# Language Change, Contact, and Koineization in Pacific Coast Athabaskan

by

Justin David Spence

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Andrew Garrett
Professor Leanne Hinton
Professor William Hanks
Professor Victor Golla

Fall 2013

Language Change, Contact, and Koineization in Pacific Coast Athabaskan



by Justin David Spence

# Abstract

Language Change, Contact, and Koineization in Pacific Coast Athabaskan

by

Justin David Spence

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Andrew Garrett, Chair

The Pacific Coast Athabaskan (PCA) languages are part of the Athabaskan language family, one of the most geographically widespread in North America. Over a millennium ago Athabaskan-speaking groups migrated into northwestern California and southwestern Oregon from a northern point of origin several hundred miles away, but even after several centuries separated from other languages in the family and in contact with neighboring non-Athabaskan populations their languages changed only incrementally and maintained an essentially Athabaskan character. Beginning in the mid-19th century, disruptions associated with colonization brought closely-related PCA varieties into intimate contact both with each other and with English, leading eventually to their current state of critical endangerment. This dissertation explores these diachronic developments and seeks to understand (a) how the PCA languages are related to each other and to the rest of the Athabaskan language family, (b) the social and structural dynamics of dialect contact between Athabaskan varieties from the mid-19th to the mid-20th century, and (c) the linguistic consequences of bilingualism as Athabaskan communities shifted from their heritage languages to English.

Chapter 1 situates the research in a theoretical and methodological context that informs subsequent chapters. Special attention is paid to outlining competing modes of explanation for diachronic outcomes observed in dialect contact situations, comparing demographic approaches that make reference to population size or density versus indexical approaches that invoke people's strategic deployment of linguistic resources to achieve social aims. This chapter also provides an overview of the PCA languages, including their classification, the extent of their aboriginal territory, and the main sources of documentation consulted in this study.

Chapter 2 addresses questions related to linguistic phylogenesis that took place over a relatively long timespan, focusing specifically on the question of whether or not the PCA languages comprise a well-defined subgroup within the Athabaskan family. This chapter uses recent computational methods that have been adapted from the study of biological evolution to address analogous issues in historical linguistics. Using a combination of lexical, phonological, and morphological characters, the chapter considers how results obtained with this computational approach compare with findings from traditional historical-comparative methods, where it has been suggested that tree-like branching structures are inappropriate models of Athabaskan linguistic relationships. The main empirical finding of this chapter is that the PCA languages emerge as a well-supported subgroup of Athabaskan under a variety of conditions: with lexical

different model assumptions about rates of evolution across lineages. Other findings of interest include topological differences obtained using multi-state versus binary coding schemes and the extent to which lexical, phonological, and morphological characters give convergent results in phylogenetic analysis.

Chapters 3 and 4 address the linguistic effects of contact between Athabaskan populations that resulted when indigenous people of California and Oregon were dispossessed and consolidated on a small number of reservations in the mid-19th century. Some current approaches to dialect contact predict that koineization leading to leveling of structural differences between input dialects is virtually inevitable, and that the direction of leveling will favor majority variants. Other theories emphasize speakers' agency in selecting the linguistic variables in their environment and allow for a wider variety of outcomes. A close examination of parameters of variation in Hoopa Valley, California (chapter 3), and Siletz, Oregon (chapter 4) shows overall that structural leveling of dialect differences did occur in this period. However, leveling applied unevenly, with some parameters of regional variation persisting into the last generation of speakers to acquire native fluency in the PCA languages, in some cases even for dialects that were a tiny minority. These results suggest that a theory allowing for the maintenance of dialect differences via social-indexical considerations does a better job of explaining the full range of observed outcomes than one relying on deterministic majority-rule demographic principles alone.

Chapter 5 considers the linguistic consequences of contact between English and Hupa, the best-documented PCA language. The chapter focuses on lexical innovations in Hupa introduced to refer to the plethora of new technologies that became part of everyday life in Hoopa Valley in the 19th century. The main finding is that Hupa speakers borrowed very few lexical items from English, preferring instead to adapt existing Hupa words or to coin new ones using productive morphological resources. In non-lexical domains, an examination of main clause constituent order reveals that younger generations of Hupa speakers tended to use a postverbal nominal position less frequently than earlier generations had. However, it is argued that this is due not to direct bilingual interference from English, but rather to pragmatic contraction disfavoring the use of marked word orders. Overall, despite a high incidence of Hupa-English bilingualism already by the turn of the 20th century and the fact that wholesale language shift was underway in this period, Hupa appears to have undergone very few lexical or structural changes that were borrowed or otherwise directly modeled on English. This is interpreted as a continuation of conservative stances towards exogenous linguistic innovations that were a part of pre-contact linguistic practice both within the Athabaskan language family and regionally across California.

Chapter 6 concludes, summarizing the main findings of the dissertation and outlining directions for future research.

.

# Contents

## Acknowledgments

Every dissertation is a labor of love. Not on the part of the author, for whom it is many hours of toil punctuated by fleeting moments of clarity and excitement. Rather, it is the product of the love and support of people who had faith that the project would someday be completed or who, if they didn't, were kind enough not to say so. It would be impossible to name everyone who supported this effort on my behalf over the past six years, but I want to acknowledge here a few of the most important contributors. Foremost among these are the members of my dissertation committee: Andrew Garrett, for his thoughtful comments on all aspects of this research and for setting such a shining example of linguistic scholarship; Leanne Hinton, for her boundless enthusiasm for this project and all things related to California's indigenous people; Bill Hanks, for doing such a great job of straddling the disciplinary divide between linguistics and anthropology; and Victor Golla, who has so generously shared his vast knowledge of the Athabaskan language family with me. I also want to acknowledge other faculty in the UC Berkeley Department of Linguistics who have shaped my thinking more than they might suspect, especially Sharon Inkelas, Keith Johnson, and Lev Michael.

Special appreciation is due to Verdena Parker, who has spent so many hours patiently trying to teach me something about the Hupa language. Verdena and her husband Otis were always so welcoming when a gaggle of linguistics students would show up on their doorstep, often on short notice. Without their energy, enthusiasm, and sense of humor much of this work would not have been possible. I would also like to thank the other members of the aforementioned gaggle for their friendship and collaboration over the past few years: Amy Campbell, Kayla Carpenter, Ramon Escamilla, and Lindsey Newbold. I am also grateful to the language teachers and community linguists in Hoopa Valley whose dedication to revitalizing Hupa have been a constant source of inspiration for me, especially Danny Ammon, Melodie George-Moore, Silis-chi-tawn Jackson, and Jackie Martins.

Many people in the California language revitalization community have also inspired me with their commitment to seeing that indigenous languages of the region continue to be spoken in the future. This has included people associated with the Kawaiisu Language and Cultural Center: Julie Turner and Laura Grant, who keep the wheels turning, and elders Luther Girado, Lucille Hicks, and Betty Hernandez, all of whom taught me that laughter is the best antidote to elicitation-induced brain freeze. Others who have offered various forms of support and enouragement include Jocelyn Ahlers, Loren and Pyuwa Bommelyn, Marina Drummer, Lori Laiwa, Wes Leonard, Mike and Perry Lincoln, L Frank Manriquez, Vince Medina, Debbie Morillo, Carly Tex, and Matt Vestuto. Thanks to your efforts, someday a child or grandchild or great-grandchild of today's speakers and second-language learners will be able to tell their own story, in their own way, using whatever language suits them. I thank you all for giving me the opportunity to contribute to this enterprise in whatever small ways I can.

Other current and former members of the UC Berkeley linguistics community who have stimulated my thinking over the past few years include Rosemary Beam de Azcona, Jess Cleary-Kemp, Mara Green, Iksoo Kwon, Florian Lionnet, Hannah Pritchett, Daisy Rosenblum, Ruth Rouvier, Clare Sandy, Ronald Sprouse, John Sylak-Glassman, and Elisabeth Wehling. I am especially indebted to Will Chang for his patient explanations of Bayesian inference, without which chapter 2 of this dissertation would not have been possible. Special thanks to Paula Floro and Belen Flores, the best department administrators one can imagine.

Before joining the linguistics program at Berkeley, I got my first serious exposure to work-

ing with endangered languages during a stint at Diwurruwurru-jaru Aboriginal Corporation in Australia's Northern Territory. I am grateful to a number of people there for sparking my interest in this sort of work: Barbara Bobby, Rob Chapman, Sarah Cutfield, Jenny Denton, Greg Dickson, Mildred Hector, Felicity Meakins, and Sophie Nicholls.

My graduate studies at Berkeley have been funded thanks to the generosity of the University of California and the National Endowment for the Humanities (which supported my work as a research assistant in the Survey of California and Other Indian Languages under a Documenting Endangered Languages grant awarded to Andrew Garrett). This research benefited immensely from a 10-week Smithsonian Institution fellowship in the Department of Anthropology at the National Museum of Natural History. Thanks to my sponsors there, Gabriela Pérez-Báez and Ives Goddard, and to the wonderful staff at the National Anthropological Archives, especially archivist Leanda Gahegan. Portions of this research were presented at the 2009 and 2010 Dene/Athabaskan Languages Conferences, the 39th Annual Meeting of the Berkeley Linguistics Society, and the 2011 and 2013 winter meetings of the Society for the Study of Indigenous Languages of the Americas. I appreciate comments from those audiences, which have helped make this research much better than it would have been otherwise. Portions of chapter 2 will be published in the proceedings volume for BLS 39, and portions of chapter 4 have been published in volume 15 of the *Reports of the Survey of California and Other Indian Languages* series.

Finally, on a more personal note, I want to thank my parents, Bill and Trish Spence, and my sister Aimee, who somehow survived my moody teenage years and now uncomplainingly endure my moody adulthood. Profound thanks to Catalin Kaser for everything we shared and for seeing me through so many good years, and to Abigail Craig for her love and patient kindness in the dénouement of this project.

# Chapter 1: Introduction

## 1. Overview of Issues

The Pacific Coast Athabaskan (PCA) languages were once spoken in a region stretching from present-day Mendocino County in northern California to Douglas County in southwestern Oregon. Athabaskan-speaking people are estimated to have first arrived in the area approximately 1300 YBP in one or more migration events originating in southwestern Canada (Golla 2011:257-258). Despite hundreds of miles of overland travel and inevitable encounters with other linguistic stocks of North America both en route and upon arrival, these Athabaskan populations transmitted their language more or less faithfully transgenerationally for over a millennium. While linguistic differentiation in this period resulted in several more or less distinct Athabaskan languages in the region, and some traces of language contact are detectable (Golla 2000, Conathan 2004), the changes were small enough for the languages to maintain an essentially Athabaskan character. They are immediately identifiable as descended from the same Proto-Athabaskan source as languages spoken as far away as Alaska and the southwestern United States.

Beginning in the mid-19th century, this chain of transgenerational transmission was weakened and ultimately broken in the wake of the colonial encounter with Euro-American society. Tensions fueled by an influx of settlers led to the dispossession of many indigenous people of the region and their subsequent relocation to geographically restricted areas on a small number of reservations and rancherias. The long-term linguistic consequences of these disruptions on the PCA languages are by now all too familiar, since similar relocations of indigenous peoples were taking place across the American west in the same period. Assimilationist pressures – reservation life, adaptation to a wage economy, and institutional policies backed by the threat of military force – created an imperative for indigenous people to communicate in new ways, both with each other and with agents of the foreign culture that had uprooted them. Regional lingua francas, initially Chinook Wawa in some areas, eventually English in all, gained new prominence in everyday life and gradually replaced traditional forms of speech. By the mid-20th century, active knowledge of the Pacific Coast Athabaskan languages was no longer being acquired by children. With the marginal exceptions of Hupa and Tolowa, as far as is known none of these languages has any fluent first-language speakers remaining today (Pierce and Ryherd 1964, Golla 2011:74,80-81), although some of them have growing numbers of people who have acquired second-language proficiency in their heritage languages.

The Pacific Coast Athabaskan languages, as leaves on the Athabaskan family tree, instantiate the full evolutionary cycle of a linguistic phylogeny, understood as the set of diachronic relationships obtaining between two or more languages descended from a common proto-language. This cycle includes linguistic differentiation of a relatively homogeneous population of speakers due to uneven diffusion of innovations and migration; changes triggered by internal pressure or external contact with other phylogenetic units; bilingualism and obsolescence under wholesale language shift; and renewal via contemporary language reclamation and revitalization efforts. These evolutionary processes are the fundamental object of study of historical linguistics writ large, the common link connecting traditional comparative approaches, dialectology, sociolinguistics, and contact linguistics. The present study, taking PCA as a linguistic phylogeny as the starting point for analysis, explores its diachronic development at two very different granularities: macroevolutionary, inferring the development of the PCA languages over the course of many centuries, and microevolutionary, considering recent developments

manifest in the past century and a half of documentation. Each of these time scales presents its own set of empirical and theoretical problems, which are outlined in this introductory chapter: §2 addresses macro-scale issues related to phylogenesis, while §3 and §4 discuss recent contact between Athabaskan groups and between Athabaskan and English, respectively. An overview of the Pacific Coast Athabaskan languages, including orthographic and citation conventions used in this dissertation, is provided in §5.

## 2. Phylogenetic Classification

The narrow macro-scale question addressed in chapter 2 of this dissertation is whether or not the Pacific Coast languages are a well-defined subgroup of the Athabaskan family. PCA is uncontroversially classified into two subgroups, outlined in more detail in §5 below: a California group consisting of Hupa, Mattole-Bear River, Wailaki and other Eel River dialects, and Kato; and an Oregon group consisting of several varieties including Tututni, Chasta Costa, Chetco-Tolowa, Galice-Applegate, Upper Coquille, and Upper Umpqua. At issue is whether or not the California and Oregon groups are jointly descended from a common Proto-PCA ancestor or whether they are no more closely related to each other than to the rest of the Athabaskan family. Hoijer (1960) presented evidence that the languages comprise a subgroup of Athabaskan, a view endorsed to some extent by Krauss (1973). More recently, Golla (2011) has argued for the opposite view, implying that the innovations common to the California and Oregon languages identified by Hoijer are relatively unimportant, and moreover that the presence of two Athabaskan groups in nearly adjacent territory hundreds of miles from the rest of the family is in effect a historical accident.

Chapter 2 considers this question in detail, bringing to bear computational methods adapted from the biological sciences that have become prominent in historical linguistic research over the past decade. Beyond the narrow question of the status of the PCA languages in particular, two issues with broader significance are addressed. The first is methodological, focusing on a major division between these recent computational approaches and traditional paper-and-pencil historical studies invoking the Comparative Method. In traditional approaches, arguments based on non-lexical data, especially shared idiosyncratic morphological innovations, are always given the highest priority in establishing diachronic linguistic relationships, since such non-lexical features are borrowed relatively rarely (cf. Ringe, Warnow, and Taylor 2002:65-69 and Thomason and Kaufman 1988 for discussion, and Goddard 1975 for application of these principles in the North American context). Most recent computational studies, however, focus more narrowly on lexical data alone. This is due in part to the relative accessibility of lexical versus non-lexical data, but also to the fact that the theoretical assumptions justifying the use of biological models – that language change is analogous in crucial respects to genetic evolution – may not be valid in non-lexical domains. While not overcoming all possible objections to including non-lexical characters in the dataset, chapter 2 attempts to account for both lexical and non-lexical data in a computational analysis and is intended as a step towards resolving the methodological dissonance currently separating the two approaches.

Second, the study is set against a backdrop of skepticism among some Athabaskan scholars concerning the appropriateness of traditional family-tree models in Athabaskan historical linguistics. This is stated most strongly by Krauss (1973), who maintains that the entire Athabaskan family is an extended dialect network with no well-defined language boundaries. According to this view, linguistic innovations originating in different Athabaskan speech communities have diffused across geographically contiguous areas – sometimes more widely,

sometimes less so – ultimately leading to a network of cross-cutting isoglosses with "each language – each community – being a unique conglomerate" of changes (Krauss and Golla 1981:69). This perspective is common in Athabaskanist historical research, and may run counter to uniformitarian principles sometimes argued to govern the typical course of language change (Labov 1994:22, Ringe, Warnow and Taylor 2002:60-63). The fundamental issue is whether divergence events (leading to traditional branching family trees) are more common or significant in language histories than convergence events (i.e., network-like reticulations across nodes in a tree) (cf. Labov 2007, Babel et al. 2013, and Kalyan and François (to appear) for discussion). The computational methods invoked in this study bear on this question insofar as they quantify in their output the strength of tree-like phylogenetic signals. It thus becomes possible to assess the status of family tree models across conditions – with lexical versus non-lexical datasets, and with different assumptions about rates of linguistic evolution – in a quantitatively rigorous way. Well-supported subgroups across conditions can be considered likely to be part of the true evolutionary history of a language family; those which are not well-supported can be treated with considerably more skepticism. A key finding of chapter 2 is that PCA is well-supported in precisely this way. Results from elsewhere in the family, especially the Northern Athabaskan languages of Alaska and Canada, are more varied and thus may be indicative of the network-like structure advocated by Krauss (1973) and Krauss and Golla (1981).

## 3. Dialect Contact

After considering the question of the relatedness of the Oregon and California Athabaskan languages in chapter 2, subsequent chapters focus on the history of each of these groups in the documented historical period. Chapters 3 and 4 pay special attention to contact between speakers of closely related Athabaskan varieties that took place on reservations established in Hoopa Valley, California (chapter 3) and Siletz, Oregon (chapter 4). Although language obsolescence was the ultimate outcome for all of the PCA languages in this period, a cornerstone of this research is the notion that the shift to English was gradual and the PCA languages did not disappear overnight. Rather, they continued to be spoken well into the 20th century: most were still being learned by children for the first half-century or more after reservations were established, and at least three generations of speakers were multilingual to varying degrees in Athabaskan languages and English.[1] There was thus ample opportunity for contact-induced changes to take place that were due to the mutual influence of the Athabaskan languages and dialects on one another. Indeed, it would be surprising if such changes did not occur, since the overall demographic profile of resettlement on reservations – abrupt relocation of distinct speech communities to a restricted geographic area – is broadly similar to other cases of large-scale migration that led to the emergence of mixed varieties of English, Norwegian, and Hindi in cities and colonies (Trudgill 2004, 2008a; Kerswill 1994, 2002; Kerswill and Williams 2000, 2005; Kerswill and Trudgill 2005).

The dynamics of contact between Athabaskan varieties in the reservation setting can therefore be viewed through the lens of theories of dialect contact, broadly construed to include "contact between varieties of language that are mutually intelligible *at least to some degree*"

[1] For some people, this multilingualism included non-Athabaskan indigenous languages of the region as well, and Chinook Wawa in Oregon.

(Trudgill 1986:1, emphasis added).[2] Such theories seek to explain three broad classes of interrelated phenomena when two or more dialect groups come into extended contact with each other. The first is the overall direction of change: whether dialects undergo convergent structural leveling and become more similar to one another in a process known as koineization, whether their differences are maintained, or whether they diverge and become less similar to one another over time.[3] Second, since a combination of outcomes is sometimes found – typically the maintenance of some features with concurrent leveling of others – another area of interest lies in explaining why some dialect features are the targets for leveling sooner and others later or not at all. Finally, the third class of phenomena to be explained by theories of dialect contact is how competition between dialect features is resolved in the case of koineization: why particular dialect variants survive while others perish. These last two phenomena can be characterized as the selection of linguistic variables as targets for change on the one hand, and the selection of particular values for each variable on the other.

A large body of research has examined cases where the outcome of dialect contact is koineization. This has led to the following theory (adapted from Kerswill and Trudgill 2005) according to which the formation of koines involves three distinct stages corresponding to successive generations of speakers living in a mixed dialect community:

| | |
|---|---|
| Stage I | Dialect divisions are largely maintained, the adult founders of new communities having relatively little linguistic influence on one another. |
| Stage II | Selection of variants gets underway: extreme inter- and intra-speaker variation as children are exposed to a range of linguistic input in the home and in the community at large. |
| Stage III | Variation is vastly reduced as the speech community settles on a relatively stable set of linguistic norms. |

In the first stage, adult speakers of diverse dialects come into contact with each other, but their speech undergoes only rudimentary leveling due to the fact that linguistic behaviors already established in adulthood are difficult (although by no means impossible) to change. In the second stage, the first generation of speakers raised in the new community starts the process of

[2] Here and throughout this thesis I will use the terms "language contact" to refer to contact involving Athabaskan and non-Athabaskan languages (especially English) and "dialect contact" to discuss interactions between Athabaskan-speaking groups. Some of the Oregon Athabaskan varieties in contact at the Siletz reservation were distinct enough to be considered separate languages, but most authors agree that there was a fairly high degree of mutual intelligibility between them nonetheless (discussed further in section 4). In any case, it is not certain that one would expect different principles to be operative in contact involving dialects versus separate languages. Weinreich, Labov, and Herzog (1968:158) maintained that "[i]n principle, there is no difference between the problems of transference between two closely related dialects and two distantly related languages." Thomason and Kaufman (1988:9) point out that "[j]ust as it is often difficult to tell whether two speech forms are dialects of one language or separate languages, so the borderline between dialect interference and foreign interference is often fuzzy."

[3] I will often use the terms "convergence," "(dialect) leveling," and "koineization" interchangeably. Some authors (e.g., Kerswill and Williams 2005) consider dialect leveling to be a structural correlate of koineization, which is a broader sociolinguistic process.

selecting from the linguistic variables in their environment. This stage is marked by extreme variation across the community as a whole and in the speech of individual speakers. Finally, the third generation settles on a stable set of linguistic conventions, typically in tandem with the development of social cohesion among speakers known as "focusing" in the sense of Le Page and Tabouret-Keller (1985). Old dialect distinctions may persist, but with new social-indexical values – so-called "reallocation" of erstwhile regional indices.

This three-stage model can be taken as a theoretical baseline for understanding dialect contact situations. The timeline of early documentation of Pacific Coast Athabaskan, outlined in more detail in §5, is conveniently spaced approximately one generation apart (1850s – 1880s – 1900 and forward), offering snapshots of the languages at intervals that are well-suited to exploring the model's main prediction: in the usual course of affairs, koineization will result when dialect groups come into prolonged contact with each other, differences between them leveling out within a couple of generations after a new community is established. Deviations from this prediction – either koineization that proceeds faster or slower than the model predicts, or non-convergent outcomes (dialect maintenance or divergence) – therefore require special explanation. Kerswill (2002) argues that only "a small set of social and linguistic factors" are implicated in such exceptional cases. Koineization may be faster than expected if the degree of structural difference between input varieties is small. It can be delayed by a lack of social integration in a new community, i.e., impediments to focusing can mitigate convergent pressure.[4] Clearly, however, the prediction of the three-stage model is that koineization is the expected outcome of contact, especially where focusing takes place. This is stated most clearly in Trudgill (2004:88-89), who maintains that "focusing implies koineization."

