein, W. (1992). The present perfect puzzle. *Language*, *68*(3), 525-552.

Kojima, G. & Bucaklisi, I. A. (2003). *Lazuri grameri*. Istanbul: Chiviyazilari.

KONDA. (2006). *Who are we? Social structure survey.* Retrieved from https://konda.com.tr/wp-content/uploads/2017/03/2006_09_KONDA_Social_Structure-1.pdf

Kontovas, N. (2015). *Lazca sözlük projesi*. Retrieved from https://www.nenapuna.net/

Koskenniemi, K. (1983). *Two-level morphology: A general computational model for word-form recognition and production.* Helsinki, Finland: University of Helsinki, Department of General Linguistics.

Kutscher, S. (2008). The language of the Laz in Turkey: Contact-induced change or gradual language loss? *Turkic Languages, 12*(1), 82-102.

Kutscher, S., Mattissen, J., & Wodarg, A. (Eds.). (1995). *Das Mut'afi-Lazische*. Köln, Germany: Humboldt University of Berlin.

Lacroix, R. (2009). Ditransitive constructions in Laz. *Linguistic Discovery, 9*(2), 78-103.

Landau, I. (2009). *The locative syntax of experiencers.* Cambridge, MA: MIT Press.

Laz. (n.d.). In *Encyclopedia of World Cultures*. Retrieved from https://www.encyclopedia.com/humanities/encyclopedias-almanacs-transcripts-and-maps/laz

Living Tongues Institute for Endangered Languages, (n.d.). *Talking dictionaries*. Retrieved from http://livingtongues.org/talking-dictionaries/

Minahan, J. B. (2016). *Encyclopedia of stateless nations: Ethnic and national groups around the world.* Santa Barbara, CA: Greenwood.

Moseley, C. (Ed.). (2010). *Atlas of the world's languages in danger*. Paris, France: The United Nations Educational, Scientific and Cultural Organization.

Oflazer, K. (1994). Two-level description of Turkish morphology. *Literary and linguistic computing, 9*(2), 137-148.

O'Rourke, B. & Walsh, J. (2014). New speakers of Irish: Shifting boundaries across time and space. *International Journal of the Sociology of Language, 2015*(231), 63-83.

Osborne, B., Jackson, P., Walsh, F., & Sanders, T. (Producers), Jackson, P. (Director). (2002). *The Lord of the Rings* [Motion picture]. United States: New Line Home Entertainment.

Öztürk, B. (2007). *The morphological and syntactic properties of case and agreement in Laz.* Istanbul, Turkey: Boğaziçi University Research Fund.

Öztürk, B. (2012). *Thematic suffixes and verb classes in Laz*. Istanbul, Turkey: Boğaziçi University Research Fund.

Öztürk, B. (2019) The loss of case system in Ardeshen Laz and its morphosyntactic consequences. *STUF - Language Typology and Universals, 72*(2), 193-219.

Öztürk, B. (to appear). Possessor applicatives in Pazar Laz. *Revue Roumaine de Linguistique*.

Öztürk, B. & Pöchtrager, M. (Eds.). (2011). *Pazar Laz*. Münich, Germany: LINCOM Europa.

Partee, B. H., Ter Meulen, A., & Wall, R. E. (1993). Finite automata, regular languages and type 3 grammars. In *Mathematical Methods in Linguistics* (pp. 453-484). Dordrecht, The Netherlands: Springer.

Rosen, G. (1844). *Über die Sprache der Lazen*. Berlin, Germany: Lemgo et Detmold.

Schuster, M., Johnson, M., & Thorat, N. (2016). Zero-shot translation with Google's multilingual neural machine translation system [Web log post]. Retrieved from https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Snoek, C., Thunder, D., Loo, K., Arppe, A., Lachler, J., Moshagen, S., & Trosterud, T. (2014). Modeling the noun morphology of Plains Cree. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 34-42). Baltimore, MD: Association for Computational Linguistics.

Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure*. Cambridge, United Kingdom: Cambridge University Press.

Tantuğ, A. C., Adalı, E., & Oflazer, K. (2006). Computer analysis of the Turkmen language morphology. In T. Salakoski, F. Ginter, S. Pyysalo, T. Pahikkala (Eds.), *Advances in Natural Language Processing* (pp. 186-193). Berlin, Germany: Springer.

Taşkın, N. (2011). *Representing and performing Laz identity: 'This is not a rebel song!'*. (Master's thesis). Boğaziçi University, Istanbul, Turkey.

Taylan, E., & Öztürk, B. (2014). Transitivity in Pazar Laz. *Acta Linguistica Hungarica, 61*(3), 271-296.

Taylor, C. (2015). Introduction: A Navajo hope. In *How Star Wars Conquered the Universe: The Past, Present, and Future of a Multibillion Dollar Franchise*. New York, NY: Basic Books.

Tuite, K. (1988). *Kartvelian morphosyntax: Number agreement and morphosyntactic orientation in the South Caucasian languages* (PhD thesis). University of Montreal, Montreal, Canada.

Vamling, K., & Tchantouria, R. (2010). Language use and attitudes among Megrelians in Georgia. In K. Vamling (Ed.), *Caucasus Studies 2* (pp. 81-92). Malmö, Sweden: Malmö University, School of International Migration and Ethnic Relations.