Trudgill (2004) goes a step further in his study of new varieties of English that developed in British colonies of the southern hemisphere. Trudgill subtitles his monograph "the inevitability of colonial Englishes," by which he means not only that the fact of koineization is inevitable, but so too are its outcomes, in some cases at least. Considering what he calls "tabula rasa" dialect contact, in which speakers of diverse dialects migrate en masse to a new geographic location where none has pre-established claims to legitimacy, he maintains that koineization in these circumstances proceeds essentially deterministically:

> "...given sufficient linguistic information about the dialects which contribute to a mixture, and given sufficient demographic information about the proportions of speakers of different dialects, it is possible, within certain limitations, to make predictions about what the outcomes of the mixture will be, at least in broad outline." (Trudgill 2004:26)

According to Trudgill, in the development of new English dialects in the southern hemisphere "the minority simply and mechanistically accommodated to the majority" (2004:148). Trudgill's determinism is demographic, with dialect features found among the largest number of speakers supplanting minority variants, thus providing an explanation for the selection of particular variants that existed in the input dialects. In Trudgill's view, socially contingent factors such as "prestige," "stigma," "identity," and "ideology" play no role in the outcomes of koineization in tabula rasa situations, except perhaps in the initial dialect mixing among Stage I adult speakers who first come into contact with each other.

[4] Kerswill gives the example of Spitsbergen, a mining community in Norway where the population is transient and families do not stay long enough for a stable koine to develop.

Although Trudgill (2004) restricts his deterministic theory to tabula rasa contact situations, he situates his discussion in a broader debate in the literature concerning the role of what he has called "identity factors" in language change (2008a). In this more recent discussion of new dialect formation in colonial settings, Trudgill (2008a) seems to imply that such identity factors are irrelevant in explaining the trajectory of dialect contact in general. While Trudgill acknowledges that contact-induced changes can take on social meanings as indices of particular reference groups, he considers this to be a development that is parasitic on linguistic changes that have already occurred. Trudgill thus denies that individual or collective intentions to create or preserve social meanings are responsible for initiating changes or causing them to proceed in a particular direction. Invoking the uniformitarian principle of Labov (1994), he maintains that "[w]e do not need [identity] as an explanatory factor at any moment in human history."

A central component of Trudgill's (2004, 2008a) arguments concerning the irrelevance of identity factors in dialect contact are his views on micro-interactional dynamics. At least since Trudgill (1986), many studies of dialect contact invoke the notion of interactional accommodation as the mechanism that leads to transgenerational dialect leveling (cf. Bourhis and Giles 1977, Kerswill 2002). According to this view, during face-to-face interactions people make minute, short-term adjustments to their speech patterns to match those of their interlocutors. Such minor accommodations eventually lead to long-term changes in how people talk, thus providing a bridge between short-term linguistic behaviors (in principle observable in real time) and long-term community-wide changes (often observed only post hoc). Since their linguistic systems are relatively malleable, children raised in mixed dialect communities are particularly susceptible to such modifications as they are exposed to a wide variety of dialect features in the speech of adults and their peer groups (Kerswill and Williams 2000, 2005). Trudgill (2008b) is explicit in claiming that children in colonial situations are either unaware of or unresponsive to the ideological "baggage" that adults bring with them to a new community (a view that has been criticized by Kerswill 2006 on empirical grounds). The social-indexical values of dialect variants, in Trudgill's view, either are not transmitted to children or do not have the same interactional force as they do for adults.

According to Trudgill (2008a), accommodation is automatic: people will engage in interactional tuning with interlocutors regardless of their feelings about them or their attitudes about the social categories to which they belong. While some experimental research supports this position (e.g., Delvaux and Soquet 2007), studies such as Babel (2009a, 2012) suggest that the degree of interactional accommodation is correlated with the solidarity that a speaker feels towards an interlocutor (although this may not be within speakers' conscious control). Nonetheless, Trudgill maintains that the only obstacle to koineization when mixed dialect communities are formed is infrequency of interaction, citing Labov (2001):

> "It does not seem likely that the linkage between linguistic and social structure involves the association of forms or frequencies with particular groups, and calculations of the consequences of adopting their speech forms. As always, it is good practice to consider first the simpler and more mechanical view that social structure affects linguistic output through changes in frequency of interaction." (Labov 2001:506)

In other words, identity factors can impact the outcomes of dialect contact only indirectly by influencing the frequency with which members different reference groups interact with each other, not by the invocation of social categories in the course of interaction.

Trudgill's demographic determinism stands in contrast to theories of language variation and change that afford a prominent role to speakers' perceptions of their interlocutors and beliefs about the social categories to which they belong. Eckert (2008), for example, outlines an ethnographically-oriented theory of linguistic practice whereby language ideologies are key factors in determining speakers' selection of linguistic variables in their environment. Social categories – especially those relevant within a local speech community – are ever-present in the choices that speakers make in the course of interaction as they position themselves in a social-indexical field. Similarly, Thomason (2007) argues that people can make deliberate linguistic choices reflecting ideologies that they want to emphasize when they speak. Recent computational simulations of social network behavior by Fagyal et al. (2010) explicitly model the impact that Bloomfield's (1933:345) "relative prestige" can have on the adoption of linguistic variants. In all of these views, which can be subsumed under the label of indexical approaches, individuals' construals of their social world can be invoked when people position themselves (whether consciously or unconsciously) according to their understandings of the categories indexed by particular linguistic choices. Although these theories are not designed to account for dialect contact situations in particular, it is clear that they have broad applicability with implications for many different kinds of language change, including koineization.

Such identity factors are often invoked in explanations for non-convergent outcomes of dialect contact (Tuten 2008), especially when minority communities face assimilatory pressure from encroaching dialect groups and wish to maintain "symbols of social attitudes" (Sapir 1949:85) via acts of what Schilling-Estes and Wolfram (1999) call "linguistic self-defense." Perhaps most relevant for present purposes is Furbee's (1995) analysis of dialect contact among Native American groups who were resettled on reservations and its relationship to language obsolescence. Furbee, following Bell (1984), distinguishes "positive accommodation" leading to conversational convergence versus "negative accommodation" whereby speakers maintain or increase linguistic differences between themselves and their interlocutors. A key stage in the obsolescence of Chiwere (a Siouan language of the American midwest) was the development of what Furbee calls "family dialects" as social fragmentation, part of the legacy of colonization, created pressure for people to distance themselves linguistically from others:

> "As speakers began to devalue the speech of those outside their own families, they would feel that they would have to 'speak down' to non-family members if they were to accommodate in Chiwere, since for them their own dialect would be the acrolect and all others would be some kind of basilect. With the increase in factionalism, there was less and less reason to accommodate." (Furbee 1995:4)

According to Furbee, language obsolescence begins with polarized ideological stances within a community that motivate divergent linguistic behavior. Such hyperlocal innovations accrue social-indexical meanings that are eventually avoided by using another language altogether, a "language of accommodation" (English in the case of Chiwere).

Crucially, in Furbee's view the lack of convergent accommodation is not due to infrequency of interaction among speakers (although the social fragmentation she describes is perhaps indicative of a lack of community focusing in the relevant sense). Far from being unaware of linguistic innovations due to isolation from other speakers, people simply choose not to follow the example of innovative speakers they are not sympathetic to. Furbee's theory is of interest in the present context because it creates an explicit link between the outcomes of dialect contact

and the broader nexus of contact with a colonial language in reservation settings, precisely the circumstances that the Pacific Coast Athabaskan groups encountered in the reservation setting.

Also to be included among the family of indexical approaches to dialect variation and contact is an environmental approach developed by Hill (2001). Like Trudgill (2004), Hill seeks to explain not only the fact of language change but also its direction – why some dialect groups' variants are adopted by others rather than vice-versa. Hill considers the question in terms of how receptive to variation and innovation different groups are. She identifies two major stances that speech communities can take: "localist," where speakers emulate a particular linguistic model, versus "distributed," where the model speakers aim for is more fluid, making speakers of a dialect less resistant to linguistic innovations. Materialist in orientation, Hill's theory treats linguistic variables as a kind of currency linked to access to life-sustaining resources. Groups with secure access will tend to adopt relatively conservative localist stances disfavoring variation as a way to assert their primary claims on resources; groups with less secure access will tend to favor distributed stances encouraging variation, with particular variants invoked contextually by speakers to project an affiliation with those who control access. Groups with distributed stances to variation will therefore be more likely to adopt innovations introduced by other groups; those with localist stances will be less likely to do so. Hill's theory is indexical insofar as these stances generalize over uses of language that are indices of people's claims on resources, but it is tightly constrained by the environmental milieux in which variation develops and explicitly does not invoke abstract notions such as prestige. Hill also emphasizes that her theory is not based on micro-interactional considerations but instead involves "broad stances toward variation that are constrained by the total contexts in which speakers live" (2001:278).

Hill's theory is of interest here in part because it was explicitly developed as an alternative to demographic approaches similar to the one developed by Trudgill (2004) for tabula rasa contact, specifically Trudgill's (1983:72-78) gravity model of diffusion whereby linguistic innovations spread from areas of high population density to areas of low population density (cf. Wolfram and Schilling-Estes 2003 for discussion). More recently, demographic considerations have also factored into Labov, Rosenfelder, and Fruehwald's (2013) study of Philadelphia English, where the size of a population undergoing linguistic changes is invoked to support arguments to the effect that micro-interactional dynamics are inadequate to explain long-term diachronic trajectories. Hill (2001) argues that such demographic explanations are highly particular to urbanized societies with large differences in population density and cannot explain the diffusion of changes in less populous societies with more uniform densities – i.e., the societies that existed over the majority of human prehistory. Hill points out the need for a theory of directed language change that can apply to both situations – such a theory would seem to be especially warranted by the very uniformitarian principles that Trudgill (2004, 2008a) and Labov, Rosenfelder, and Freuhwald (2013) invoke.

The demographic determinism of Trudgill (2004, 2008a,b) and indexical approaches such as Eckert (2008), Thomason (2007), Fagyal et al. (2010), Furbee (1995), and Hill (2001) thus offer two competing modes of explanation for understanding the outcomes of dialect contact. One of the primary goals of chapter 3 and especially chapter 4 of this dissertation is to explore the extent to which the two approaches can account for the outcomes that are observed in contact involving PCA languages in the reservation setting: the overall direction of contact (koineization vs. maintenance vs. divergence), the selection of particular linguistic variables as targets of leveling or divergence, and the asymmetric selection of particular variants from the input dialects where leveling occurs. Chapter 3 presents evidence that contact in Hoopa Valley

involving Hupa and closely related dialects did result in koineization, tracing the leveling of one dialect difference through the available documentation. A second parameter of variation in the Hupa-speaking community, one less clearly linked to regional dialect divisions, did not undergo leveling in this period, although some changes in its distribution are observed in the documentary record. While the parameter of variation that was the target of leveling was also one that was indexically linked to a stigmatized group on the reservation, that group was also a minority. It is suggested that in this case the indexical and demographic approaches to the selection of variants under koineization perform equally well. At the same time, however, it is argued that variation along these lines probably extends back into pre-contact times, and that a strictly demographic account of the diffusion of linguistic features does not make the correct predictions about which groups exhibited variation.

The case study of variation and dialect contact in Hupa and related dialects in chapter 3 sets the stage for considering similar issues at the Siletz Reservation in Oregon in chapter 4. The crucial difference there is that the vast majority of the indigenous groups who were resettled at Siletz were from a different region, so no single group's language had an a priori advantage over any of the others due to historical continuity in the locus of contact. This situation can be interpreted as instantiating the tabula rasa dialect contact scenario developed by Trudgill (2004, 2008a). Nor did any single group at Siletz have a significant population advantage over the others. Thus, two of the factors that favored the majority dialect in Hoopa Valley are removed. While koineization in this period affected most of the Athabaskan varieties present at Siletz, some minority dialects survived nonetheless. This is taken as counterevidence to Trudgill's claim that koineization is inevitable in tabula rasa contact, since demographic determinism predicts that such varieties would have been at an extreme disadvantage. While it is difficult to rule out the possibility that the retention of minority variants was due to a lack of community focusing, it is argued that this is not to be equated with infrequency of interaction: all of the Athabaskan groups lived side-by-side at Siletz and intermarried with one another, and it is hardly plausible that there wasn't significant daily communication between them.[5] In this case, it appears that a theory excising social-indexical considerations as explanatory factors in dialect contact has difficulty accounting for the full range of observed outcomes.

## 4. Contact with English

Dialect contact between Athabaskan-speaking groups in the reservation period was set against a backdrop whereby all indigenous languages of California and Oregon eventually underwent language obsolescence under wholesale shift to English. Chapter 5 focuses on this broader nexus of language contact between English and Hupa, the PCA language with the most robust documentation available. The research question addressed is what impact the colonial encounter with English-speaking populations had on the Hupa language in the decades before it ceased to be acquired by children. This includes changes that were a direct consequence of Hupa-English bilingualism – borrowing and effects that might be ascribed to incipient language attrition – as well as more general responses to functional pressure to adapt linguistically to the technological

[5] Naturally, the documentary record is such that there will be severe limits on the certainty with which these issues can be considered. It is impossible to know the actual details of how frequently or in precisely what ways various groups interacted with each other. However, the situation is probably not much worse than the cases considered by Trudgill (2004, 2008a), and similar kinds of reasonable speculation can be engaged in, especially where supported by statements in various historical sources.

changes that were introduced into Hupa society after contact.

Many of the changes that occurred in Hupa in this period were concentrated in the lexicon. The discussion in chapter 5 is cast in light of Brown's (1999) comparative study of lexical acculturation in Native American languages, which provides a framework for quantifying how heavily indigenous groups borrowed words denoting the large number of new technologies and other items introduced by the colonial powers they came into contact with. One of the main findings of chapter 5 is that Hupa borrowed very few English words for this purpose and instead relied heavily on the strategy of coining new vocabulary. This conservatism is very much in keeping with trends in California as a linguistic area (Golla 2011:5) and across the Athabaskan family both in pre-contact times (Sapir 1921) and in more recent colonial contact (Brown 1999:85). Brown (1999) favors explanations for different degrees of lexical borrowing in the Americas in terms of regional differences in the incidence of bilingualism, which he correlates with differences in the strategies of colonization favored by French- and English-speaking powers on the one hand, and Spanish- and Russian-speaking powers on the other. This view is examined critically in light of the Hupa data. An alternative view is developed whereby Hupa's linguistic conservatism with respect to English is an extension of pre-contact stances disfavoring borrowing, understood in terms of Hill's (2001) theory outlined above.

While there was relatively little lexical borrowing from English into Hupa, the loanwords that do exist bear on a current debate concerning the phonetic versus phonological basis for adapting loanwords to recipient language phonology. There are two broad views in the literature. According to one, loanwords are adapted based on perceptual cues in the acoustic signal of source language words, including subphonemic properties (cf. de Jong and Cho 2012 and Chang 2012 for discussion). In the other, developed especially in LaCharité and Paradis (2005), loanwords are adapted to the recipient language phonology via a more abstract process of matching phonological features without reference to their phonetic implementation. The treatment of aspiration – phonemic in Hupa, allophonic in English – sheds light on this debate, Hupa for the most part preserving subphonemic aspiration in loanwords from English.

Also of interest is the fact that many new lexical items in Hupa introduced in this period failed to diffuse widely throughout the Hupa-speaking community. There are many cases where two or more competing neologisms were coined to label introduced items, leading to a fairly high degree of lexical variation. It has been argued in the literature that such extremes of variation are indicative of imminent language attrition, including in Furbee's (1995) discussion of Chiwere outlined above. An alternative view proposed here, however, is that such variation might have been a longstanding fact of Hupa linguistic culture. Semantically transparent nominalizations comprise most of the Hupa nominal lexicon and similar lexical variation is found even for items associated with pre-contact Hupa culture. Thus, while Hupa did indeed have a high degree of lexical variation immediately prior to undergoing obsolescence, the two phenomena may not be causally related.

If lexical borrowing from English into Hupa is rare, it is even more difficult to find clear cases where Hupa borrowed non-lexical features from English. One domain where this might be expected is word order, since this is often one of the first areas where non-lexical contact effects can be detected (Thomason and Kaufman 1988:55). A study of main clause constituent order in Hupa texts from the early 20$^{th}$ century to the present reveals statistically significant changes in tendencies with respect to the placement of subjects in postverbal position. However, it is argued that these changes are not modeled directly on English's SVO word order: while postverbal subjects become less common, postverbal objects do not become more common, and

some speakers show a significant decrease in the use of postverbal objects as well. This is interpreted as a trend towards less frequent use of marked word orders, and hence as a species of pragmatic attrition rather than interference directly modeled on English patterns.

## 5. Overview of Pacific Coast Athabaskan

This section summarizes information about the Pacific Coast and other Athabaskan languages that is relevant to understanding the issues addressed in subsequent chapters, including conventions used for orthographies and citing sources. The discussion of PCA draws heavily on the comprehensive treatment in Golla (2011:76-82), which provides descriptions of each language's dialects, pre-contact territory, and phonological inventories, inter alia. Readers are referred to Golla's study for more complete information about each language and exhaustive bibliographical references.

### 5.1. The Athabaskan Language Family

The Pacific Coast Athabaskan languages are members of the Athabaskan language family, itself part of the larger Na-Dene stock of North America that also includes Eyak and Tlingit of southeastern Alaska (Sapir 1915, Krauss 1973). The Athabaskan proto-language is believed to have been spoken in eastern Alaska and western Canada as late as 2500 YBP. Subsequent migrations of Athabaskan-speaking groups out of this homeland are estimated to have occurred between 1000 and 2200 YPB (Krauss and Golla 1981:68, Golla 2011:68-69, 257), resulting in three main geographic divisions within the family: Northern Athabaskan in Alaska and western Canada; Southern Athabaskan, also known as Apachean, in the American southwest; and the Pacific Coast languages in California and Oregon.

The Northern Athabaskan and Apachean languages appear in this dissertation mainly in connection with the long-range comparative study presented in chapter 2. Languages from these groups were selected based on the availability of robust lexical and non-lexical data and represent only a subset of the approximately 40 languages commonly identified within the family. For the Northern Athabaskan group, this includes the Alaskan languages Ahtna, Dena'ina, Koyukon, and the Minto variety of Lower Tanana; Gwich'in (Kutchin), which straddles the border between eastern Alaska and western Canada; and the Canadian languages Beaver, Dakelh (Carrier), Dëne Sųłiné (Chipewyan), Hare, and Tsuut'ina (Sarcee). The Apachean languages included in the study are Chiricahua, Jicarilla, Lipan, Navajo, and San Carlos. Some of the structural details of these languages relevant to the computational analysis in chapter 2 are summarized in Appendices A and B of that chapter. For more detailed discussion, readers are referred to Hoijer (1963, 1971) and Krauss and Golla (1981), and to Krauss (1973) for an overview of the Athabaskan family as a whole and a critical evaluation of Sapir's (1915) original Na-Dene hypothesis.

Frequent reference will be made to reconstructed Proto-Athabaskan vocabulary, which is taken from Krauss (2005) except where otherwise noted; in many cases Krauss' transcriptions have been normalized for consistency with the conventions employed elsewhere in the present work. The Proto-Athabaskan segment inventory is presented here for ease of reference, as reconstructed by Krauss and Golla (1981) but also normalized for consistency. The Proto-Athabaskan obstruents are summarized in Table 1.1:

| | stops/affricates | | | continuants | |
|---|---|---|---|---|---|
| | unaspirated | aspirated | glottalized | voiceless | voiced |
| dental | d | t | t’ | s | z |
| | dz | ts | ts’ | - | - |
| lateral | dl | tɬ | tɬ’ | ɬ | l |
| palatal | dž | tš | tš’ | š | ž |
| labialized | dž$^{w}$ | tš$^{w}$ | tš$^{w}$’ | š$^{w}$ | ž$^{w}$ |
| front velar | ĝ | k̂ | k̂’ | x̂ | ŷ |
| back velar | ɢ | q | q’ | x̣ | ɣ |
| labialized | ɢ$^{w}$ | q$^{w}$ | q$^{w}$’ | x̣$^{w}$ | ɣ$^{w}$ |
| laryngeal | ʔ | - | - | h | - |

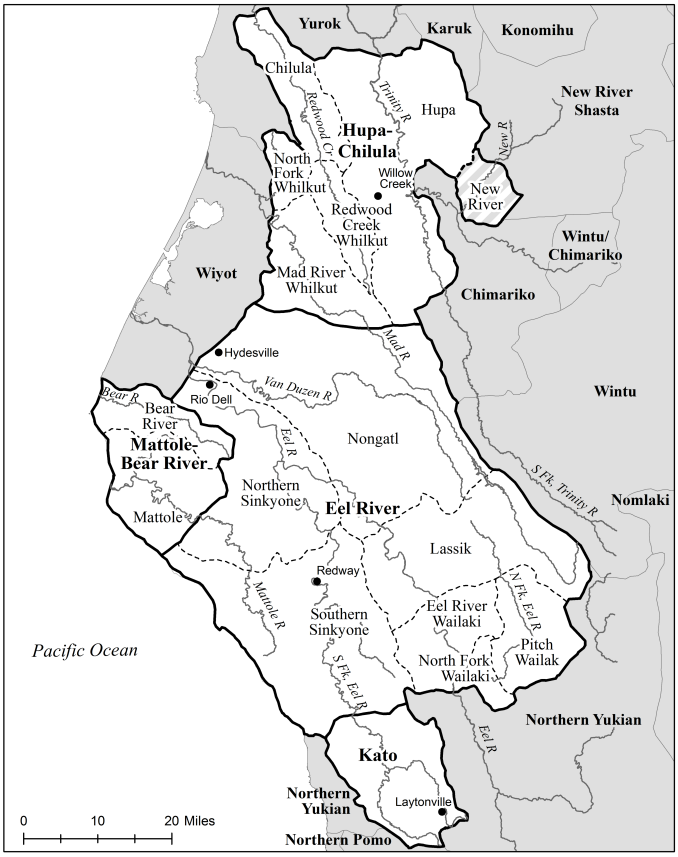
Table 1.1: Proto-Athabaskan obstruents

Proto-Athabaskan also had sonorants /n w y w̃ ỹ/, where /w̃/ and /ỹ/ are nasalized counterparts of /w/ and /y/ respectively. Proto-Athabaskan is reconstructed as having five phonemic full vowels /i e a o u/ (which may be long or short depending on context) and three reduced vowels /ə ɑ ʊ/ (which are inherently short). Vowels could also be articulated with glottal constriction, represented here with the diacritic /◌̰/, a property that developed into tone in many languages (Krauss 2005, Kingston 2005). Proto-Athabaskan forms are sometimes reconstructed with nasalized vowels as well. Readers should be aware of differences between Athabaskanist versus IPA and Americanist transcription conventions, especially the use of /b d g/ for what are in many Athabaskan languages voiceless unaspirated consonants, and the use of /p t k/ for voiceless aspirated sounds. For the most part I follow the Athabaskanist conventions, but sometimes explicitly represent aspiration where it adds clarity to the discussion at hand (e.g., the analysis of loanword adaptation in chapter 5).

## 5.2. California Athabaskan

The California Athabaskan languages were spoken in pre-contact times in a region stretching from Humboldt and western Trinity Counties in the north to Mendocino County in the south, as shown in Map 1 (from Golla 2011:76):[6]

[6] Maps 1 and 2 are used with the generous permission of Jacquelyn Bjorkman, GIS analyst at the Information Center for the Environment at the University of California, Davis.




Map 1: California Athabaskan languages (Golla 2011:76)

Golla (2011) distinguishes four California Athabaskan languages. From north to south in Map 1, they are Hupa-Chilula, Mattole-Bear River, the Eel River dialect network, and Kato. Hupa-Chilula is a composite name for Hupa, spoken in Hoopa Valley on the lower Trinity River, and Chilula, a variety from lower Redwood Creek to the west of Hoopa Valley. Chilula and Whilkut, from the upper Redwood Creek and Mad River regions south and west of Hoopa Valley, are often identified as separate dialects of Hupa; detailed discussion of these and other Hupoid divisions is found in §2 of chapter 3. I will sometimes use the term "Hupa proper" to refer to the lower Trinity variety to the exclusion of other Hupa dialects. Similarly, Mattole and Bear River were dialects of a single language that will henceforth be referred to simply as "Mattole." The Eel River dialect network includes several varieties formerly spoken on the Eel and Van Duzen Rivers, canonically discussed according to groupings known as Nongatl, Sinkyone, Lassik, and Wailaki (Golla 2011:79). Of these, Wailaki has the most extensive documentation and is the basis for the analysis in chapter 2, although information about other Eel River dialects has been accounted for where available as well.

As noted by Golla (2011:77), Hupa is by far the best-documented of any Pacific Coast Athabaskan language. The earliest attestations are found in 19th century wordlists such as Gibbs (1851-1852, 1853b), Crook (1852-1861), Azpell (1870), Powers (1870s), and Curtin (1888-1889). Extensive linguistic and ethnographic documentation was conducted between 1900 and 1910 by P.E. Goddard, who published several linguistic and ethnographic studies of Hupa proper (1901, 1903, 1904, 1905, 1907, 1911, 1928) and the Redwood Creek dialects (1914a-b). Edward Sapir collected a large amount of text material in 1927, published only recently thanks

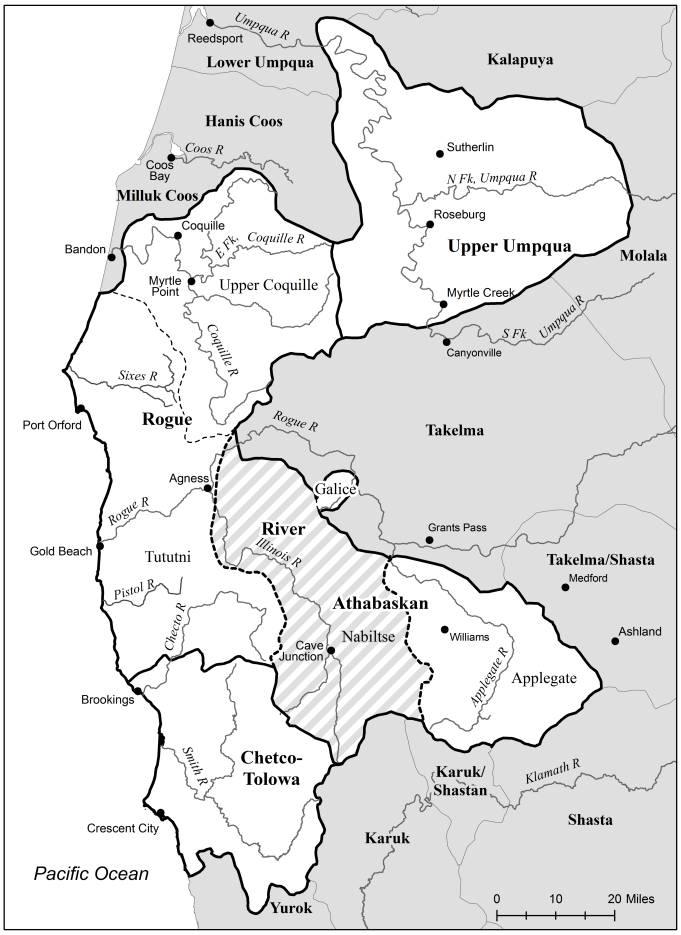
to the efforts of Victor Golla and Sean O'Neill (Sapir and Golla 2001). Victor Golla conducted fieldwork in the early 1960s that informed his dissertation research (1970) and other publications (1964, 1977, 1984, 1985, 1996a, 1996b); his field recordings and microfilm copies of his field notebooks and are also publicly accessible (Golla 1963). Other published and archival sources of Hupa documentation consulted in this study include Reichard (1922), Bright (1950-1952), Woodward (1953), Gordon (1996), and Gordon and Luna (2005). Since 2005 graduate students based at the University of California, Berkeley (the present author among them) have been working with contemporary speaker Verdena Parker to document her knowledge of the language, and recordings produced as part of that project have been a key source of information as well.

Documentation of other California Athabaskan languages is less robust than it is for Hupa. The main sources are Li (1930a) for Mattole, which also includes some of Li's Wailaki material; for Wailaki, Goddard (1923) and the unpublished field notes of Li (n.d.), one text from which was published as Seaburg (1977); and Goddard (1909, 1912) for Kato. Independent scholar C. Hart Merriam collected vocabularies of all of the California Athabaskan languages, including many of the Eel River dialects, now archived with Merriam's other papers at the Bancroft Library (Merriam 1850-1974), a small portion of which is published in Merriam (1979). Merriam seems to have been careful to record variant pronunciations where he found them, providing a key source of information about dialect variation in chapter 3. Other sources, especially unpublished field notes from Goddard held by the American Philosophical Society, have not been analyzed systematically for the present study; a comprehensive list of additional references for these languages can be found in Golla (2011).

Unsurprisingly, the quality of these sources of documentation is quite variable. Problems with early wordlists transcribed when the study of Native American languages (and linguistic science more generally) were in their infancy are obvious, although it should be noted that the material found in Curtin's vocabulary are fairly accurate for the period. By his own admission, Goddard did not consistently transcribe phonemic contrasts, especially in his rendering of glottalization (Goddard 1908, 1909:68, 1923). Merriam used an idiosyncratic transcription system based on English spelling conventions that can be difficult to interpret. Woodward is also inconsistent in a manner that often suggests transcription errors. For the most part, however, the Hupa linguistic variables considered in chapter 3 involve contrasts that even untrained ears are likely to have recorded accurately: velar stops versus fricatives, and word-initial /n/ versus /ʔ/.

## 5.3. Oregon Athabaskan

Several varieties of Oregon Athabaskan have been identified in the literature, each associated with pre-contact villages located along watercourses on or near the coast of southwestern Oregon and extending across the border into California (Miller and Seaburg 1990). These are shown in Map 2, also from Golla (2011:71):




Map 2: Oregon Athabaskan languages (Golla 2011:71)

According to Golla (2011), the varieties in bold in Map 2 (Upper Umpqua, Rogue River Athabaskan, and Chetco-Tolowa) are distinct languages; Rogue River Athabaskan is a dialect network with several distinct local varieties, many of which came to be known collectively as "Tututni" by the mid-20th century.

Figure 1.1 summarizes the classification that Golla (2011:70-75) provides for Oregon Athabaskan; abbreviations used when citing data in chapter 4 are given in square brackets:

| | |
|---|---|
| **Upper Umpqua** | |
| **Rogue River dialect network** | |
| Upper Coquille | [Coq] |
| Lower Rogue | |
| Sixes River | [Six] |
| Euchre Creek | [Euc] |
| Tututni-Joshua | [Tut] |
| Mikwanutni | [Mik] |
| Chasta Costa | [ChC] |
| Pistol River | [PiR] |
| Galice-Applegate | [Gal] |
| **Chetco-Tolowa** | |
| Chetco | [Cht] |
| Tolowa | [Tol] |

Figure 1.1: Classification of Oregon Athabaskan in Golla (2011)

Figure 1.1 is arranged roughly north to south, with Upper Umpqua the northernmost member of the Oregon group (in the region around present-day Roseburg) and Chetco-Tolowa the southernmost (straddling the Oregon-California border as far south as Del Norte County). Within the Rogue River dialect network, the northernmost variety was Upper Coquille, which lay between Upper Umpqua to the north and the Lower Rogue dialects to the south. The Lower Rogue dialects included Sixes River and Euchre Creek along the coast north of the Rogue River; Tututni-Joshua, originally spoken in villages near the mouth of the Rogue River, with Mikwanutni and Chasta Costa farther upstream; and Pistol River on the coast to the south of the Rogue River, midway between Tututni-Joshua and Chetco-Tolowa. The Galice-Applegate dialect was spoken in two inland areas that were adjacent to the non-Athabaskan Takelma people in pre-contact times; they were separated from other Oregon Athabaskan groups by a variety known as "Nabiltse" that is poorly documented but may be another dialect most closely related to Galice-Applegate (Golla 2011:72-73). Most speakers of these Oregon Athabaskan varieties were relocated to the Siletz Reservation in the mid-19th century, the focus of chapter 4. The exceptions were the Tolowa (who remained more or less in situ in California) and Upper Umpqua (who were relocated to the Grand Ronde reservation adjacent to Siletz).

Although Golla (2011) identifies three distinct Oregon Athabaskan languages, it should be pointed out that both Pierce and Ryherd (1964:142) and Bright (1964:101) reported a high degree of mutual intelligibility between Chetco-Tolowa and the Rogue River dialects. This is corroborated in statements from speakers interviewed in the 20th century, one reporting that Chetco people "talk half-way RR [Rogue River]" (Harrington 1981, reel 25, frame 989), another that the Chetco and Tolowa people can understand the Rogue River dialects but "talk different" (Barnett 1934).[7] On the other side of the coin, Golla (2011) includes Galice-Applegate ("Galice" henceforth) in the Rogue River group but considers it a highly distinctive member; Sapir (1914:274) reported that Galice "differed so much from Chasta Costa [in the Lower Rogue group] as to be but partly understood, if at all, by speakers of the latter," and Drucker (1937:283) considered the Galice dialect to be "the most divergent of all the groups," including

[7] Barnett's information came from Billy Metcalf, who, although self-identified as a Chetco, seems to have spoken a variety that was indistinguishable from the Rogue River varieties, as discussed in chapter 4.

Chetco-Tolowa. The status of Galice and Chetco-Tolowa is revisited in chapter 4, but the available evidence suggests that these two groups were peripheral members of a shallowly differentiated continuum centered on the Tututni/Lower Rogue and Upper Coquille varieties. Reports suggest that within this area at least some speakers had passive competence in multiple varieties. The variety spoken on the Upper Umpqua River, however, seem to have fallen outside this continuum, and most likely constituted a distinct Oregon Athabaskan language (see chapter 2, Appendix C).

Published accounts of Oregon Athabaskan languages include wordlists of Upper Umpqua (Scouler 1841, Hale 1846) and Tututni (Hubbard, published as Taylor 1860); article-length grammatical sketches of Chasta Costa (Sapir 1914), koineized Tututni (Golla 1976), and Galice (Hoijer 1966); analyses of Tolowa phonology (Bright 1964, Collins 1989) and morphology (Collins 1985, Givón 2000, Givón and Bommelyn 2000); a dictionary of Tolowa (Bommelyn 1989); a list of Galice noun and verb stems (Hoijer 1973); and short texts in Galice (M. Jacobs 1968) and Chetco (E. Jacobs 1968, 1977). Dorsey (1889, 1890) provides a few general grammatical and lexical notes about the Oregon Athabaskan languages he studied. Landar (1977) published wordlists for Tututni, Mikwanutni, and Galice: the first of these originated with J.P. Harrington and the others are Landar's transcriptions of audio recordings made by Hoijer and Morris Swadesh.

While the published Oregon Athabaskan resources are substantial, a large quantity of unpublished material also exists. The most important archival materials, both in quality and breadth of coverage, are Dorsey's vocabularies from his visit to Siletz in 1884 (Dorsey 1884a-n) and Harrington's notes from field trips to Grand Ronde and Siletz in 1940 and 1942 respectively (referenced in this study as Harrington 1981, the date when the microfilm copies of these notes were published). Dorsey's materials are especially important because he documented all known Oregon Athabaskan varieties, often with multiple speakers, and also because his transcriptions, like Curtin's for Hupa, are relatively sophisticated for the period. Harrington also worked with several speakers, and his phonetic transcriptions are, as usual, of the highest quality. Much less reliable than Dorsey and Harrington is Everette's (1882) documentation of Tututni. Nonetheless, Everette's data are an important complement to Dorsey's since they were collected in the same period and used the same Bureau of American Ethnology vocabulary schedule, providing numerous points of direct comparison from different transcribers.[8] Other unpublished sources consulted for this study include early wordlists from Kautz (1855) and Abbott (1858); field notes from Waterman (n.d.), Goddard (1902-1903), Barnett (1934), and Drucker (1934); and a draft lexicon of Tututni prepared by Golla (2008).

## 5.4. Transcription Conventions

Given such a wide variety of documentation collected over so long a time span for such a large number of languages, some of the data in this thesis have been normalized for ease of comparison; other data retain more or less the same orthography as found in the original sources. In general, the earlier the source and the more idiosyncratic the transcriptions, the more likely it is that forms have been normalized. Normalization has often involved simple transliteration, and

[8] Also of special importance are unpublished texts collected by Melville and Elizabeth Jacobs (1930s) now held by the University of Washington library, and Pliny Earle Goddard's field notes from a trip to Siletz in the first decade of the 20th century at the Columbia University library (recently discovered by Athabaskan scholar Chris Cox). These materials have not been consulted in the present study.

for the most part no systematic attempt has been made to correct transcriptions where a given researcher might not have transcribed certain distinctions consistently. For example, in many cases it is possible to guess where Goddard (or any number of other researchers) failed to transcribe glottalization, or where it is exuberantly over-transcribed, but the normalized forms rarely reflect these judgments, especially where it is peripheral to the point at hand. Where forms have been normalized, in many cases representations of the original transcriptions are also provided in angle brackets, although these are sometimes omitted where they interfere with legibility in tables.

For the most part, Oregon Athabaskan normalized forms in chapter 4 use the orthography presented in §5.1 above, modulo some additions reflecting changes within the subgroup (e.g., /θ/ and /ʂ/ for interdental and "retroflex" fricatives). Normalized Hupa forms in chapter 3 and chapter 5, including examples from the texts in Sapir and Golla (2001), are rendered in the practical community orthography found in the Hupa Language Dictionary (Golla 1996b), which employs conventions that are familiar to people approaching the languages without previous training in linguistics. This includes <ch> and <j> for palatal affricates, <wh> for /ʍ/, <u> for centralized short /a/ (i.e., [ə]), <ng> for /ŋ/, and apostrophe for glottal stop. Readers are referred to Golla (1996b) for more complete details. Hupa material transcribed by Goddard has been left as in the original source; conventions that might be unfamiliar to some readers are Goddard's use of <tc> for a postalveolar/palatal affricate, <ñ> for a velar nasal, italic <*w*> for a voiceless labiovelar fricative, capital <L> for a glottalized lateral affricate and small-caps <ʟ> for a voiceless lateral fricative. The Bureau of American Ethnology alphabet used by many of the 19th century sources, especially Everette (1882), Dorsey (1884a-n, 1890, 1891), and Curtin (1888-1889) may also present difficulties for some audiences. Likely points of confusion are <q> (a voiceless velar fricative, not a stop as in later sources), <rx> (a voiced velar fricative, usually transcribed as either <r$^{x}$> or <$^{r}$x> by Dorsey), <c> (a voiceless palatal fricative), <ç> (a voiceless interdental fricative), and a digraph <çl> invented by Dorsey to represent a voiceless lateral fricative (Curtin used <hl>). Readers unfamiliar with these conventions are referred to Powell (1880) and the concordance in Golla (2011:283-286) for guidance.

A final notational point concerns the representation of parameters of variation, especially in chapters 3 and 4, where it is not always clear whether the variation is phonemic or sub-phonemic. Square brackets are generally used when discussing the sounds in isolation, but this should not be considered a commitment to their being allophonic rather than categorical. Hence "[k] ~ [x] variation" can be considered as shorthand either for "variable phonetic realizations [k] and [x] of a single phonological category" or as "variation involving two categories with characteristic phonetic instantiations [k] and [x]." This point is non-trivial, since two different claims about the nature of the variation is implied in the two cases, but in general the data considered in this study are insufficient to be able to resolve this question definitively. The choice of square brackets implies simply that at least there is a reasonable degree of certainty that the sounds that were transcribed were acoustically different from one another even if their categorical status is unclear.

## 5.5. Citation of Sources

Citation conventions used for examples are often ad hoc and depend in part on how a given resource was accessed. Many of the Hupa examples cited in chapter 3 are drawn from texts or digital images of archival documentation that are cumbersome to cite in full for each instance;

conventions that deviate from the standard YEAR:PAGE schema are as follows:

| Source | Example | Interpretation |
|---|---|---|
| Goddard (1904, 1914b), Golla (1984) | 325.7 | page 325, line 7 |
| Golla (1963) | 2.118 | notebook 2, page 118 |
| Sapir and Golla (2001) | 13.28 | text 13, line 28 |
| Merriam (1850-1974) | 52/38 | reel 52, frame 38 of the microfilm copy of Merriam's papers |
| Woodward (1953) [file slips] | Woodward.001.012.0082 | digital image 82 of item Woodward.001.012 in the Survey of California and Other Indian Languages |
| Bright (1950-1952) | | |
| Woodward (1953) [texts] | Woodward.002.001, IV.3 | text IV, line 3 of notebook Woodward.002.001 |

Table 1.2: Conventions for citation of select Hupa archival data

Recordings with Verdena Parker made by the Hupa Language Documentation Project are cited by the identifier of the digital recording, e.g., VP-310513-05.

# Chapter 2: Pacific Coast Athabaskan in Comparative Perspective

## 1. Introduction

Historical linguistic research over the past decade has been invigorated by the application of computational approaches to phylogenetic inference developed in the biological sciences. These methods have shed new light on previously intractable problems, and in some cases sparked new controversies, in language families as diverse as Indo-European (Gray and Atkinson 2003; Atkinson and Gray 2006; Nakhleh, Ringe, and Warnow 2005; Nakhleh et al. 2005; Bouckaert et al. 2012), Bantu (Holden and Gray 2006, Marten 2006), Austronesian (Bryant 2006, Dunn et al. 2008), and Pama-Nyungan (Bowern and Atkinson 2012). This chapter applies these methods to the Athabaskan language family, specifically addressing the question of the historical status of the California and Oregon Athabaskan languages, both with respect to each other and with respect to the rest of the family. These languages are canonically discussed as a single Pacific Coast group due to their geographical proximity to one another and isolation from other Athabaskan-speaking groups. However, there is disagreement in the literature about whether PCA is a meaningful phylogenetic subgroup or is instead a regional grouping with no higher-order significance within the family.

While the main goal of this chapter is to shed light on the status of PCA, the discussion is set against a backdrop of broader theoretical and methodological interest. First, previous research has suggested that applying branching tree-like models to the Athabaskan family has not been successful. Krauss (1973) in particular has argued that the patterns of lexical and phonological isoglosses found throughout the family (especially the Northern Athabaskan languages – cf. Krauss and Golla 1981) indicate a historical development involving local innovations that spread more or less widely within a dialect continuum. Given this historical scenario, it is of interest to determine how computational models perform where there are no well-established subgroups to compare the results against – and indeed, where it has been suggested that it is futile to seek such subgroups in the first place. The Bayesian phylogenetic inference models used in the present analysis are explicitly geared toward providing metrics for the degree of certainty associated with particular subgroups in tree topologies and therefore can quantify the extent to which meaningful tree-like structures can be found within the family.

Second, one of the goals of the present study is to incorporate both lexical and non-lexical evidence in the computational phylogenetic analysis of PCA. Most current computational historical linguistic research focuses on lexical evidence alone, for reasons to be discussed in more detail below. However, in traditional historical linguistics there are sometimes conflicts between evidence from different domains – lexical, phonological, and morphological. Lexical evidence is usually given lowest priority in arguments for subgrouping because words are considered more prone to borrowing than non-lexical linguistic features. Shared innovations in phonology or morphology, especially where they are highly idiosyncratic, are considered the strongest evidence that similarities between languages are due to a shared genetic inheritance instead of linguistic contact events or parallel development (Ringe, Warnow, and Taylor 2002:65-69). There is thus a deeper theoretical question of why different linguistic domains sometimes do not give consistent diachronic results, mismatches reflecting social and structural realities that demand explanation. Excluding non-lexical evidence from analyses makes it impossible to even consider such questions. Nonetheless, while there are by now some well-established (if not completely uncontroversial) methods for modeling lexical change, the same models are arguably inappropriate for non-lexical characters, and results based on them pre-

sented in §5 will therefore be more tentative.

Another methodological point of interest involves the way linguistic features ("characters" in the evolutionary idiom) are coded. Gray and Atkinson (2003) presented a method for coding lexical datasets according to which characters with multiple states (e.g., where there are more than two cognate sets associated with a semantic category in a wordlist) are recoded in a binary scheme, with each language assigned state '1' if a cognate in a given meaning is present in the language and state '0' otherwise (either the cognate is not found in the language at all, or it has some other meaning). Evans, Ringe, and Warnow (2006) argue that this recoding creates dependencies among characters that have unknown consequences and potentially lead to biased results. Pagel and Meade (2006), however, have argued that dependencies among characters introduced under binary recoding will have only a minor impact on computational results and merely create scaled versions of the best topology – trees with shorter branch lengths and higher posterior probabilities for subgroups, but without major changes to the subgroups themselves. The present study considers results based on both multi-state and binary codings in light of Pagel and Meade's arguments.

The chapter is structured as follows. Section 2 summarizes previous attempts to classify the PCA languages. Section 3 provides an overview of the languages and data sources used in the computational study. Section 4 provides an overview of the computational techniques employed and details of particular coding decisions that were made, highlighting especially problems surrounding uncertainty and ambiguity in the data and the analysis of non-lexical characters. Section 5 presents results obtained from lexical, non-lexical, and combined datasets. The main finding is that the Pacific Coast Athabaskan languages do indeed emerge as a subgroup, one that is well-supported with different data types (lexical vs. non-lexical), data codings (multi-state vs. binary), and assumptions about rates of change across lineages (non-clock, strict clock, and relaxed clock evolutionary models). Other findings of interest include differences in tree topologies under multi-state and binary codings of the data and the effects of including non-lexical characters in the dataset (modulo some caveats about the appropriateness of the models). Section 6 summarizes these findings and concludes.

## 2. Classifying Pacific Coast Athabaskan Languages: Previous Work

### 2.1. Hoijer (1960)

The first systematic attempt to classify the PCA languages is found in Hoijer (1960).[1] Focusing

[1] Some of the earliest statements concerning the historical relationships obtaining among the PCA languages are found in various writings of Pliny Earle Goddard. Goddard (1903:91) noted that Hupa differed noticeably from other PCA languages and from the rest of the Athabaskan family, its unique character instantiated across linguistic domains: lexical, phonological, and morphological. He elsewhere observed that "the remaining languages of the Pacific Division of the Athapascan are rather closely connected and grade into each other" (1905:10), explicitly including languages of both the California and Oregon divisions in this comparison, implying an extended dialect network of the kind envisioned by Krauss (1973) for PCA and elsewhere in the family. Goddard seems to have considered the Pacific Coast languages a phylogenetic unit, but one that in his view included an unnamed Athabaskan language of the interior of British Columbia (1911:92) (perhaps the poorly-attested Nicola, as in Morice 1906:504). This connection between the geographically closest Northern Athabaskan languages and the PCA languages is also made by Harrington (1943), who claimed that PCA was closely related to Chilcotin. Harrington's claim, however, is rejected by Krauss (1973:919) and Krauss and Golla (1981:83), although both of those studies endorse the idea that PCA languages share some affinity with Athabaskan languages of British Columbia.

on the development of Proto-Athabaskan stem-initial consonants, the basis for comparison in most Athabaskan historical linguistic research, Hoijer maintained that the phonological data he considered "clearly differentiates a PCA sub-stock of the greater Athapaskan family and ... strongly supports the hypothesis that this sub-stock has two main divisions, the Californian ... and the Oregonian" (1960:969). He considered the following evidence crucial:

(a) the absence of what Hoijer calls "pitch accent" (elsewhere "tone phonemes", i.e., lexical tone) in PCA languages
(b) merger of the Proto-Athabaskan back-velar aspirated stop and fricative *q and *x̣
(c) retention of a labiovelar obstruents (*x̣$^{w}$, *q$^{w}$, *ɢ$^{w}$)
(d) the sound change *ɣ > g in some languages
(e) merger of *s and *z
(f) merger of *s/*z with *dz, and *ts in some languages
(g) merger of *š and *ž
(h) partial merger of *tš and *dž with *s (and *z, *dz, *ts) in some languages
(i) merger of *tɬ and *dl with *ɬ and *l respectively

Some of these pieces of evidence are stronger than others, in some cases in light of information about the PCA languages unavailable to Hoijer. Each one deserves careful scrutiny and they are considered in turn here.

(a) All PCA languages lack lexical tone. However, insofar as lexical tone is innovative in Athabaskan (Krauss 2005, Kingston 2005), this is a shared retention and therefore cannot be considered strong evidence for a PCA subgroup.[2]
(b) All PCA languages merge Proto-Athabaskan *q and *x̣, a change not found elsewhere in Athabaskan. They merge as /x/ in most Oregon languages and in Hupa, and as /k/ in Galice and in the rest of the California languages. Golla (2011:74-75, 81) reconstructs voiceless velar fricatives in the Oregon Athabaskan and California Athabaskan proto-languages, implying a subsequent fortition to /k/ in Galice and in all of California Athabaskan except Hupa. This scenario is consistent with a view whereby the merger of *q and *x̣ as /x/ occurred in a PCA proto-language.
(c) The evidence Hoijer presents does not support a hypothesized PCA subgroup. According to the data in Hoijer's tables, Proto-Athabaskan labiovelars are retained only in the Oregon languages (and *ɢ$^{w}$ only in a subset of those). Hoijer takes this as evidence that Proto-PCA retained the plain velar ~ labiovelar contrast, which was subsequently lost in California Athabaskan. However, this evidence presupposes the existence of the subgroup it is intended to support and is equally consistent with a view whereby the merger took place in a California Athabaskan subgroup with no higher-order connection to the

[2] Quasi-tonal phenomena are found in the Oregon languages. Bright (1964) described a pitch accent system for Tolowa, and Golla (1976:220) reported non-contrastive pitch phenomena conditioned by syllables ending in laryngeal consonants that appears to be the precursor of a low-marking tone system. Harrington's (1981) documentation of Oregon Athabaskan contains numerous minimal pairs where a coda laryngeal segment in Upper Coquille corresponds to falling pitch on a preceding vowel in Chasta Costa and Tututni, although it is unclear at present how systematic the phenomenon is.

Oregon languages.[3] This is again a reflection of the fact that shared retentions cannot be considered strong evidence for establishing subgroups.

(d) According to Hoijer's data, *ɣ shifts to /w/ in Hupa, variably to /g/ and /ɣ/ in Mattole, and to /g/ in the rest of California Athabaskan. In Oregon, it shifts variably to /g/ and /w/ in Galice and is retained in the other languages. The irregular outcomes in both Mattole and Galice are conditioned by considerations that are not strictly phonological: /g/ is found in all verb stems in both Mattole and Galice; /w/ in Mattole and /ɣ/ in Galice is found in noun stems and postpositions, perhaps with further conditioning in Mattole based on the nature of a preceding agreement prefix. This pattern, given Hoijer's inclusion of Galice in the Oregon division, is difficult to interpret as evidence for a higher-order PCA subgroup. The shared conditioning based on lexical category in Galice and Mattole, and the existence of /w/ in Galice and Hupa, could be evidence for an early set of shifts in a hypothesized Proto-PCA, but one that was innovative in Galice and California Athabaskan – implying that these languages are more closely connected to one another to the exclusion of the other Oregon languages.

(e) The voicing contrast in Proto-Athabaskan dental fricatives *s and *z is lost in all of PCA (although see Appendix C for discussion of Upper Umpqua) and retained in most other languages in the family (Hare is an exception – cf. Appendix A).

(f) The further merger of *s/*z with the corresponding Proto-Athabaskan dental affricates is found in all of Oregon Athabaskan and is considered by Hoijer as "partial" in all of California Athabaskan except Hupa, where Proto-Athabaskan *ts and *dz are retained as such. Li (1930a:10) maintained that /s/ and /ts/ appear "indiscriminately" in Mattole as reflexes of Proto-Athabaskan *s, *z, *ts, and *dz, and could find no principle governing their distribution.[4] Li relates this to the appearance in Wailaki and probably Kato of an affricated allophone of /s/ that appears when preceded by /l/ or /ɬ/, i.e., an epenthetic stop breaking up adjacent dental continuants.[5] Since the variation in Mattole and the epenthesis in Wailaki and Kato appear to apply regardless of the Proto-Athabaskan source of /s/, it is conceivable that the merger was complete and the epenthesis is a more recent development. If so, the merger of these affricates probably applied in all of PCA except Hupa.

(g) Loss of the voicing contrast in *š and *ž is found in all of PCA, where they are realized as /š/ in all languages except Hupa, where they are /ʍ/, and Mattole, where they are palatalized/front-velar /x̂/. This change, however, can be effectively collapsed with (e)

---

[3] Moreover, Golla (1976) reports that labiovelar consonants in Tututni are not retentions from Proto-Athabaskan but rather are innovations in the vicinity of labial consonants and vowels – this appears to be the case in other Oregon languages as well. This raises the possibility that the merger of proto-Athabaskan labialized and non-labialized velars is a shared innovation in PCA, but there are sufficient ambiguities in the data to render this uncertain.

[4] According to Li, in Mattole [ts] is pronounced such that "the release for the t- is so faint that it is hardly audible," with some forms phonetically [s]. The precise nature of the variation is unclear – whether some words are consistently [s] (implying underlying /s/) and others consistently [ts] (implying underlying /ts/), or whether the same words might sometimes have [s], other times [ts], perhaps both for underlying /s/ and/or /ts/. Li's description seems more suggestive of the latter scenario.

[5] A similar epenthetic affrication of /θ/ (< *s, *z, *ts, *dz) following /ɬ/ at least is also found in Harrington's (1981) Upper Coquille field notes, but not in his documentation of Chasta Costa.

above as a loss of voicing contrast affecting all coronal fricatives.

(h) As in (c), Hoijer's own data do not support his conclusion that there is a PCA subgroup. *dž and *tš develop into a series often described as "retroflex" in most of Oregon Athabaskan (merging with /š/ in Galice). According to Hoijer, in some Oregon varieties these undergo a partial merger with *s (apparently in some stems but not others). They are retained as /dž/ and /tš/ in all of California Athabaskan. It is therefore unclear why Hoijer considers this evidence for a Pacific Coast Athabaskan subgroup, since it is also consistent with the view that the Oregon and California languages were distinct offshoots of Proto-Athabaskan, with the changes affecting *dž and *tš applying in one but not the other.

(i) The merger of lateral affricates with corresponding fricatives is found in all of PCA. This change can be collapsed with (f) above, only without additional loss of the voicing contrast between /l/ and /ɬ/. This suggests a single Proto-PCA sound change merging Proto-Athabaskan dental affricates with the corresponding fricatives that was incomplete in Hupa, affecting the laterals but not the plain dentals.

On balance then, Hoijer's claim for a Pacific Coast Athabaskan subgroup is supported only weakly by (a) (a shared retention of tonelessness and hence irrelevant for subgrouping) and not at all by (c) or (h). (d) and (f) are consistent with the PCA subgroup hypothesis but not with the California ~ Oregon split Hoijer proposes insofar as Hupa retains contrasts lost in the Oregon languages and all the rest of the California languages. His strongest evidence, then, is (b) (merger of *q and *x̣), (e) (merger of *s and *z), (g) (merger of *š and *ž), and (i) (merger of *dl and *tɬ with *l and *ɬ respectively). The merger of voiceless and voiced coronal fricatives in (e) and (g) can be collapsed, and (i) can be collapsed with (f).

## 2.2. Hoijer (1962)

In subsequent work Hoijer (1962) re-affirmed the existence of a PCA subgroup on the basis of phonological data, but also presented conflicting lexicostatistical evidence. This was based on pairwise comparison of three California languages (Hupa, Kato, Mattole) and three Oregon languages (Tolowa, Upper Coquille, and Galice), probably using Swadesh-100 wordlists similar to those presented in Hoijer (1956). With an arbitrary threshold of 77% shared cognacy required for two languages to be considered part of the same subgroup, Hoijer found that the Oregon Athabaskan languages had over 90% shared cognacy with one another, whereas shared cognacy rates among the California languages were at 75% and 76%. Shared cognacy rates between the California and Oregon languages were for the most part below the 77% cutoff. Thus, Hoijer found that lexical evidence failed to support a PCA subgroup or even a California Athabaskan subgroup.[6] Hoijer's results point out the tension noted in §1 between conflicting evidence from different linguistic domains. Here, the lexical evidence supports one classification, whereas phonological evidence supports a very different one.

## 2.3. Krauss (1973)

In his comprehensive overview of the Athabaskan languages, Krauss (1973) endorses Hoijer's (1960) conclusion that there is a PCA subgroup, maintaining that "it is clear that Pacific Coast

[6] Hoijer is inconsistent in this article, since he elsewhere says that "the Pacific languages all show cognate percentages of 77 percent or more" (1962:195).

Athapaskan split off from a relatively restricted or linguistically homogeneous sector of the North, and that a number of important phonological and morphological innovations are unique to and universal in PCA" (p. 919). He nonetheless expresses reservations: "Hoijer's own data themselves clearly show however conflicting cross-classification, necessitating arbitrary weighting of one type of change over another in evaluating degrees of divergence" (p. 924). Moreover, while tentatively accepting a high-level split of California and Oregon branches within PCA, Krauss maintains that additional phylogenetic structure is unsupported and that "... further evidence from other sources, mostly unpublished, makes any Stammbaum most inappropriate as a model for description of PCA relationships" (Krauss 1973:924).

This discussion reflects Krauss' broader claim that Athabaskan linguistic relationships are ill-represented with branching family-tree models: "it is not possible to classify the Athapaskan languages meaningfully solely or even largely on the basis of the Stammbaum model ... Athabaskan must be viewed as a dialect complex with many convergence ('wave') as well as divergence (Stammbaum) relationships" (1973:943).[7] According to this view, shared innovations in Athabaskan are due to diffusion of linguistic features across a large dialect network, and "[t]he most important differences among Athapaskan languages are generally the result of separate innovations from different points of origin, each language – each community – being a unique conglomerate" (Krauss and Golla 1981:69). Krauss demonstrates this in light of phonological data showing that several Proto-Athabaskan obstruent groups develop with little isogloss bundling making it hard to discern well-defined subgroups.[8]

While Krauss may be correct that Hoijer's data and interpretations were inaccurate, the core of his objections can again be cast as one of conflicting evidence: here, in the form of cross-cutting isoglosses even within a single linguistic domain, leading to what he considers arbitrary weighting of some pieces of evidence over others. However, as Bloomfield points out in his discussion of incompatible isoglosses in dialect geography, it may be possible to establish principled reasons for preferring some isoglosses – some linguistic features – over others in establishing degrees of linguistic relatedness (1933:341-342). Elsewhere Bloomfield (1933:318) points out that the Comparative Method works only with the assumptions of uniformity in the proto-language and of communicative isolation of the speech communities that underwent differentiation, preventing reticulations across nodes in the family tree. Bloomfield emphasizes that these must be understood as convenient fictions; as Atkinson and Gray have put it more recently, "[m]odels are lies that lead to the truth" (2006:94). Thus, while Krauss is certainly correct that there are cross-cutting isoglosses within the Athabaskan family that tend to obscure a phylogenetic signal, this does not mean that a family tree model is entirely inappropriate – only that the differentiation events implied by a tree are just one part of the overall story

[7] This is tempered somewhat in Krauss and Golla (1981:69), who allow for the possibility that "[a] geographically isolated subgroup such as Pacific Coast Athapaskan or Apachean can be treated as a historical unit," but they maintain that "for Northern Athabaskan it is relatively useless to search for the kinds of extensive correlations of morphological, phonological and lexical features that allow the establishment of a 'subgroup' with an assumed common prototype."

[8] Accordingly, Krauss is severe in his criticism of Hoijer's broader attempts to establish Athabaskan subgroups: "due to the obvious inadequacy of the criteria, and what should be an equally obvious inadequacy of the theory based exclusively on divergence (and ignoring all diffusion, convergences, and simple parallel innovations or retentions), and extremely frequent inadequacy of the data and often interpretation of the data, the results are predictably absurd from a genetic point of view" (Krauss 1973:944).

(Labov 2007, Babel et al. 2013, Kalyan and François (to appear)). Moreover, a lack of structure involving some languages in a family need not require that there is a lack of structure across the board.

## 2.4. Golla (2011)

Most recently, Victor Golla has argued that the Pacific Coast Athabaskan languages do not comprise a subgroup of Athabaskan. While acknowledging that there are some innovations shared between the California and Oregon languages, and between the poorly-attested Kwalhioqua-Tlatskanai of the Lower Columbia River, he discounts these as relatively unimportant. He suggests instead that the concentration of Athabaskan languages in the region is in effect a historical accident:

> "While internally showing only modest differentiation, these three enclaves of Athabaskan speech are sharply distinguished from one another by extensive phonological, grammatical, and lexical differences and share few common innovations beyond a general tendency to simplify and restructure inherited features. Rather than being treated as a classificatory node, the Pacific Coast Athabaskan languages are probably best viewed as a geographically defined group of isolated and divergent forms of Athabaskan that, at most, partially share a common history of southward migration from Canada." (Golla 2011:69)

Golla further points out that the PCA languages lack shared lexical innovations indicative of a common arrival in a new region prior to differentiation: "There is no linguistic evidence (such as shared lexical innovations for the distinctive flora, fauna, and culture features of the region) to suggest that the ancestors of the California, Oregon, and Lower Columbia Athabaskans had a common homeland along the Pacific Coast. It is thus a plausible assumption that the Pacific Coast Athabaskan migration occurred in two or more stages or pulses" (Golla 2011:257). This can again be viewed as a problem of evaluating conflicting evidence – in this case, a small number of shared innovations versus a relatively large number of differences, and emphasizing a particular kind of lexical evidence as a key indicator of a common social and linguistic history.

## 2.5. Resolving Conflicting Evidence

The debate surrounding Hoijer's efforts to classify the PCA languages, and the rest of the Athabaskan family more broadly, reflect a basic tension between different criteria that could be invoked to establish subgroups. Hoijer (1962) found competing classifications when considering lexical versus phonological features within PCA. Conflicting evidence in the form of cross-cutting phonological isoglosses, and what he considers the arbitrary weighting of some shared features over others, is also at the heart of Krauss' lukewarm acceptance of Hoijer's (1960) arguments for a PCA subgroup and his insistence that PCA is not amenable to subgrouping below a California-Oregon split. Golla (2011) dismisses non-lexical innovations in PCA as mere simplifications and restructurings, and values instead lexical evidence from particular semantic domains – those that would be indicative of a shared migration into a new physical and cultural environment.

Resolving the tension between potentially conflicting evidence from different linguistic domains is an area of longstanding interest in historical linguistics. The computational methods

used in the present study are well-suited to addressing this issue: although there remains a danger of biasing results through an inappropriate selection of linguistic features relevant to inferring phylogenies, in principle all characters in a dataset contribute to the overall picture, even when some of the characters imply conflicting interpretations of the data. This is because the models invoked here do not include any inherent bias preferring some features over others (although it may be appropriate to impose such biases, such as the infinite weighting of non-lexical characters in Nakhleh et al. 2005). Rather, the methods seek to discover the phylogenetic tree (or set of trees) that render the observed data most likely given a model of evolution. Conflicting data lead to weak support for particular subgroups, and it is possible to quantify the degree of confidence in branching events reflected in trees produced as part of the analysis.

## 3. Languages and Sources

### 3.1. Lexical Data

The core lexical data for this study were harvested from the glottochronological study in Hoijer (1956), which contains Swadesh-100 lists for fifteen Athabaskan languages: the Northern languages Beaver, Dakelh (Carrier), Dëne Sųłiné (Chipewyan), Gwich'in (Kutchin), Hare, and Tsuut'ina (Sarcee); the Apachean languages Chiricahua, Jicarilla, Lipan, Navajo, and San Carlos; and the Pacific Coast languages Galice, Hupa, Kato, and Mattole. Nine other languages were added to this set: the Pacific Coast languages Tututni (Golla 2008), Tolowa (Bommelyn 1989), and Wailaki (Li n.d.), the Northern languages Ahtna (Kari 1990), Dena'ina (Kari 2007), Koyukon (Jetté and Jones 2000), and Minto (Krauss 1963, Kari 1994), and the Na-Dene languages Eyak and Tlingit (Krauss 1963). The PCA data were supplemented by information from various published and archival sources: Bright (1964), Collins (1989), Dorsey (1884i), Goddard (1902-1903, 1923), Harrington (1981), Hoijer (1973), Landar (1977), Merriam (1850-1974), Waterman (n.d.), and Tuttle (n.d.). Lexical data for some Northern Athabaskan languages were also confirmed with dictionaries available from the Alaska Native Language Archive website: Jones (1978), Minto Nenana Athabaskan Noun Dictionary (1974), and Peter (1979).

### 3.2. Non-Lexical Data

Beyond the 15 languages considered by Hoijer (1956), languages were added to this study that had both accessible lexical data and grammatical descriptions deemed robust enough to allow non-lexical characters to be harvested. Many Athabaskan languages are not adequately described in relevant areas of grammar to allow for this kind of analysis. This is true for many Pacific Coast Athabaskan languages as well, but analysis of the texts in Goddard (1909, 1923), Seaburg (1977), and M. Jacobs (1968) supplemented the available published descriptions (non-existent for Wailaki, limited to Goddard 1912 for Kato). Table 2.1 summarizes the main language-specific sources consulted for non-lexical information:

| Northern Athabaskan | |
|---|---|
| Ahtna | Kari (1989, 1990) |
| Beaver | Goddard (1916, 1917), Randoja (1990) |
| Dakelh (Carrier) | Morice (1932), Poser (2005) |
| Dena'ina | Tenenbaum (1978) |
| Dëne Sųłiné (Chipewyan) | Li (1933, 1946), Cook (2004) |
| Gwich'in (Kutchin) | Leer (1996) |
| Hare | Rice (1989) |
| Koyukon | Jetté and Jones (2000), Thompson (1993) |
| Minto | Tuttle (1998) |
| Tsuut'ina (Sarcee) | Li (1930b), Hoijer and Joël (1963), Cook (1984) |
| Pacific Coast Athabaskan | |
| Galice | Hoijer (1966, 1973), Jacobs (1968), Tuttle (n.d.) |
| Hupa | Golla (1964, 1970, 1996a), Sapir and Golla (2001) |
| Kato | Goddard (1909, 1912) |
| Mattole | Li (1930a) |
| Tolowa | Bright (1964), Givón and Bommelyn (2000) |
| Tututni | Golla (1976), Sapir (1914) |
| Wailaki | Goddard (1923), Seaburg (1977), Li (1930a) |
| Southern Athabaskan | |
| Chiricahua | (Hoijer 1938, 1945), also Leer (1987) for Navajo |
| Jicarilla | |
| Lipan | |
| Navajo | |
| San Carlos | |

Table 2.1: Language-specific sources for non-lexical characters in the study

Several comparative studies were also invaluable, especially Cook (1996), Golla (1964, 2011), Hoijer (1960, 1963, 1971), Howren (1971), Kari (1993), Krauss (2005), Krauss and Golla (1981), Krauss and Leer (1981), Leer (1979, 2000), Rice (2000), Rice and Saxon (2005), Story (1989), Thompson (1996), and Wisniewski (n.d.). Information concerning the particular non-lexical characters selected for this study (described in Appendices A and B to this chapter) is for the most part unavailable for Tlingit and Eyak (modulo some sound correspondences for Eyak and Athabaskan sonorants published in Krauss and Leer 1981); no attempt was made to code non-lexical data for these languages, since they were included in the study mainly in order to produce outgroup-rooted trees for the lexical analysis.

## 3.3. Limitations of the Data

More than 40 Athabaskan languages have been identified in the literature, many of them with significant dialect variation. Only 22 of these languages (plus Eyak and Tlingit) were included in this study, due mainly to practical considerations. As already noted, languages were added largely on the basis of having both accessible lexical data and reasonably robust grammatical descriptions (or, in the case of PCA, adequate archival and text material to allow such characters to be inferred). This decision has meant that Upper Umpqua, a distinct language in the

Oregon group according to the classification of Golla (2011:70), has not been included in the main analysis since the documentation is too sparse for morphological characters to be gleaned with any confidence. Appendix C, however, considers the status of Upper Umpqua in light of lexical and phonological evidence. In addition to keeping the amount of data collection and coding in the project to a manageable scale, there was also the consideration that integer-valued "standard" data in MrBayes, the software used for this study, is limited to a maximum of ten states per character. Since each language in the study has a certain number of unique innovations, including a larger set of languages would tend to push characters over the 10-state limit and make them unusable under a multi-state character coding.

Restricting the set of languages in this way exposes this study to one of Krauss' (1973) main criticisms of Hoijer's phylogenetic research on Athabaskan: discrete, well-defined boundaries distinguishing Athabaskan languages from one another are an artifact of the fact that documentation was collected at widely-dispersed checkpoints, thereby obscuring internal diversity and linguistic features that have diffused regionally within Krauss' Athabaskan dialect complex. This criticism can be countered to some extent insofar as the present study has accounted for a moderate amount of dialect diversity by coding some characters as polymorphic (as discussed in more detail below). Moreover, the dataset is surely adequate to allow reasonable inferences related to the status of PCA, the main research question addressed in this chapter. However, Krauss' point is well-taken, and therefore for Northern Athabaskan especially the results reported in §5 must for this reason be considered provisional.

## 4. Computational Methods

### 4.1. Overview

The present study uses Bayesian phylogenetic inference to address questions of subgrouping related to Pacific Coast Athabaskan. The fundamental premise of this and other recent computational approaches is that at least some linguistic features are heritable traits that evolve in a manner that can be understood as analogous to evolution in biology. As discussed by Gray and Atkinson (2003), such models offer major improvements over the largely discredited lexicostatistics and glottochronology of yesteryear (Lees 1953, Swadesh 1955, Bergsland and Vogt 1962).[9] Some of the main objections to those earlier approaches centered around (a) the information loss associated with collapsing characters into summary scores of percentage similarity (as in Hoijer's 1956 study of Athabaskan discussed above) and (b) assumptions about the universality of rates of change. As Gray and Atkinson point out, objection (a) is largely overcome by the fact that many current models allow phylogenies to be derived directly in terms of the linguistic characters themselves rather than from summary similarity scores. Moreover, (b) is no longer a major concern because models can allow variation in the rates of change both for different linguistic features and for different lineages.

The methodology adopted here is summarized schematically as follows, following the discussions in Nichols and Warnow (2008), Atkinson and Gray (2006), Pagel and Meade (2006), and Bowern and Atkinson (2012).[10] First, a data matrix is assembled according to the procedure

[9] It is perhaps noteworthy that Athabaskan languages figured in these early approaches, as in Hoijer's (1956) glottochronological study (with the response by Hymes 1957). Also notable is Dyen and Aberle's (1974) reconstruction of the Proto-Athabaskan kinship system using clustering algorithms.

[10] A detailed overview and critique of Bayesian approaches to phylogenetic inference can be found in Felsenstein (2004, ch. 16).

outlined below whose rows are languages ("taxa") and whose columns are linguistic features of interest ("characters" or "sites"). This matrix is the input to software that performs the phylogenetic analysis – in the present study, the program MrBayes (Ronquist et al. 2012) was used. The goal of the analysis is to infer the evolutionary tree that makes the observed data – the set of states associated with various languages in the data matrix – most likely given a model of evolution. Ideally the likelihood of every tree would be inspected and the winning candidate(s) selected, but the large number of possible trees makes this computation effectively impossible. Therefore, Markov chain Monte Carlo (MCMC) sampling is employed to generate a set of trees from which the best phylogeny is inferred. An analysis proceeds by making incremental changes to a randomly-selected starting tree. At each step, if changes increase the likelihood of the data given the evolutionary model, the tree is added to the sample set. If the changes decrease the likelihood, the tree is retained in the sample some relatively low proportion of the time. At the end of the analysis – after a sufficient number of trees have been generated – the sample reflects the posterior probability distribution of the universe of possible trees. The sample can then be summarized to indicate the extent to which particular features (tree topologies and branch lengths) are found in the sample. The proportion of trees that a subgroup appears in can be taken as an approximation of its posterior probability – i.e., the degree of certainty it is part of the true evolutionary history of the language family.

## 4.2. Lexical Characters

### 4.2.1. Coding

As input to MrBayes, a data matrix was constructed whose rows were the twenty-four languages in the study and whose columns were the meanings in the Swadesh-100 list. Each cell in the matrix contained an integer indicating the cognate set(s) found in a given language in a given meaning, as shown in Table 2.2 for the meanings 'blood', 'two', and 'cloud' with a subset of the languages:

| | 'blood' | 'two' | 'cloud' |
|---|---|---|---|
| Navajo | 0 | 0 | 0 |
| Hupa | 1 | 0 | 1 |
| Kato | 1 | 0 | 1 |
| Tututni | 0 | 0 | 1 |
| Gwich'in | 3 | 0 | 0 |
| Eyak | 0 | 1 | 0 |
| Tlingit | 2 | 2 | 2 |

Table 2.2: Multi-state character coding

Taking 'blood' as an example, most Athabaskan languages and Eyak have a reflex of *dəɬ in this meaning, so these languages were assigned the integer 0 indicating that they all have the same state for this character. The California Athabaskan languages all have a word cognate with Hupa *tse:lin* in this meaning, so these languages were assigned character state 1. Gwich'in (Kutchin) (as coded by Hoijer 1956) and Tlingit both have distinct words for this meaning, so they were coded with unique integers '2' and '3' for their unique states.

For lexical data obtained from Hoijer (1956), Hoijer's original cognacy judgments were sometimes modified. This was especially the case for the PCA languages, but also sometimes for other languages in the family, often in light of the cognate sets published in Krauss and

Leer (1981) and Krauss (2005) or Leer's (2011) unpublished comparative lexicon. Cognacy judgments for Minto, Tlingit, and Eyak were taken from Krauss (1963). For Northern languages not included in Hoijer (1956), some of the remaining cognacy judgments were found in the original sources listed above, whereas others were my own. To deal with the 10-state limit in MrBayes, the following coding procedure was used. Where there were more than ten states within Athabaskan proper (i.e., excluding the Na-Dene congeners Tlingit and Eyak), the character as a whole was discarded – this affected one lexical character ('ash', item 66 on the Swadesh-100 list). Where the offending states were unique and found only in Eyak and/or Tlingit, which were included in the study mainly to produce outgroup-rooted trees, those languages were coded as having missing data. This affected seven lexical characters total for Tlingit (three characters), Eyak (one character), or both (three characters).

This procedure was repeated for all items and all languages, yielding a data matrix with 100 items for 24 languages. After discarding the item 'ash' with more than the 10 states allowed for integer-valued standard type data in MrBayes, an additional five items were uninformative under the multi-state coding retaining all polymorphisms.[11] The main lexical results reported in §5.1 are therefore based on a matrix of 94 informative characters coded for 24 languages, or 2256 data points. Of these, 41 (1.8%) had missing data coded using the character '?' (cf. Atkinson and Gray 2006).[12]

### 4.2.2. Polymorphic Characters

Nakhleh, Ringe, and Warnow (2005:384) and Nakhleh et al. (2005) argue that polymorphic characters – characters where at least one language is coded as having more than one state – are problematic for computational analysis because there are no widely-accepted and realistic models of how such characters evolve. Their solution was to remove from their dataset any character having at least one language with more than one state. The size of their dataset made this move relatively unproblematic: after removing 40 polymorphic characters, they still had 297 to work with.

Discarding entire characters was not feasible in the present study due to the relatively small number of characters and the large number of polymorphic or otherwise ambiguous characters, in part a reflection of the fact that Athabaskan does not have the same well-established canon of cognate sets in accessible published sources that Indo-European does. However, it is also in part in response to Krauss' (1973) critique of studies that simply discard evidence that is inconvenient for or inconsistent with a strictly branching view of language change. Since determining the extent to which branching structure can be detected in Athabaskan is one of the goals of the present study, retaining uncertainties (especially true polymorphisms) was desirable. From a technical perspective, this was mostly unproblematic: MrBayes can handle multi-state data with polymorphic and ambiguous character codings, and the binary recoding discussed below has as one of its principle advantages its ability to handle such data.

---

[11] Uninformative characters have the same state for all languages, sometimes with additional polymorphic states as well. For example, given a set of taxa most of which are coded unambiguously with state 0, the polymorphic coding {0,1} for a taxon is treated by MrBayes the same as if it were also coded as 0 (i.e., the character is uninformative in the analysis). Some characters became informative when polymorphisms were removed, so under some codings there were as many as 96 informative lexical characters in the dataset.

[12] Because of the 10-state limit in the software, the solution of Nakhleh et al. (2005) for dealing with missing data (coding missing data points with unique character states) was not generally available in the present study.

Overall, 199 out of 2376 cells in the full lexical data matrix (8.4%) had multiple states encoding uncertainty of one sort or another.[13] This included actual polymorphisms, cases where a language showed evidence of having more than one word fitting a given meaning in the Swadesh list. For example, for the item 'cold (of weather)', the Koyukon Dictionary (Jetté and Jones 2000) has three different roots: *-kk'ŭtl*, *edzoo*, and *edlee*, the last of which is also found in the Upper Inlet dialect of Dena'ina. Koyukon was assigned one state in the data matrix for each of these roots. A sub-case of polymorphism is dialect variation. For example, all dialects of Koyukon have a reflex of Proto-Athabaskan *dəɬ for 'blood', but the Central dialect also has a word built on the root *-kko* (literally, 'thing contained in a receptacle' – Jetté and Jones 2000). Koyukon was therefore assigned both state '0' and a unique state '4' for this character. Dialect variation is well-described especially for the Alaskan Athabaskan languages in the dictionaries prepared by Kari (1990, 1994, 2007) and Jetté and Jones (2000), and for some languages it could be inferred by comparing various sources of documentation (e.g., Li n.d., Merriam 1850-1974, and Goddard 1923 for Wailaki).[14]

Second, polymorphic coding was used for cases where cognacy judgments were deemed to be uncertain. These uncertainties received the same coding as true polymorphisms; for simplicity, I will refer to these cases as "polymorphic" as well. As noted, in some cases I disagreed with Hoijer's judgments; where this disagreement was tentative, multiple states were coded. In other cases the disagreement was inherent in the sources: different researchers sometimes offered conflicting judgments, and Krauss (1963) coded some pairwise cognacy judgments involving Hupa, Navajo, Minto, Eyak, and Tlingit as being '+/-' cognate with each other when they were similar enough to be considered possible cognates but not similar enough to be certain. Several items were treated as polymorphic where one or more languages had a morphologically complex word built on the same root as a simplex one in other languages – the idea being that a morphologically complex word might or might not be cognate with a simplex one depending on whether strict word equations for cognacy were enforced versus permitting relatively lax root equations. For example, Hoijer (1956) included for many Athabaskan languages a reflex of Proto-Athabaskan *k̲ə̂n 'stick, base' in the Swadesh list item 'tree'. Many Northern Athabaskan languages have this root in this meaning but with a d-initial prefixal element, presumably innovative. These languages were coded as polymorphic for this character, reflecting the fact that morphological complexity has been introduced that could render them non-cognate with other languages under an (extremely) strict word-based cognacy regime.

Nonetheless, since Nakhleh et al. report that their computational models performed better with polymorphic characters removed, analyses were also run under alternative codings that excluded some or all polymorphisms. Due to the large number of uncertainties in the dataset, the characters coded in this way were not left out entirely. Instead, where possible a language coded as having more than one state for a given character had one of its states designated as a default. This was done most often where one of the states was unique (e.g., due to a dialect variant) and hence uninformative for subgrouping. For polymorphisms that were the result of

[13] This figure excludes the discarded item 'ash' but includes the non-informative characters. It is reduced to 181 out of 2256 (8%) when the non-informative characters are excluded.

[14] Not all documented variation was included, however. For example, the so-called "elite replacements" in Dena'ina (Kari 2007) were excluded from consideration: as the term implies, these are clearly recent developments in the language that are nearly always restricted to particular dialects, and including them would only have served to proliferate non-informative unique character states in the dataset.

disagreements in the literature about cognacy judgments, default states were also sometimes assigned based on the relative reliability of sources, usually more recent ones based on better documentation. Still other defaults were assigned where there was dialect variation known to have resulted from diffusion from a neighboring language (e.g., the Upper Inlet dialect of Dena'ina discussed in Kari 1977). Finally, in cases of morphological complexity default states were sometimes assigned where one state would be informative for subgrouping. In the case of 'tree', for example, the more restrictive word equation isolating languages with the shared innovative prefixal element was set as the default state.

Overall, 123 polymorphisms in the full lexical dataset had default states designated, and alternative analyses were conducted on datasets that included only these defaults. Under one such alternative coding, remaining polymorphisms were retained; under another, languages with residual polymorphisms for a given character were recoded as having missing data. This was done only for multi-state codings, since polymorphisms are unproblematic under binary recoding. Some of the results obtained using these alternative codings are reported in §5.

### 4.2.3. Binary Recoding

Multi-state characters, where each character can have more than two states, can be transformed into binary characters according to a procedure introduced by Gray and Atkinson (2003) and Atkinson and Gray (2006) for Indo-European. The basic procedure is to consider each state for each multi-state character – i.e., each cognate set associated with a given meaning in a wordlist – as a binary-valued character in its own right. For example, as noted the item 'blood' is a character with five states: reflexes of *dəɬ (state 0) in Eyak and most Athabaskan languages, *tse:lin* in Hupa and the other California Athabaskan languages (state 1), and three unique states (2-4) for Tlingit, Gwich'in (Kutchin), and the Central dialect of Koyukon. This single multi-state character is then transformed to five binary-valued characters, one per cognate set, with each language assigned state 0 or state 1 for a character depending on whether it has that cognate in the relevant meaning. Hupa, for example, has state 1 ('present') for the cognate set *tse:lin* in the meaning 'blood', and state 0 ('absent') for the other four binary-valued characters.

This technique has been criticized by Evans et al. (2006:124), who point out that one of the foundational assumptions of the models invoked in most computational historical linguistic studies requires that characters evolve independently of one another. Because the innovation of a new association between a meaning and a form (whether by semantic shift of an existing word, or by neologism creating an entirely new cognate set) typically involves replacement of an existing association between that meaning and another form, binary recoding of multi-state characters introduces many dependencies between characters: if a language has state 1 ('present') for a given cognate in a particular meaning, it will usually have state 0 ('absent') for all other cognates associated with that meaning. For example, the innovation of *tse:lin* for the meaning 'blood' in California Athabaskan languages in all cases went hand-in-hand with the loss of the reflex of *dəɬ in that meaning. The fact of having *tse:lin* makes it highly likely (if not inevitable) that a language will not have a reflex of *dəɬ. Under the Gray and Atkinson binary coding, a language having state 1 for *tse:lin* is considered independent of its having state 0 for all other cognates in this meaning. Evans et al. (2006) argue that these changes are not independent of one another, leading to "extreme violations of the independence assumption" underlying the computational models, making this binary recoding "patently inappropriate."

Atkinson and Gray (2006) respond to this criticism by pointing out first of all that assumptions of independence also generally don't hold in the biological sciences where these models

were developed, yet nonetheless the models produce accurate phylogenies. Moreover, in their view the concern over dependencies between characters is overstated – they disagree with the assertion that violations of independence in their binary recoding are in fact "extreme." They also question whether multiple cognates in a given meaning are really atypical, pointing to the existence of polymorphic characters, which imply that the innovation of an association between a cognate and a meaning in some cases does not entail the loss of an existing association. They maintain that one of the merits of their cognate-based binary coding scheme is how easily it handles such polymorphisms.

Atkinson and Gray (2006) also maintain that Evans et al. (2006) do not demonstrate any actual problems with their results that are due to non-independence of characters. They follow Pagel and Meade's (2006) arguments that non-independence of characters has two effects. First, a binary recoding will tend to shorten the branch lengths of the tree returned by the analysis: given *k* states for a given multi-state character, branch lengths in a binary recoding are predicted to be scaled by the proportion 2/k, an effect that will be multiplied across characters in the dataset (although Pagel and Meade also point out that the actual effect may also depend also on the number of states associated with different characters and the likelihood of particular character-state pairings for a given branch length). Second, non-independence will tend to inflate posterior probabilities at each node but "should have only a small effect on the best topology." To test these predictions, the present study created two binary recodings of the original multi-state data matrix, one in which all states were retained, and another in which all unique states were discarded (Gray and Atkinson 2006:104). Results of the latter coding are reported in §5. Of particular interest is the fact that the consensus tree topology returned by the binary recoding is not just a scaled version of the one discovered under the corresponding multi-state coding. Instead, more substantive differences in the topology are found as well. This issue is discussed in §5.

### 4.3. Non-Lexical Characters

One of the goals of this study is to explore the consequences of including characters from different linguistic domains in the dataset – in particular, what impact the inclusion of non-lexical information has on phylogenies returned by the computational methods employed here. However, non-lexical data are problematic for a variety of reasons. These issues are summarized in this section, followed by an overview of the characters and coding used in this study, comparing especially the non-lexical characters found in Nakhleh et al. (2005) and related studies of Indo-European by the Computational Phylogenetics in Historical Linguistics research group (Ringe, Warnow and Taylor 2002; Nakhleh, Ringe and Warnow 2005). Details of the specific phonological and morphological characters used in this study can be found in Appendix A and Appendix B respectively.

#### 4.3.1. General Issues

As noted in §2 for attempts to classify the PCA languages, there is a fundamental tension in historical linguistic research surrounding the nature of the linguistic features that can be used to establish subgroups. The traditional view is that populations in contact with one another can easily transfer lexical items between their languages, whereas non-lexical features, especially highly idiosyncratic grammatical ones, diffuse across linguistic boundaries only with difficulty if at all (Ringe, Warnow, and Taylor 2002:65-69). Accordingly, comparative evidence based on non-lexical characters is considered the strongest kind for establishing phylogenetic relation-

ships: observed similarities are unlikely to be due to borrowing, and highly idiosyncratic grammatical changes are unlikely to be due to independent development. Some researchers, e.g. Nakhleh, Ringe, and Warnow (2005), go so far as to maintain that barriers to structural borrowing are effectively absolute; they point out moreover that the relative difficulty of structural borrowing is accepted to some extent even by those who maintain that in principle any linguistic feature can be borrowed. Thomason and Kaufman (1988), for example, establish a hierarchy of borrowability correlated with the "intensity" of contact: non-lexical features are borrowable, but only when there is a high degree of intensity in the contact situation.

By contrast, many current computational historical linguistic studies rely on lexical characters alone, which has led historical linguists who are wary of the effects of borrowing to be skeptical of the methodology. However, Atkinson and Gray (2006) argue that borrowing is less of a problem than sometimes supposed, since many borrowings can be weeded out from a dataset by the same procedures used in non-computational work. More recently, Greenhill, Currie, and Gray (2009) (reported in Bowern and Atkinson 2012) have argued that borrowing rates must approach 40% in order to impact detection of a phylogenetic signal. Moreover, while Atkinson and Gray (2006) concede that morphological and phonological characters might in principle be useful for phylogenetic analysis, they obtained tree topologies using lexical characters alone that are highly similar to those inferred by applying the Comparative Method to data from multiple linguistic domains.[15] The upshot of these arguments is that the problems associated with lexical data may not be as pernicious to constructing accurate phylogenies as previously thought, and it is unclear that results are any different when non-lexical characters are considered. However, the latter point especially remains to be verified in a wider range of cases.

The focus on lexical data in computational studies is in part a matter of convenience, since in general the basic vocabulary items in Swadesh or similar wordlists are relatively easy to assemble for a large number of languages. By contrast, many language families are not well enough described grammatically to permit compiling robust datasets of non-lexical characters. However, there are more substantive reasons to focus on lexical rather than non-lexical characters. First, Evans, Ringe, and Warnow (2006:121-122) point out that many evolutionary models are time-reversible, meaning that the evolutionary processes work the same way whether time is run forwards or backwards. They note moreover that many kinds of linguistic change do not have the property of time-reversibility, specifically mentioning phonemic mergers as a kind of irreversible linguistic change. This makes time-reversible models inappropriate for modeling many kinds of sound change. Second, there is the potential problem of non-independence: Whereas most lexical changes are arguably independent of one another, non-lexical changes can be implicated in correlated historical events: phonological chain shifts, for example, or broad morphological restructurings affecting multiple specific targets.

The problem of non-reversibility is also discussed by Dunn (2013), who points out that many phonological and morphological changes are one-time events. For the phonological characters coded by the Computational Phylogenetics in Historical Linguistics project, for

[15] In running alternative analyses using the results obtained by Ringe, Warnow, and Taylor (2002) as topological constraints, Atkinson and Gray (2006) implicitly concede that topological differences may result when non-lexical characters are included in analyses even in Indo-European. This is not a major problem for them, however, since their primary aim is to assign a date to the root of the Indo-European tree rather than discovery of the topology per se.

example, "a language either has or has not undergone a regular sound change (or set of regular sound changes) at some point in its prehistory" (Nakhleh, Ringe, and Warnow 2005:394-395). Having undergone a change, even if further changes obscure or undo its effects, a language cannot revert to the state of not having undergone it: transitions from state 1 ('present') to state 0 ('absent') are ruled out in principle, a particular instantiation of non-reversibility. Such binary-valued, non-reversible phonological characters have at most one state transition in the history of a lineage and therefore they do not have rates of change in the way that lexical replacement does, where lexical meanings are stable sites of evolution that can gain and lose associations with cognate sets multiple times diachronically. Although the irreversibility of changes can be stipulated computationally, from a conceptual standpoint phonological changes are arguably inappropriate characters in rates-across-sites models like the one employed here. Similar considerations apply for many morphological changes as well, although the exponence of morphological categories can perhaps be treated as stable sites for rate-based evolution.

Despite these pitfalls, some computational studies have attempted to account for both lexical and non-lexical data. In a series of papers on Indo-European by the Computational Phylogenetics in Historical Linguistics group, Ringe, Warnow and Taylor (2002), Nakhleh et al. (2005), and Nakhleh, Ringe, and Warnow (2005) include 22 phonological characters and 13-17 morphological characters in their dataset, all rare or idiosyncratic enough to be considered a good basis for phylogenetic inference. Because these studies primarily use a Maximum Compatibility evolutionary model, the issue of imposing rates of evolution on isolated historical events (as found in a rates-across-sites approach) does not arise.[16] Their non-lexical characters are carefully selected to be as informative as possible for subgrouping and are assigned infinite weights so that any tree incompatible with them is rejected. Further details about their phonological and morphological characters are included in §4.3.3 and §4.3.4 below.

Although the inclusion of non-lexical characters in rates-across-sites Bayesian inference models is not unproblematic, it is not unprecedented. Nakhleh et al. (2005) explicitly compared results obtained with different computational methods, including those used by Gray and Atkinson (2003), using their Indo-European lexical and non-lexical characters. Nakhleh et al. found that topologies returned when non-lexical characters were included were not identical to those obtained from lexical characters alone, suggesting that one cannot in general rely on lexical data for a complete picture of the historical relationships obtaining within a language family. Thus, while the present study offers no good solutions to the problems associated with using non-lexical characters in a Bayesian inference model, it can be taken as a step, albeit imperfect, towards considering the full range of linguistic data available for inferring linguistic phylogenies. This is an essential part of resolving the current disconnect between computational approaches and traditional paper-and-pencil historical linguistics where shared phonological and morphological innovations, rather than lexical similarities, remain the gold standard in establishing subgroups.

### 4.3.2. Coding

The present study coded a total of 31 phonological characters and 14 morphological ones, all but one with a binary scheme indicating whether or not a particular feature or change is present

[16] Maximum Compatibility seeks the tree (or set of trees) that have the minimal amount of homoplasy – either backmutation (changing from one state to another and then back again) or parallel evolution (the same change occurring in separate lineages) (Nichols and Warnow 2008).

or absent in a given language. (The lone exception, #4 in Appendix B, is related to the morphological exponence of plurality in the Athabaskan verb template, which is expressed by one of three prefixes.) With 45 characters and 22 languages, there were 990 cells in the data matrix, of which 98 (10%) were coded as missing (a number that was greatly inflated due to the handling of dependencies between characters). 31 cells (3.1%) had polymorphisms (or other uncertainties coded as polymorphisms), which were treated in the same way as in the lexical dataset: default states were assigned to ten of them and additional computational analyses were run on datasets that filtered the polymorphisms to accept only default states.

Some pairs of characters were related to each other in distinguishing the presence vs. absence of a change on the one hand and the outcome of the change on the other. To reduce the impact of non-independence among such characters, languages that did not undergo a change at all were coded as missing for the related change indicating its outcome. For example, the presence or absence of the merger of *q and *x̣ was treated as one binary-valued character (#6 in Appendix A, with state 1 in all of PCA and state 0 elsewhere). The final outcome of the change (merger as /k/ as in most of California Athabaskan, as /x/ as in most of Oregon Athabaskan) was treated as another character (#7 in Appendix A), with PCA languages coded as state 0 if the outcome was /x/ and as state 1 if the outcome was /k/; non-PCA languages were coded as missing for this character. Similarly, some Athabaskan languages permit the Proto-Athabaskan optative *-u- to be used with first person subjects, while others allow both first and second person subjects (all allow third person subjects). This was coded as two binary-valued characters, one for the first person and one for the second person (#13 and #14 in Appendix B). Since no language allows second person subjects to the exclusion of first person subjects, those coded with state 0 (‘absent’) for the first person were coded as missing for the second person.

Although the binary coding of these non-lexical characters is suggestive of a state of affairs whereby state 0 reflects an ancestral state and state 1 an innovative one (as with many of the directed changes in the non-lexical characters used in Nakhleh et al.), there is no guarantee in the evolutionary model used here that states known from the Comparative Method to be ancestral will in fact be inferred as such. It is possible to obtain post hoc reconstructions of the probabilities of states at ancestral nodes for particular characters by re-running analyses in MrBayes with topology constraints (Ronquist, Huelsenbeck, and Teslenko 2011:78-80). This has not been done in the present study due to the amount of time required to re-run analyses that are already computationally intensive. This is not unusual for work using this methodology, which is more concerned with inferring phylogenies than with reconstruction per se. However, it is important to realize that resolving the disconnect between computational work and the results of traditional historical linguistics requires more than finding the same subgroups in tree topologies. Ultimately inferences about particular linguistic features hypothesized for proto-languages must be in alignment as well.

### 4.3.3. Phonological Characters

The phonological characters assembled by Don Ringe and Ann Taylor and employed by the Computational Phylogenetics in Historical Linguistics (CPHL) Project for the most part are ones that are known to be clade-defining within Indo-European. Moreover, they are unlikely to have arisen through parallel development: they are either typologically unusual and idiosyncratic individual sound changes, or combinations of common sound changes that are “reasonably unlikely to have co-occurred by chance” (CPHL Project 2007a:1). Their phonological character P20, for example, combines three sound changes that occurred in Celtic: merger of *ē with *ī,

and merger of *ō with *ū in final syllables but with *ā elsewhere. This approach has been criticized for selecting characters that in effect pre-determine the outcomes of their research; the CPHL researchers counter that incorporating results obtained using traditional methods is an advantage and that they "do not expect that pretending ignorance would lead to more realistic results" (CPHL Project 2007a:6) . Nonetheless, this does limit the broader applicability of their method: as Dunn et al. (2008:713) put it, the amount of specialist knowledge incorporated in their studies "effectively restricts this methodology to very well-understood language families."

Details of the 31 phonological characters used in the present study are given in Appendix A. In general, they differ in key respects from the ones used by the CPHL Project. They are all single changes coded as either having occurred in each language (state 1) or not (state 0), or, as noted above, in some cases they indicate the specific outcome of a change (such as merger of *q and *x̣ as /k/ or /x/ in PCA). Most of them lack the property of being idiosyncratic: many are unconditioned beyond a restriction to stem-initial or stem-final position, and many are typologically common (e.g., palatalization and coda nasal deletion). However, while the possibility of independent development of particular changes is not ruled out, it is expected that the likelihood calculations in the MCMC analysis will produce the same effect as the grouping of common changes employed in the CPHL studies. Thus, if several languages have collectively undergone a large number of common sound changes to the exclusion of other languages, trees that group these languages together will be preferred; the analysis will in effect determine which combinations of changes are "reasonably unlikely to have co-occurred by chance." The advantage of this approach is that it does not bias the results on the basis of pre-determined subgroups believed to exist in Athabaskan. Indeed, in keeping with the arguments in Krauss (1973) and Krauss and Golla (1981), it is unclear that subgroups of any sort should be expected to emerge from analysis of these changes – this is one of the empirical questions that the study seeks to address.

### 4.3.4. Morphological Characters

The morphological characters used in the CPHL studies (CPHL Project 2007b) are of several different types. One of them (M1: organization of the verb system) involves high-level systemic properties in the morphology of Indo-European languages. Others identify the presence or absence of a morphological category (e.g., M3: thematized aorist). Still others encode the morphological expression of a category, either the binary presence or absence of a specific innovative morpheme (e.g., M13: gerundive *-ndo-*) or multiple states corresponding to different morphological exponents. There is also one character related to the function of a particular morpheme (M4: *-ské/ó- as iterative, inchoative, or causative), and two related to whether or not specific morphological changes took place (M14: syncretism of 3sg. and 3pl. and M15: replacement of 2sg indicative by optative in strong preterite). For the most part the 14 morphological characters in the present study (described in more detail in Appendix B) are of the same general nature as the ones found in CPHL (2007b). One character, for example, codes the function of the Proto-Athabaskan morpheme *yə- (obviative in most Athabaskan languages, inanimate subject in Hupa and Kato; #11 in Appendix B). Another is related to the presence vs. absence of syncretism of the 1st person and 2nd person plural object agreement prefixes (#3 in Appendix B).

However, two characters are interestingly different from those used in the CPHL Project (2007b) in relating to the relative order of morphological elements in the notoriously complex

Athabaskan verb template. The suitability of using templatic ordering differences for subgrouping may not be accepted by all researchers, as there are two views in the Athabaskanist literature. The first is that templatic morphology of the kind found in Athabaskan is highly stable diachronically. Rice (2000) calls this property "global uniformity" – the fact that despite several centuries of separation the Athabaskan languages' verb templates all retain roughly the same key elements in roughly the same order. Rice contrasts this with what she calls "local variability" – the fact that local idiosyncrasies in the template can be found nonetheless. This diachronic stability of templatic morphology has made it possible to identify homologies with Eyak and Tlingit (e.g., Krauss 1965) that have led to the general acceptance of Na-Dene as a historical reality, and it has been crucial to Vajda's (2010) efforts to demonstrate a connection between Na-Dene and the Yeniseian languages of Siberia.

By contrast, Kari (1989) has suggested that the ordering of elements in the Athabaskan verb is susceptible to change: "[t]he very nature of the Athapaskan verb, where discontinuous constituents are altered by layers of derivations that are themselves discontinuous constituents, lends itself to reanalysis in individual languages and dialects. Prefixes can be anticipated to merge, slip, reduplicate, metathesize, and/or fuse" (p. 434). However, a tendency for reanalysis and templatic slippage in general does not mean that any particular such change would be exceedingly likely, and as with phonological changes it can be expected that an accumulation of shared changes of this sort would be evidence for common historical development. To my knowledge such differences have not been used as a basis for inferring historical relationships within the family, and while only two have been coded in the present study this seems like an area with rich potential for future research (provided adequate data can be assembled for a large enough number of languages).

Although the morphological characters in this study can be considered useful for diachronic analysis, it is important to point out that they were selected opportunistically on the basis of information available in the sources, not because they are necessarily the best characters for this purpose. Different ones may be more appropriate in this regard, but phenomena are not always well documented in all languages. For example, variation in the so-called "qualifier zone" (Kari 1993) of the verb template is well-described in many Alaskan Athabaskan languages, but not elsewhere in the family. Similarly, as Rice (2000:236-244) points out, Hupa and Kato stand out in the family as having object prefixes occurring closer to the verb stem than 3rd person deictic subject prefixes, but at present it is unclear whether or not the same ordering is found in the rest of PCA. Overall, additional analysis of individual languages is needed in order to assemble a more robust set of morphological features that will make phylogenetic inferences more precise. Nonetheless, the ones used in this study can be considered a step towards the inclusion of substantive morphological characters (as opposed to abstract typological ones, as in Sicoli and Holton 2013) in a computational analysis of Athabaskan.

## 4.4. Analysis and Software Settings

Analyses were conducted with the software MrBayes (Ronquist et al. 2012). Results were obtained using the lexical characters alone, the non-lexical characters alone, and the lexical and non-lexical characters combined. Each dataset was analyzed under the following codings:

1. Multi-state coding, all polymorphic states included
2. Multi-state coding, polymorphisms filtered to default states where available, residual polymorphisms retained

3. Multi-state coding, polymorphisms filtered to default states where available; residual polymorphisms coded as missing
4. Binary recoding, all character states retained
5. Binary recoding, unique character states excluded

In addition, three different evolutionary models were used: a non-clock model placing no constraints on the rates of evolution across lineages, a strict clock model that assumes all lineages evolve at a constant rate (analogous to the "glotto-clock" of glottochronology), and a relaxed clock model constraining rates of evolution across lineages but allowing for variation. Strict and relaxed clock models produce rooted trees, whereas non-clock models do not and must be rooted on an outgroup. Tlingit was designated as the outgroup to root trees produced from analysis of the lexical and combined lexical plus non-lexical datasets. Only the strict and relaxed clock models were run on non-lexical data since these characters were not coded for Tlingit or Eyak. For the most part the results reported here focus on conditions (1) and (5) run with the non-clock model (where lexical characters were used) and the relaxed clock model (for the non-lexical dataset). Interesting results obtained under other conditions are sometimes noted as well.

One of the main pitfalls in Bayesian phylogenetic inference is that there are no criteria guaranteeing that the universe of possible trees has been adequately sampled – that is, knowing how many generations are required to obtain a statistically reliable result, when the analysis is said to have "converged" (Nichols and Warnow 2008). One heuristic diagnostic for convergence is to run simultaneous analyses and monitor the average standard deviation of split frequencies, which is generated on the fly by MrBayes – effectively a measure of how similar the results of the analyses are (Ronquist, Huelsenbeck, and Teslenko 2011:21). For the present study, two simultaneous runs (the default in MrBayes) of between 4 million and 10 million generations were conducted until this convergence diagnostic fell below .01. Post hoc analysis of two additional convergence diagnostics – the Potential Scale Reduction Factor (PSRF) and Effective Sample Size (ESS) of model parameters – was conducted. PSRF values for each parameter should approach 1 as an analysis converges, and following Bowern and Atkinson (2012) an ESS greater than 2000 was sought. Where these diagnostics fell short, additional generations were added until they achieved the desired threshold. For analyses using relaxed clock models, ESS $>$ 1000 was ultimately accepted due to the large number of model parameters and the number of generations required to bring each analysis above the more stringent threshold.

Results reported in §5 are given in the form of majority-rule consensus trees, which show nodes occurring in 50% or more of trees in the sample after the first 25% were discarded as burn-in. The integer annotated on a given node indicates the percentage of trees it occurs in, which can be taken as an approximation its posterior probability, i.e., its level of support (Pagel and Meade 2006). Particular nodes will be said to have, for example "90% support", which can be interpreted to mean that they occur in 90% of the trees in the sample or, equivalently, that they have a posterior probability of .9. As Bowern and Atkinson (2012:829) point out, there are no well-established criteria for considering a node well-supported versus merely adequately supported. Their standards are adopted here, and a node with 80% support (i.e., posterior probability greater than .8) is considered well-supported. Subgroups not occurring in the consensus tree at all (with posterior probability less than .5) are considered unsupported.

As discussed in §4.3.1, a rates-across-sites model may not be appropriate for non-lexical

characters in part because sound changes and morphological changes are typically irreversible: A language has either undergone a change, or it has not, and having undergone it a language cannot revert to an earlier state. To account for this, some analyses were re-run by imposing an irreversibility constraint on these characters with the ctype command in MrBayes. Adding this constraint had a negligible effect on the results, so the discussion is based on results obtained without it. Reflecting the fact that lexical, phonological, and morphological changes are not necessarily the result of related historical processes, datasets with characters from more than one of these domains were partitioned and analyzed with some model parameters unlinked, specifically, the *statefreq*, *shape*, and *ratemultiplier* parameters in MrBayes.[17] This has the effect of allowing different linguistic domains to evolve partially independently of one another within each lineage.

## 5. Results

It is important to emphasize that the main question of interest in the following discussion is whether or not the Pacific Coast Athabaskan languages emerge as a well-supported subgroup of the Athabaskan family. The tree topologies and branch lengths presented here raise a host of other questions of potential interest within the family. Some of them relate to the tree topology within PCA itself, others to how PCA is related to other languages in the family, and still others to how non-PCA languages are related to one another. These additional points of interest will be noted in due course, but in general no special effort will be made to explain them.

### 5.1. Results from Lexical Dataset

Figure 2.1 shows the majority-rule consensus tree obtained from the 94 informative lexical characters in the dataset with all polymorphisms included, run with a non-clock model of evolution. This is the least complicated analysis insofar as it has only one type of data, no manipulations to the underlying multi-state coding, and the smallest number of model parameters in MrBayes.[18]

[17] *Statefreq* estimates the frequencies of character states, *shape* estimates the shape of the gamma distribution from which characters' rates of change are drawn, and *ratemultiplier* is a partition-specific coefficient on rates of change. I am grateful to Will Chang for this suggestion, and more generally for his patient explanation of many nuances of Bayesian inference.

[18] Trees were created with FigTree v. 1.4.0 (Rambaut 2006-2012).

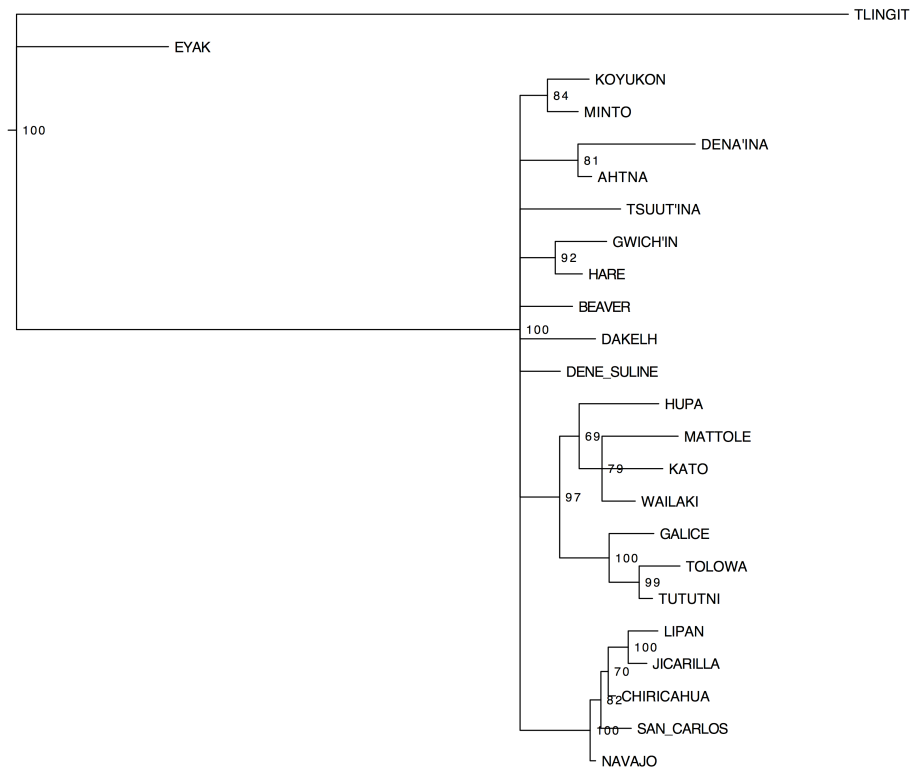



Figure 2.1: Majority-rule consensus tree for the lexical dataset coded as multi-state characters with all polymorphisms included (non-clock evolutionary model)

The main point of interest in Figure 2.1 is that a Pacific Coast Athabaskan subgroup emerges from the analysis, one with a high degree of support (97%). The Apachean languages also emerge as a subgroup with strong support; any other result would be shocking, since these languages are so similar to one another (Hoijer 1956). The Northern languages show a highly unresolved, rake-like branching – Gwich'in (Kutchin) and Hare, Koyukon and Minto, and Dena'ina and Ahtna appear as clades with over 80% support (over 90% for Gwich'in and Hare), but the rest of the Northern group does not resolve into a branching structure. This is consistent with the view of Krauss (1973) and Krauss and Golla (1981) that a highly structured family tree is an inappropriate model of Northern Athabaskan linguistic relationships.

Figure 2.1 can be taken as support for the view that the Pacific Coast Athabaskan languages are a well-defined subgroup within Athabaskan. Within PCA, the Oregon Athabaskan languages emerge as a subgroup with strong support (100%). A further subgroup including Tututni and Tolowa to the exclusion of Galice also receives strong support, favoring views such as Sapir (1914) maintaining that Galice was among the more divergent of the Oregon Athabaskan languages. Interestingly, there is only moderate support for a California Athabaskan subgroup comprised of Hupa, Mattole, Kato, and Wailaki (69%); the Mattole, Kato, and Wailaki subgroup falls just below the threshold for strong support at 79%. This is consistent with Hoijer's (1962) observation that a California Athabaskan subgroup is not well-supported

by lexical data.[19]

Analysis of the lexical characters was also conducted using the binary recoding procedure introduced in Gray and Atkinson (2003) and discussed in Atkinson and Gray (2006). The binary recoding was done two ways: one including all character states, and another according to which all unique character states were removed (which is intended to reduce the impact of non-independence introduced in the recoding – cf. Atkinson and Gray 2006:104). The consensus tree from the latter recoding is shown in Figure 2.2:

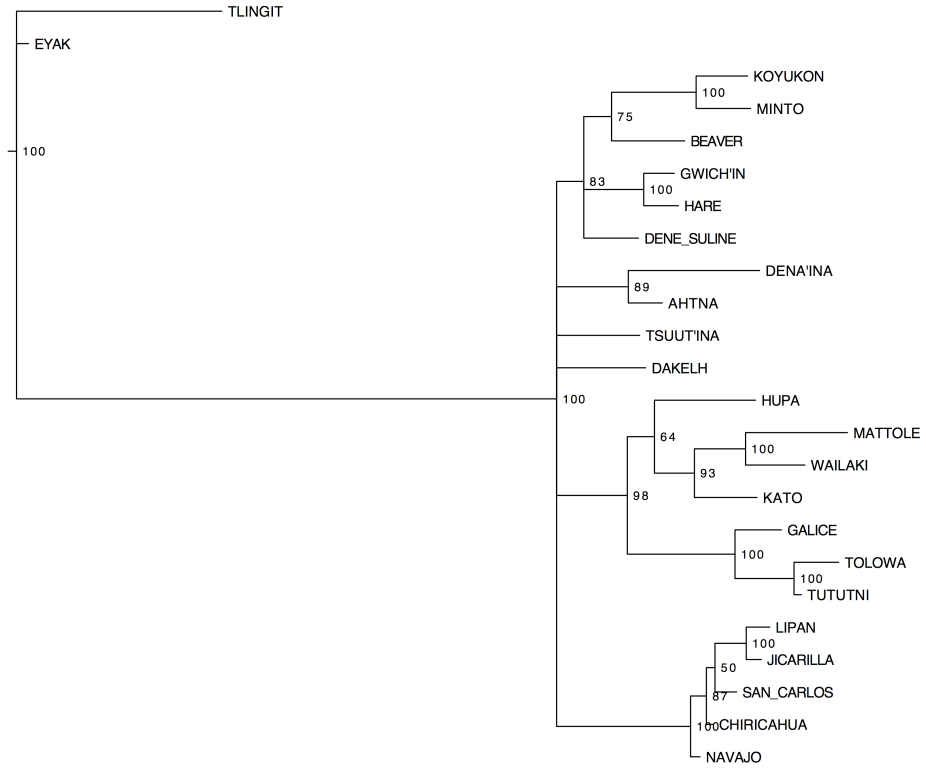



Figure 2.2: Majority-rule consensus tree for the lexical dataset coded as binary characters with unique character states removed (non-clock evolutionary model)

As discussed in §4.2.3, the binary recoding procedure introduces dependencies among characters, since in the usual course of affairs a language having one cognate in a given meaning implies that it does not have any other cognate in that meaning. According to Pagel and Meade (2006), the introduction of non-independent characters has two main effects: shortening branch lengths and increasing posterior probabilities. They argue that phylogenies produced with multi-state and binary codings will be scaled versions of each other, and non-independent characters introduced by recoding should not have a large impact on the best topology. Com-

[19] Analysis of the lexical dataset recoded with polymorphisms filtered to accept their default states (with remaining polymorphisms either retained or removed) produced consensus trees with only minor differences from the one in Figure 2.1. The topologies were identical except for some minor restructuring within California Athabaskan, and the emergence of a subgroup including PCA and Apachean languages (with relatively weak 60-61% support). The PCA subgroup had slightly lower (93%) support when polymorphisms lacking default states were excluded entirely.

parison of Figure 2.1 and Figure 2.2 suggests that, contrary to Pagel and Meade's predictions, binary recoding had a non-trivial impact on the topology returned by the analysis. For example, within PCA, a well-supported split between Kato and Wailaki + Mattole emerges in California Athabaskan. More strikingly, many of the Northern Athabaskan languages are subsumed under a well-supported (83%) subgroup, which also contains a subgroup with moderate support (75%) consisting of Beaver, Koyukon, and Minto.[20]

The differences between the topologies in Figures 2.1 and 2.2 may be due to the inflationary effect of non-independence on posterior probabilities anticipated by Pagel and Meade (2006), with some nodes achieving the 50% threshold required to be included in the consensus tree in Figure 2. Differences between Figure 2.1 and 2.2 are nearly consistent with this insofar as in most cases Figure 2.2 adds more structure than is found in Figure 2.1 but does not otherwise alter the topology. Note, however, that there is a minor restructuring in Apachean, with Lipan and Jicarilla grouped with Chiricahua in Figure 2.1 but with San Carlos in Figure 2.2. It is also noteworthy that one node in Figure 2.2 – the California Athabaskan subgroup subsuming Hupa, Kato, Wailaki, and Mattole – has slightly lower support (64%) than the corresponding node in Figure 2.1 (69%), although this may be due to the exclusion of unique character states in the binary recoding.

Alternatively, differences in topologies might be due to a disproportionate influence of characters with large numbers of states. (A similar point is made by Pagel and Meade with respect to the effect of binary recoding on branch lengths, and perhaps is implied for topology as well.) Pagel and Meade's argument concerning topology relies on the idea that given a tree, a model of evolution, and a set of characters only some of which are independent of one another, the true likelihood of the data can be obtained by raising the likelihood obtained from the non-independent characters to a power less than 1 (Pagel and Meade 2006:176), a calculation that does not depend on any particular characters in the dataset. However, binary recoding of characters with many states will introduce more dependencies than will binary recoding of characters with fewer states. Such uneven distribution of dependencies in the dataset may cause characters with more states in the multi-state coding to contribute disproportionately to likelihood calculations in a binary recoding. Trees that are well-supported by characters with many states will therefore be included in the MCMC sample set more often than they would if only independent characters were considered in the calculations, and certain nodes will receive support strong enough to alter topologies returned through summary consensus trees. The upshot is that one cannot assume that the best topology obtained under a binary recoding will simply be a scaled version of the topology obtained under the corresponding multi-state coding. This result suggests that it may be necessary to reconsider Pagel and Meade's arguments about the effects of character dependencies introduced by binary recoding.

[20] The shortening of branch lengths for Tlingit and Eyak in Figure 2.2 relative to Figure 2.1 is due to the exclusion of unique character states, many of which are found in those two languages. When unique character states are retained in the binary recoding, the long branch lengths for Tlingit and Eyak are restored. The topology is still roughly as in Figure 2.2, but with an additional well-supported (88%) subgroup subsuming all of the Northern Athabaskan languages.

To facilitate comparison with results obtained with the non-lexical dataset, analysis of the full lexical dataset (with all polymorphisms included) was also conducted using relaxed and strict clock models of evolution. The resulting consensus tree for the relaxed clock analysis is shown in Figure 2.3:[21]

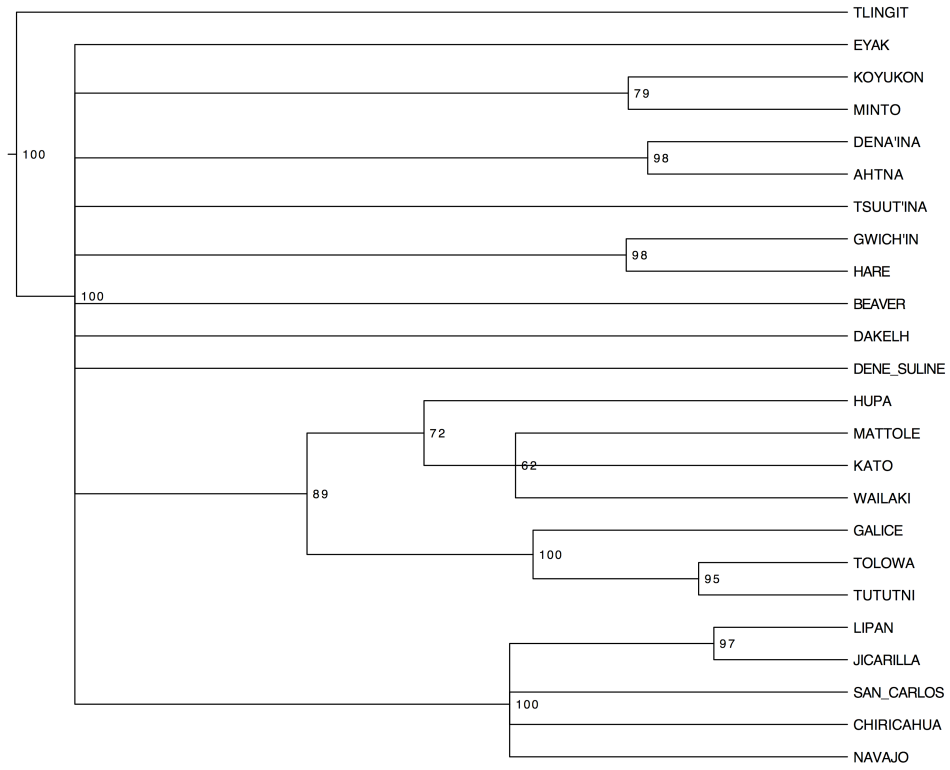



Figure 2.3: Majority-rule consensus tree for the lexical dataset coded as multi-state characters with all polymorphisms included (relaxed clock evolutionary model)

Notice that the subgroups in Figure 2.3 are similar to those in Figure 2.1. PCA again emerges as a subgroup, one with less support (89%) than under the non-clock model but still comfortably above the 80% threshold to be considered well-supported. Northern Athabaskan has the same rake-like branching as found in Figure 2.1, but the model does not produce the high-level Eyak-Athabaskan split, and there is less branching within Apachean. Moreover, the branch lengths within Athabaskan in Figure 2.3 are much longer relative to the root, suggesting a much earlier split for Athabaskan than is implied in Figure 2.1.

[21] The strict clock model produced much shorter relative branch lengths in Athabaskan than the relaxed clock model, implying a later split within the family. There were differences in the tree topologies as well. The strict clock, for example, finds a clade that includes the Northern and Apachean languages to the exclusion of PCA, with 67% support (i.e., the same order of support for California Athabaskan as found in the non-clock model). Bayes factors analysis shows that the relaxed clock model performed better than the strict clock model (BF > 5), in a range that is considered "substantial" but not "strong" support for one model over another (Bowern and Atkinson 2012:830). The relaxed clock model performed somewhat better than the non-clock model on this dataset (BF > 2), at a level that is considered only weak support for one model over another.

## 5.2. Results from Non-Lexical Dataset

Results obtained from analysis of the non-lexical characters, partitioned as described in §4.4, using a relaxed clock model are shown in Figure 2.4. Because non-lexical data were not coded for the Na-Dene languages Tlingit and Eyak, only relaxed and strict clock models of evolution were used in order to produce a rooted tree: [22]

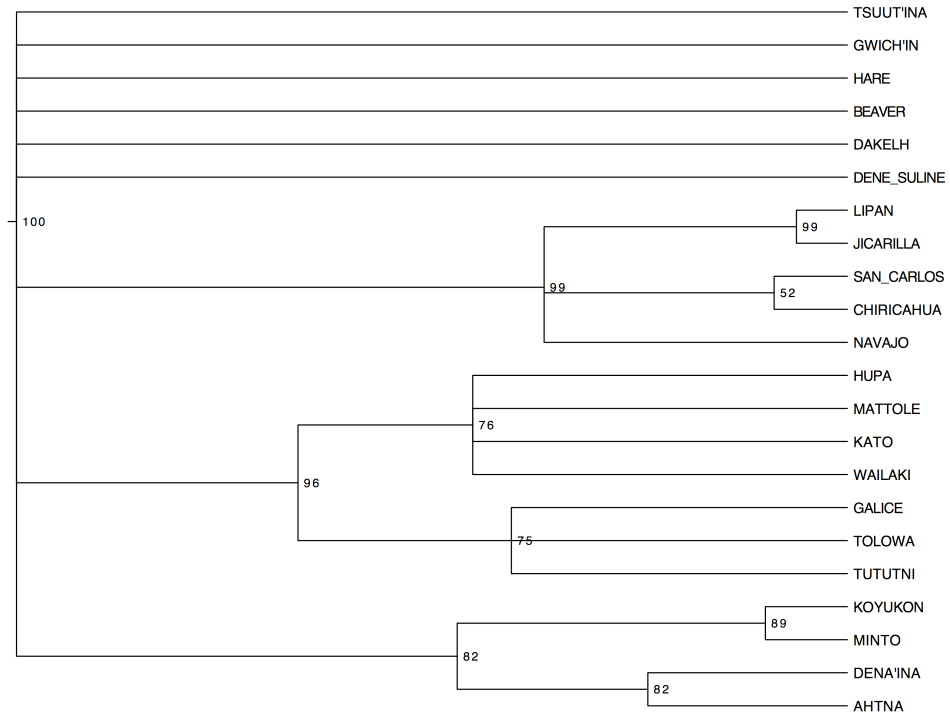



Figure 2.4: Majority-rule consensus tree for non-lexical characters (relaxed clock evolutionary model)

As with the lexical dataset, the non-lexical characters produce a Pacific Coast Athabaskan subgroup with strong support (96%). There are important differences in the topologies in Figure 2.1 and 2.4, however. For example, while Northern Athabaskan overall has the same rake-like branching as found in Figure 2.1, there is the addition of a well-supported clade combining the Alaskan Athabaskan languages (Koyukon, Minto, Dena'ina, and Ahtna). This suggests that these languages have diverged from one another lexically but are much more similar to one another phonologically and morphologically.

More germane to the present discussion are differences found within PCA. Here, the level of support for California Athabaskan is roughly the same as found with lexical characters alone (76% vs. 72% in Figure 2.3, generated with the same relaxed clock evolutionary model). However, there is much weaker support for the Oregon Athabaskan subgroup – still supported at 75%, but below the 80% threshold to be considered well-supported. Despite the caveats surrounding analysis of the non-lexical characters in the computational models invoked here,

[22] As with the lexical dataset, the relaxed clock model performed better than the strict clock model, but here only marginally so: BF $>$ 2. Analyses were run as multi-state characters with all polymorphisms included, but note that all but one of the non-lexical characters were coded as binary to begin with, so the differences between results obtained with the binary and multi-state codings are expected to be minimal.

these results raise an interesting question about why the two PCA subgroups have slightly different profiles: if the degree of support for each subgroup can be taken as inversely proportional to its internal diversity, the degree of lexical diversity within California Athabaskan is roughly the same as its non-lexical diversity; in Oregon Athabaskan, there is low lexical diversity but high non-lexical diversity.

This situation is reminiscent of theories of language contact such as Thomason and Kaufman (1988), who distinguish between "maintenance" versus "shift" profiles in language contact. The maintenance profile would be expected to lead to a high degree of lexical diversity across lineages relative to non-lexical diversity since lexical borrowing is the first and most common outcome of contact where language shift does not result. Language shift typically results first and foremost in phonological and syntactic substrate effects (Thomason and Kaufman 1988:39), leading one to expect a higher degree of non-lexical diversity relative to lexical diversity. Interestingly, however, the lexical diversity in California Athabaskan is not due to borrowing, but rather to semantic shift and neologism. This raises the possibility of two language shift scenarios, one with substrate effects concentrated in non-lexical domains and another where they are concentrated in the lexicon (and possibly in non-lexical domains as well). At present too little is known about the prehistoric language contact situations in California versus Oregon, especially in the crucial first few centuries upon the arrival of Athabaskans in the region, to develop this point further (but see Jacobs 1937; Golla 2000, 2011:68-81, 257-258; Conathan 2004:173-180). This finding does, however, suggest interesting directions for future research that only come to light by virtue of considering both non-lexical and lexical characters.

## 5.3. Results from Combined Dataset

Results were also obtained from a dataset that combined the lexical and non-lexical characters. Figure 2.5 shows the results for a multi-state coding with all polymorphisms retained, run with a non-clock model:[23]

[23] Tlingit and Eyak were included in the analysis of the combined dataset despite lacking data for the non-lexical characters (i.e., roughly 1/3 of the total characters).

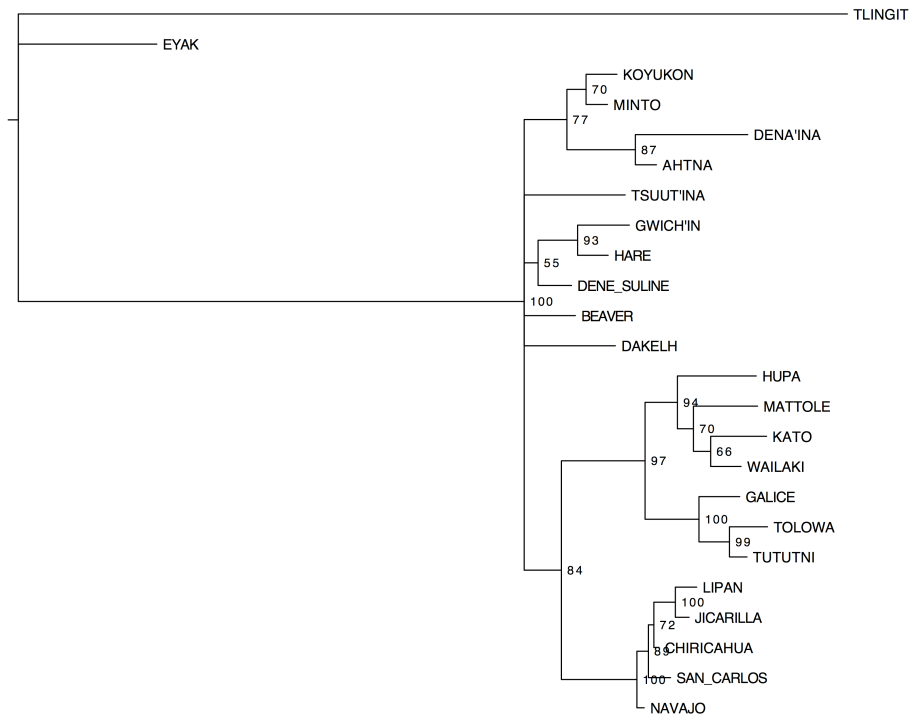



Figure 2.5: Majority-rule consensus tree for the combined lexical and non-lexical datasets coded as multi-state characters with all polymorphisms included (non-clock evolutionary model)

Once again, the PCA languages emerge as a well-supported subgroup (97%). Oregon Athabaskan receives strong support (100%), as does California Athabaskan (94%). Thus, while California Athabaskan is only weakly supported when lexical or non-lexical characters are considered separately, and Oregon Athabaskan when non-lexical characters are considered separately, combining the datasets has the effect of producing strong support for both of them. The Oregon Athabaskan result is perhaps explainable in terms of the fact that the lexical characters outnumber the non-lexical characters roughly 2:1 in the combined dataset, a margin perhaps sufficient to maintain a high level of support.[24] The result for California Athabaskan, however, is especially puzzling since the weak support for this subgroup obtained with the separate datasets does not lead to an expectation a priori that a higher level of support would be returned when they are combined. This oddity is merely noted here, and developing an explanation is left for future research.[25]

[24] By the same token, Figure 2.5 has the same Alaskan Athabaskan subgroup as obtained using non-lexical characters alone in Figure 2.4, but with lower support – unsurprising considering that the lexical characters, which do not support that subgroup when considered on their own, outnumber the non-lexical characters in the combined dataset.

[25] Figure 2.5 has another surprising difference in the topology compared with Figure 2.1 and Figure 2.4: PCA and Apachean form a subgroup within the range considered well-supported (84%). Interestingly, this subgroup disappears when a relaxed clock model of evolution is used, and Apachean is coordinate with PCA and Northern Athabaskan languages as in Figure 2.1. More interesting still, when a strict clock model is used the Apachean languages combine with the Northern languages to form a subgroup with strong (86%) support. That is, the more

## 5.4. Analysis of Characters

The Bayesian inference methods used in this study are not well-suited to discovering specific characters with the biggest effect on the topology and posterior probabilities returned in an analysis. However, a rough sense of the most important ones can be obtained using distance-based clustering methods such as NeighborNet in the program SplitsTree (Huson and Bryant 2006). Although such methods do not provide an explicit model of language evolution (Nichols and Warnow 2008:763-764), they can identify particular characters supporting partitions of the data. The combined lexical and non-lexical dataset with polymorphisms filtered (set to their default states, if available, and recoded as missing otherwise) was analyzed with NeighborNet, which produced the following graph:

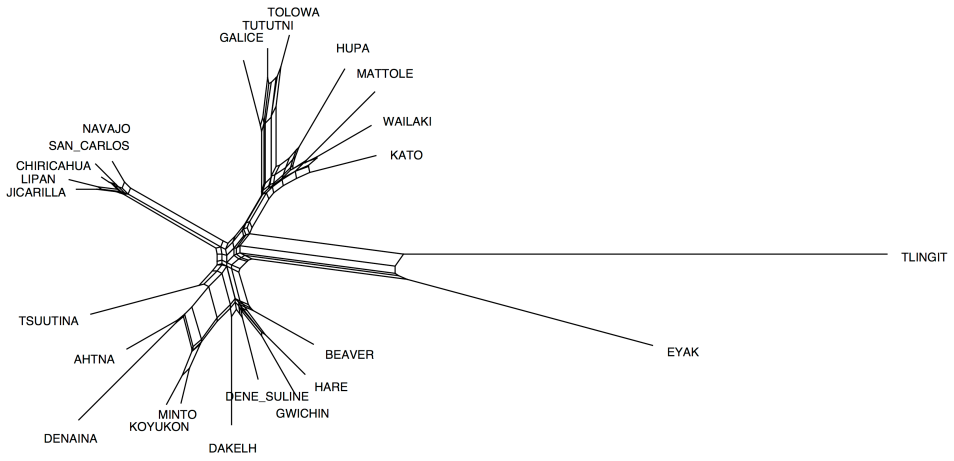



Figure 2.6: NeighborNet graph for the combined lexical and non-lexical datasets coded as multi-state characters with polymorphisms excluded

As in Figure 2.1 and the other topologies produced with Bayesian models, a Pacific Coast Athabaskan subgroup can be detected in the graph in Figure 2.6.[26] The Northern Athabaskan languages, by contrast, show a high degree of internal diversity with many conflicting cross-classifications, as indicated by the thick web-like structure near the center of the graph; this can be understood as analogous to the unresolved rake-like branching in Figure 2.1.

The data partition in Figure 2.6 separating the PCA languages from the rest of the family is supported by four characters, two lexical and two phonological. These are characters that have the same state for all of the PCA languages and a different state for all other languages. The

---

stringent the assumptions about the uniformity of rates of evolution across lineages, the more Apachean gravitates towards the Northern languages and away from PCA. Exploring the reasons for this difference in topology under different assumptions about rates of evolution is beyond the scope of the present study, but also indicates fruitful directions for further research.

[26] The PCA cluster in Figure 2.6, moreover, receives strong support (97%) under bootstrap resampling analysis. This is a technique whereby the NeighborNet algorithm is run on modified datasets with some characters replaced by copies of others. This is repeated for several iterations (5000 in the present case), yielding confidence scores indicating the level of support for various partitions in the network. Scores above 95% are generally taken as strong support for a partition.

phonological characters are the merger of *dl, *tɬ with *l, *ɬ (#3 in Appendix A) and the merger of *q and *x̣ (#6 in Appendix A).[27] The lexical characters are the meanings 'cloud' and 'mouth'. The former involves a root found as *ʔah* in California and *ʔak* in Oregon, from Proto-Athabaskan *ʔa̱:qʾ reconstructed with the meaning 'fog' (vs. *qʾʊs 'cloud'); the semantic shift of this root to 'cloud' is apparently an innovation in PCA.[28] The latter, however, involves a root *-daʔ* 'mouth', which is reconstructible to Proto-Athabaskan in that meaning (Krauss 2005:129); many non-PCA languages have a compound meaning 'lips' that is built on the same root. This is most likely a shared retention in PCA, and while it is not inconsistent with the partition shown in Figure 2.6, it is not a clade-defining innovation in the family.

Support for a subgroup from exactly three characters (or four if 'mouth' is included) might not seem like much, especially considering that none of them are morphological: lexical innovations can easily diffuse across pre-existing linguistic boundaries, and Babel et al. (2013) argue that phonological innovations can do so as well. However, this post hoc analysis of the characters is only part of the story. Re-running the combined lexical and non-lexical dataset with these four characters removed produced same topology as found in Figure 2.5:

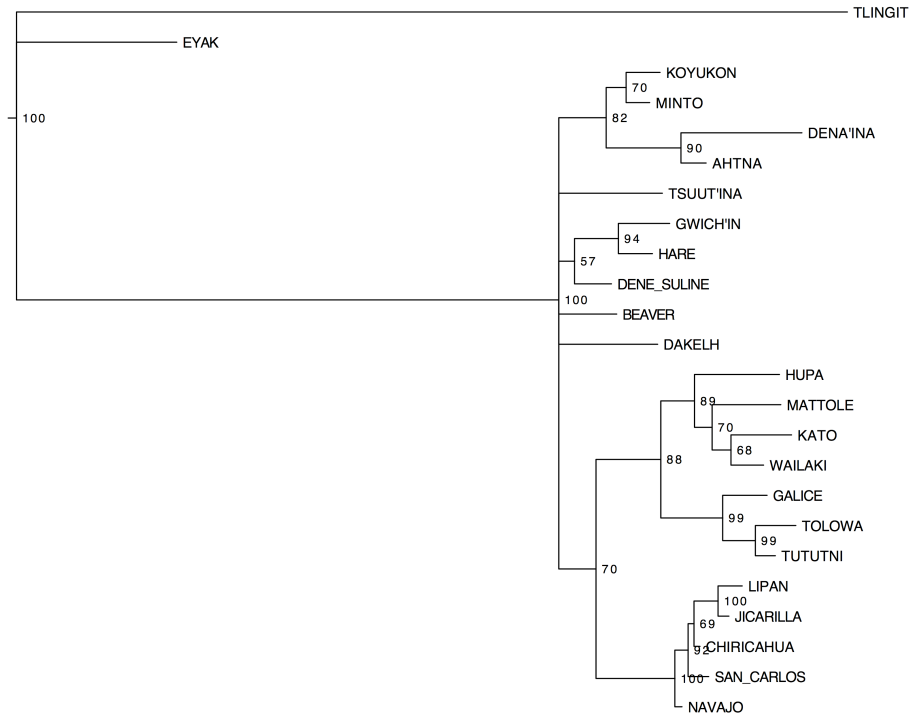



Figure 2.7: Majority-rule consensus tree for the combined lexical and non-lexical datasets excluding characters advantageous for a PCA clade and coded as multi-state characters with all polymorphisms included (non-clock evolutionary model)

[27] A third phonological character expected to lend support to the group based on Hoijer (1960), the merger of *s and *z (#10 in Appendix A), also supports the PCA partition when Hare is recoded (since the merger is apparently a very recent innovation in that language).

[28] The root is often found with an augmentative suffix in the Oregon languages, but the bare root is also attested, e.g. in Harrington's (1981) documentation of Upper Coquille, Chasta Costa, and Tututni (reel 25, frame 1126).

Posterior probabilities for some nodes was lower, but PCA still had 88% support, above the threshold to be considered well-supported. Thus, while further analysis is needed to get a more precise sense of which characters are contributing the most to the results, the explicit parametric evolutionary model in the Bayesian methodology still detects a PCA subgroup even when the characters that are most advantageous to it are removed from consideration.[29]

## 6. Discussion

The most pertinent result for the narrow goals of this chapter is that the Pacific Coast Athabaskan languages emerge as a well-supported subgroup of Athabaskan in all conditions: with lexical or non-lexical datasets, or a combination of the two; with non-clock, relaxed clock, or strict clock models; with polymorphic characters included or excluded under a multi-state coding; with unique character states included or excluded in a binary recoding. All of this can be taken as evidence in favor of a PCA subgroup within Athabaskan, as argued by Hoijer (1960). Contrary to Hoijer (1962), however, the Bayesian inference models used here suggest that the PCA clade is supported on both lexical and non-lexical grounds. The methods employed here can therefore be taken as resolving to some extent what had been taken as conflicting evidence in lexical and non-lexical linguistic domains. Hoijer (1962) is partially vindicated, however, insofar as support for a California Athabaskan subgroup within PCA is relatively weak when lexical and non-lexical characters are considered separately; surprisingly, it is stronger when the two character sets are combined. The views of Krauss (1981) and Krauss and Golla (1981) with respect to the Northern Athabaskan languages especially are also largely supported insofar as the results across conditions are quite varied, and with some data and model configurations very little branching structure emerges from the analysis. However, the well-supported subgroup subsuming the Alaskan Athabaskan languages in Figure 2.4 shows that a moderate amount of branching structure can be detected when non-lexical characters are considered.

The question of subgrouping the PCA languages is orthogonally related to the migration of Athabaskan-speaking groups into California and Oregon. The results reported here do not speak to the locus of differentiation of PCA, whether it was in situ after a single Athabaskan-speaking population migrated into the California-Oregon region, or in some other place prior to migration. Golla's (2011) arguments concerning a lack of shared lexical innovations indicative of a common migration into a new biological and cultural environment suggest that the latter scenario is correct. This is surprising, since a priori a theory of differentiation in situ is preferable: it requires positing only a single migration event and provides the simplest explanation for the concentration of a closely-related subset of the Athabaskan languages in a locale far removed from the rest of the family. Multiple migrations subsequent to the diversification of PCA is not a historical impossibility, however; the interesting question then becomes whether

[29] The lack of morphological support for a PCA clade may be a reflection of the fact that morphological characters were selected largely on the basis of the availability of data across the Athabaskan family rather than on their inherent usefulness for identifying subgroups. While some morphological properties have been identified that are found only in PCA languages (e.g., the ordering of object agreement to the right of subject agreement in Hupa and Kato discussed by Rice 2000:236-244), it is unclear at present how pervasive they are within PCA.

there are principled reasons for both divisions of the subgroup winding up in the same region.[30]

A second finding of interest in this chapter is the relevance of non-lexical information in establishing linguistic phylogenies using computational methods. Although the models invoked here may well be inappropriate for non-lexical datasets, it is nonetheless interesting that differences in tree topologies were found when the non-lexical characters were included, confirming the findings of Nakhleh et al. (2005) for Indo-European. Including non-lexical characters did not alter the basic result with respect to PCA, but a moderate amount of well-supported structure in the Northern Athabaskan languages was detected. This highlights the urgent need to develop computational methods that are more realistic models of phonological and morphological change in order to account for potential differences that result when more than just lexical data are considered.

A final result worth noting is the differences in topologies obtained under multi-state versus binary codings of the data. Contrary to the expectations of Pagel and Meade (2006), the topology obtained under a binary recoding is not simply a scaled version of the one obtained in the corresponding multi-state coding. It is suggested here that this may be due to a disproportionate influence of characters with large numbers of states compared versus characters with a smaller number of states. A formalization of this argument is left to future work.

## Appendix A: Phonological Characters

This appendix provides a summary of the phonological characters that were coded in this study. Here and in Appendix B polymorphic codings are represented in curly braces and default states are indicated with an asterisk. Data coded as missing – either because information truly was not available, or because a character isn't relevant for a given language due to dependencies discussed in §4.3.2 – are represented with '?'.

The first two phonological characters relate to the status of lexical tone:

1. Presence vs. absence of lexical tone. Proto-Athabaskan vowel constriction develops into lexical tone in many Athabaskan languages (Krauss 2005, Kingston 2005). As noted in §2.1 (item (a)), tone does not develop in the Pacific Coast Athabaskan languages. It is also absent in some of the Northern Athabaskan languages (Koyukon, Dena'ina, Ahtna), although there is evidence in one dialect of Koyukon of a vestigial tone system (Krauss and Golla 1981). Dakelh (Carrier) is treated as toneless in Krauss and Golla (1981) and is coded as such here; however, Krauss (2005:69,117) reclassifies Dakelh as having tone and a recoding is probably warranted. Languages lacking lexical tone were coded with state 0, while those with lexical tone were coded with state 1; Koyukon was coded as polymorphic.

[30] Golla's (2011) suggestion of a "partially shared" history of migration that occurred in two or more "pulses" may provide the key to understanding the dynamics of Athabaskan migration from the north. In particular, it would be expected that a migration over such a long distance wouldn't have been completed all at once, but rather would have involved intermediate migratory stages, with some groups lagging behind others. This would imply the creation of what is essentially a mobile dialect network across which some early innovations might have spread.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | 0 | Koyukon | {0,1} | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 1 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

2. High tone vs. low tone. For languages in which tone develops, Proto-Athabaskan vowel construction is realized variably as high tone or low tone, so-called high-marking vs. low-marking systems. Low-marking languages were coded with state 0 and high-marking languages with state 1. Languages lacking lexical tone (coded with state 0 for #1) were coded as missing. Beaver is reported as being high-marking in some dialects and low-marking in others (Krauss 2005:111) and was coded as polymorphic for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | ? | Ahtna | ? | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | ? | Beaver | {0,1} | Hare | 1 | Lipan | 0 |
| Kato | ? | Dakelh | ? | Koyukon | 0 | Jicarilla | 0 |
| Mattole | ? | Dena'ina | ? | Minto | 0 | Navajo | 0 |
| Tolowa | ? | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | ? | | | | | | |
| Wailaki | ? | | | | | | |

Many of the sound correspondences and changes from Proto-Athabaskan discussed in the literature affect consonants in stem-initial position:[31]

3. Merger of *dl, *tɬ with *l, *ɬ As noted in §2 (item (i)), this change is characteristic of all the PCA languages (coded with state 1) but is not found in the rest of Athabaskan (coded with state 0). Hare retains stem-initial *dl distinct from *l, but *tɬ merges with *ɬ. It was coded as polymorphic for this character but with non-merger as a default, since the merger of *tɬ and *ɬ is part of a more general merger of stem-initial aspirated affricates and voiceless fricatives that appears to be a local innovation in this language (cf. Rice 1989:101, fn. 3).

[31] Many changes in place of articulation affect entire series (plain, aspirated, and glottalized), but following Athabaskanist convention just one member of each series is represented. Changes affecting consonants more generally often have different outcomes in prefixes vs. stem-initial position (e.g., Li 1930a:35-39 for Mattole, and Hoijer 1938:79-81 for Apachean), but these are rarely considered as potential sources of evidence for establishing historical relationships.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 1 | Beaver | ? | Hare | {0*,1} | Lipan | 0 |
| Kato | 1 | Dakelh | ? | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 1 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

4. *t > k. This sound change is found in Lipan and Jicarilla Apache (Hoijer 1938), coded with state 1; other Athabaskan languages were coded with state 0.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 1 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

5. *tɬ' > t'. This sound change is found in Wailaki and Tolowa (state 1) but nowhere else in Athabaskan (coded with state 0); Golla (2011) suggests that this may be a substrate effect due to contact with the adjacent Algic languages Yurok and Wiyot.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 1 | | | | | | |

6. Merger of *q and *x̣. As noted in §2.1 (item (b)), this merger is characteristic of the PCA languages, which were coded as state 1; other Athabaskan languages were coded as state 0.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 1 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | 1 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 1 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

7. Merger of *q and *x̣ as /k/ or /x/. Some PCA languages merge *q and *x̣ as /x/ (state 0), others as /k/ (state 1). Other Athabaskan languages were coded as missing. As discussed in detail in chapter 3, these obstruents merge in Hupa proper as /x/ but as /k/ in variation with /x/ in the Whilkut dialect of Hupa, so Hupa was coded as polymorphic for this character. Although there is evidence in Tututni of variation between /k/ and /x/ in early sources (discussed in chapter 4), the details are murky and this is not part of the established understanding of the language, so Tututni was not coded as polymorphic for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | ? | Gwich'in | ? | Chiricahua | ? |
| Hupa | {0,1} | Beaver | ? | Hare | ? | Lipan | ? |
| Kato | 1 | Dakelh | ? | Koyukon | ? | Jicarilla | ? |
| Mattole | 1 | Dena'ina | ? | Minto | ? | Navajo | ? |
| Tolowa | 0 | Dëne Sųłiné | ? | Tsuut'ina | ? | San Carlos | ? |
| Tututni | 0 | | | | | | |
| Wailaki | 1 | | | | | | |

8. Fronting of affricates and fricatives *s, *ts, etc. Many languages realize Proto-Athabaskan *s, *ts etc. as interdental /θ/, /tθ/, etc. (state 1); other languages retain a plain dental or alveolar articulation. Koyukon merges plain dentals with laterals (e.g., *s > ɬ), but according to Krauss (1973:944) fronting was "no doubt an intermediate stage" in this development, so Koyukon was therefore coded as state 1; Krauss similarly suggests that fronting was a stage in the development *s > f in Hare. As discussed in chapter 4, Tututni is reported as having undergone this change to some extent (e.g., Golla 1976, Sapir 1914) and was coded as polymorphic.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 1 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | 1 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | {0,1} | | | | | | |
| Wailaki | 0 | | | | | | |

9. Merger of *s with *ts. As noted in §2.1 (item (f)), this merger is characteristic of the Oregon Athabaskan languages (state 1). It is considered by Hoijer (1960) to have applied partially in California Athabaskan languages other than Hupa. Kato was coded as polymorphic for this character (states 0 and 1). Mattole was coded as polymorphic as well, but with non-merger set as the default state based on the discussion in Li (1930a), who considers the merger in Mattole to be partial at best. Li (1930a) treats [ts] in Wailaki as due to phonologically predictable epenthesis of [t] before /s/, so was coded as state 0. Other Athabaskan languages, which do not undergo the merger, were coded as state 0.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | {0,1} | Dakelh | 0 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | {0*,1} | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

10. Merger of *s and *z. As discussed in §2.1 (item (e)), this merger is found throughout PCA (state 1) but not elsewhere in Athabaskan (state 0), with the exception of Hare, where *s and *z merge as /w/ (Rice 1989). Hare was coded as state 1 for this character, but according to Krauss and Golla (1981:79) this change occurred only in the late 19th century and hence is surely unrelated to the change in PCA, so a recoding as state 0 is probably warranted (also if the character is interpreted as merger of *s and *z resulting specifically in /s/).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 1 | Beaver | 0 | Hare | 1 | Lipan | 0 |
| Kato | 1 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 1 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

11. Fronting of *tš. Many Athabaskan languages front *tš to /ts/ (state 1); languages that do not were coded with state 0. The Oregon Athabaskan languages, most of which modify the Proto-Athabaskan palatal affricates to a retroflex fricative /ṣ/, were coded as state 0 since the modification does not lead to merger with *ts (whose reflex is /s/ in Oregon Athabaskan). Hoijer (1960, 1963) reports both /š/ and /s/ as reflexes of *tš in Galice. However, this may be a reflection of the same plain vs. retroflex distinction, which is extremely subtle everywhere in Oregon Athabaskan except Tolowa (cf. Golla 1976). Galice was coded as polymorphic for this character but with non-fronting (state 0) set as the default. According to Krauss and Golla (1981:72), Dena'ina retains *tš as such only in some dialects and among older speakers; it was coded as state 1, but a polymorphic coding is probably warranted

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | {0*,1} | Ahtna | 1 | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 1 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

12. Palatalization of *k̂. Unsurprisingly, many languages have undergone palatalization of front-velar stops, which are realized as /tš/ etc. (state 1); other languages were coded as state 0. Dialect variation is reported in Koyukon, Ahtna, and Wailaki, which were coded as polymorphic for this character.[32] In Koyukon and Ahtna, the variation is restricted to single dialects, so state 0 in the majority dialects lacking palatalization was designated as the default. The Apachean languages have /ts/ as a reflex of *k̂. They were coded as state 0 for this character since the change *k̂ > ts was achieved without palatalization as an intermediate stage (since these languages all retain *tš as such).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | {0*,1} | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 1 | Koyukon | {0*,1} | Jicarilla | 0 |
| Mattole | 1 | Dena'ina | 0 | Minto | 1 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | {0,1} | | | | | | |

13. Palatalization of *x̂. Many languages palatalize *x̂ to /š/ (state 1). This is treated separately from palatalization of *k̂ because palatalization of *x̂ precedes merger of palatal obstruents with the alveolar/dental series (cf. #11 above), so in many languages *x̂ is found as /s/. This is also the case in Apachean, but as with merger of *k̂ with *ts the change *x̂ > s was achieved without intermediate palatalization (since *š is retained as such), so these languages were coded with state 0. Hupa has /ʍ/ as the reflex of both *š and *x̂, so presumably palatalization merging *x̂ with *š preceded the shift *š > ʍ (state 1). Li (1930a:13, 27) reports for Mattole that *š is found as /x̂/ everywhere but in the 1sg pronoun and agreement prefixes. Although Li had no clear examples of stem-initial *x̂ in Mattole, he maintains that *x̂ > x̂ stem-finally. Together with stem-initial *š > x̂, this can be taken as circumstantial evidence that palatalization of *x̂ would not be the expected outcome, so Mattole was coded as state 0 for this change. Relevant examples were not located in Goddard (1909, 1912) for Kato, which was coded as missing. Tsuut'ina (Sarcee) is reported by Hoijer (1963) as having both /x/ and /š/ as a reflex of *x̂ and was coded as polymorphic for this character.

[32] Goddard's description of Kato suggests that a polymorphic coding might be warranted: "In many cases it is rather difficult to be sure whether tc [i.e., tš] is uttered or a prepalatal k with, perhaps, a glide" (1912:11).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 1 | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 1 | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | ? | Dakelh | 1 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | {0,1} | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

14. Merger of *k̲, *x̲̂ with *ts, *s without palatalization. This character picks out the Apachean development of the front-velar obstruents (cf. #12 and #13 above).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 1 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 1 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

15. Denasalization of *n before oral vowels. According to Krauss and Leer (1981), many Athabaskan languages denasalize Proto-Athabaskan nasal consonants before oral vowels, either entirely or as prenasalized /$^{n}$d/. This character reflects retention of nasality (state 0), or complete or partial denasalization (state 1). Prenasalized /$^{n}$d/ in Apachean languages other than Navajo was coded as state 1 even though this may be a recent and independent development in these languages.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

16. Nasalization of *y before nasal vowels. Krauss and Leer (1981:77) report that *y is found as /n/ in several Northern Athabaskan and some Apachean languages (state 1); languages retaining /y/ were coded as state 0. *y is also apparently sometimes realized as /m/ in Hupa and Galice, which were coded as polymorphic for this character with the nasal /m/ treated as equivalent to /n/ (state 1). Navajo is also reported as having variable outcomes and was coded as polymorphic.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | {0,1} | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | {0,1} | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 0 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | 1 | Navajo | {0,1} |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

17. *w realized as non-nasal /b/ or /v/ (state 0) or nasal /m/ (state 1) before oral vowels. Both states occur in the Apachean languages (Hoijer 1938) and Koyukon (/m/ generally in the Lower dialect, before a vowel followed by /n/ in other dialects) (Krauss and Golla 1981:74), which were coded as polymorphic for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | {0,1} |
| Hupa | 1 | Beaver | 0 | Hare | 0 | Lipan | {0,1} |
| Kato | 0 | Dakelh | 0 | Koyukon | {0,1} | Jicarilla | {0,1} |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | {0,1} |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 1 | San Carlos | {0,1} |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

The next several characters relate to Proto-Athabaskan stem vowels:

18. Raising of *a. Some languages raise Proto-Athabaskan *a to /o/ or /u/ (state 1); languages retaining /a/ as such were coded as state 0. Gwich'in, which has /i/ and /e/ as reflexes of *a (Krauss and Golla 1981:72), was coded as missing (since it neither retains *a as such nor raises it to /o/ or /u/).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | 0 | Dakelh | 0 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

19. Lowering of *u to /o/. Languages where this lowering applies were coded as state 1, others were coded as state 0. Dakelh (Carrier) is reported to have sometimes /u/, sometimes /o/ (Krauss and Golla 1981) and was coded as polymorphic for this character. Gwich'in has /io/, partially retaining the original height and partially lowered, and was coded as missing since this intermediate value could be interpreted either way.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 1 | Beaver | 0 | Hare | 0 | Lipan | 1 |
| Kato | 1 | Dakelh | {0,1} | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 1 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

20. All Proto-Athabaskan reduced vowels merge as /ə/. This change is found in Oregon Athabaskan and Dena'ina (state 1), but not in other languages considered in this study (state 0).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

21. *ə > i. This change is found in Hupa, Mattole, and the Apachean languages (state 1). It is not found in Kato. Wailaki has many words where Li (n.d.) transcribed [i] but Goddard (1923) transcribed <α> (i.e., [ə]) and was coded as polymorphic for this character. The Oregon Athabaskan languages and Dena'ina were coded as missing since they underwent the merger of all Proto-Athabaskan reduced vowels to /ə/ (state 1 for #20 entails state 0 for this character).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | ? | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 1 |
| Hupa | 1 | Beaver | 0 | Hare | 0 | Lipan | 1 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 1 | Dena'ina | ? | Minto | 0 | Navajo | 1 |
| Tolowa | ? | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | ? | | | | | | |
| Wailaki | {0,1} | | | | | | |

22. Proto-Athabaskan high back reduced vowel *ʊ realized as /u/ (state 0) vs. /o/ (state 1). Oregon Athabaskan languages and Dena'ina, which merge all reduced vowels as /ə/, were coded as missing.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | ? | Ahtna | 1 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | 1 | Beaver | 0 | Hare | 1 | Lipan | 1 |
| Kato | 1 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 1 | Dena'ina | ? | Minto | 0 | Navajo | 1 |
| Tolowa | ? | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | ? | | | | | | |
| Wailaki | 1 | | | | | | |

The next several characters relate to stem-final consonants, which have undergone various kinds of erosion in many Athabaskan languages:

23. Deletion of coda nasals. Many Athabaskan languages delete coda nasal segments (state 1), while others retain them (state 0). Tsuut'ina (Sarcee) retains coda nasals underlyingly but deletes them when they close a syllable with a short/reduced vowel (Li 1930b); it was coded as polymorphic for this character but with retention as the default.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | 1 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | {0*,1} | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

24. Result of coda nasal deletion. For languages that delete coda nasals, nasalization is sometimes retained on the preceding vowels (state 0), in other cases it is lost entirely (state 1). Languages retaining coda nasal segments (including Tsuut'ina) were coded as missing for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | ? | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | ? | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | ? | Dakelh | 1 | Koyukon | ? | Jicarilla | 0 |
| Mattole | ? | Dena'ina | ? | Minto | ? | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | ? | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | ? | | | | | | |

25. Loss of stem-final affrication. Many languages realize Proto-Athabaskan stem-final affricates as fricatives (state 1), often retaining place of articulation, but sometimes debuccalizing to /h/ in Northern languages where fricatives are also lost stem-finally. Languages retaining the affrication were coded as state 0. Wailaki and Mattole were coded as polymorphic for this character since they retain affrication for most consonants but show reduction of the glottalized lateral affricate *tɬ' to /ʔl/ stem-finally; since this affects only one of

several candidate affricates, retention (state 0) was set as the default for the polymorphism.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | 0 | Beaver | 0 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | 1 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | {0*,1} | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | {0*,1} | | | | | | |

26. Loss of final glottalization. Another instantiation of the erosion of stem-final obstruents is the loss (state 1) vs. retention (state 0) of stem-final glottalization. This is justified as a separate character because some languages retain affrication but lose glottalization, while others retain glottalization but lose affrication. Glottalization can be considered retained even if it is realized as pre-glottalization (e.g., /ʔs/ from *ts' as in Oregon Athabaskan). Kato wass coded as missing since the presence or absence of glottalization cannot be inferred from Goddard (1909, 1912), who was a notoriously poor transcriber of glottalization.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | ? | Dakelh | 1 | Koyukon | 1 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 1 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

27. Reduction or loss of final fricatives. This is found in many Northern Athabaskan languages (Krauss and Golla 1981), but only in Hare and Gwich'in (Kutchin) in the languages included in this study (coded as state 1).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

The last four phonological characters relate to morphophonological alternations in verb stems in the environment of Proto-Athabaskan prefixes with *d (a detransitivizing voice-valence "classifier" prefix and the first person plural subject prefix). Some stem-initial consonants

undergo modifications in this environment, the so-called Athabaskan "d-effect."[33] The following characters summarize four of the consonants that are affected. The d-effect is found with all four of these consonants in Dakelh (Carrier), Dëne Sųłiné (Chipewyan), and the Apachean languages; with some exceptions to be noted, it is found with none of them in PCA and the Alaskan languages. Data for Gwich'in are missing in all cases. Other consonants are affected but can be predicted based on whether or not one of these four triggers the d-effect and therefore are not justified as separate characters. Notice that these characters ignore the actual realization of stem-initial consonants, which may be subject to other phonological changes, but instead merely register the presence (state 1) or absence (state 0) of the d-effect.

28. D-effect /d/ + /ʔ/. This is found in most languages with the d-effect, where it is realized as /t'/. The d-effect with stem-initial /ʔ/ is found in Minto and Koyukon and is otherwise unattested in the four Alaskan languages included in this study (including Ahtna and Dena'ina).[34] It is found in all of Apachean and in the Canadian languages: Dakelh (Carrier), Dëne Sųłiné (Chipewyan), Hare, Beaver, and Tsuut'ina (Sarcee). Transcriptions of Kato (Goddard 1909) and Wailaki (Goddard 1923, Seaburg 1977) show that the d-effect occurs with /ʔ/-initial stems in these languages as well, but it is not found elsewhere in PCA.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 1 | Dakelh | 1 | Koyukon | 1 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 1 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 1 | | | | | | |

29. D-effect /d/ + /ɣ/. This is found in Apachean and the Canadian languages, and is absent from the Alaskan languages and PCA.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | 1 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

30. D-effect /d/ + /n/. This is absent from the Alaskan languages and PCA and present in

[33] The importance of the d-effect for understanding historical relationships within Athabaskan was suggested by Howren (1971), who noted that "the phenomenon has its roots in a period when the Northern and Apachean languages constituted in some sense a single linguistic community."

[34] In Koyukon it is triggered by a preceding /d/ even when it is separated from the stem by the vowel (Wisniewski n.d.); this is exceptional within the family.

all of Apachean. It is found in Dakelh (Carrier) and Dëne Sųłiné but not in the other Canadian languages.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 1 |
| Kato | 0 | Dakelh | 1 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

31. D-effect /d/ + /y/. This is present in Apachean, in all of the Canadian languages except Hare, and in Tututni. It is apparently absent in other PCA languages and the Alaskan languages.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 0 | Lipan | 1 |
| Kato | 0 | Dakelh | 1 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

## Appendix B: Morphological Characters

As noted in §4.3.4, morphological characters were selected more on the basis of the availability of data in a large number of languages rather than because they are necessarily the best characters for phylogenetic inference. The similarities and differences they encode are real, however, and can therefore be considered a reasonable first pass at developing a robust set of morphological characters as input for computational studies in the family. Some languages, especially Gwich'in, do not have published grammatical descriptions that are robust enough to allow values for many of these characters to be inferred.

1. Expression of first person plural subject agreement. Most Athabaskan languages retain a reflex of the Proto-Athabaskan prefix (state 0) that includes a /d/ element in a position close to the verb stem (the same one implicated in the d-effect, reconstructed variably as *əŋəʔdəD, *əŋə by Story 1989). Several Northern Athabaskan languages, however, have innovated by replacing the Proto-Athabaskan element with a historically third person indefinite prefix in a position farther away from the stem (state 1). Beaver was coded as polymorphic for this character, with the innovative prefix set as the default state: Goddard (1917:474) reported "a few" cases of a prefix that might be cognate with the old 1pl prefix, and Story (1989:513) says that the innovative prefix is the most common one in the Doig River dialect but the /d/ prefix is retained in hortatives.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 1 | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 0 | Beaver | {0,1*} | Hare | 0 | Lipan | 0 |
| Kato | 0 | Dakelh | 0 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

2. Retention versus loss of Proto-Athabaksan negative paradigm. This paradigm is retained in some Northern languages (state 0) (Krauss and Golla 1981:71, Leer 2000) and lost elsewhere in the family (state 1). According to Leer (2000), it is lost in most varieties of Gwich'in (Kutchin) but retained in what he calls a "Biblical" variety (presumably one used in Bible translations, perhaps from the late 19th or early 20th century); Gwich'in is coded as polymorphic for this character but with loss of the category coded as the default state. A potential problem with this character is that it assigns languages with a shared retention (the Proto-Athabaskan negative paradigm) to one state, and all other languages to another state. However, it is far from clear that all of the languages that lost the verb-internal negative innovated the same means of expressing negation, and a more nuanced coding is probably needed.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | {0,1*} | Chiricahua | 1 |
| Hupa | 1 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 1 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 1 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

3. Syncretism of first person plural and second person plural object prefixes. Some languages retain distinct 1pl and 2pl object prefixes (state 0), but in many languages the two have been collapsed (state 1).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 1 | Beaver | 0 | Hare | 1 | Lipan | 1 |
| Kato | 1 | Dakelh | 0 | Koyukon | 0 | Jicarilla | 1 |
| Mattole | 1 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | 1 | | | | | | |

4. Form of disjunct plural prefix. The Athabaskan verb template is divided into so-called conjunct and disjunct zones, prefixes in the former being closer to the verb stem than those in the latter. All of the languages included in this study have a plural prefix in the disjunct

zone that starts sometimes with /n/ (state 0), sometimes with /d/ (state 1), still other times with /y/ (state 2). Which one occurs is often correlated with distinct semantics – the n-/d-initial forms tend to be distributives, and the y-initial forms totalatives (especially in PCA) (Rice 2000). Hare, however, has a y-initial distributive. Rice (2000:376) suggests that these prefixes are not cognate with one another, a view supported by the fact that Koyukon and Carrier have more than one of them, and Beaver all three (Randoja 1990); these languages were coded as polymorphic for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 2 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 2 | Beaver | {0,1,2} | Hare | 2 | Lipan | 1 |
| Kato | 2 | Dakelh | {0,2} | Koyukon | {0,2} | Jicarilla | 1 |
| Mattole | 2 | Dena'ina | 0 | Minto | 0 | Navajo | 1 |
| Tolowa | 2 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 2 | | | | | | |
| Wailaki | 2 | | | | | | |

5. Order of disjunct plural and iterative. The Athabaskan disjunct zone includes an iterative/reversative prefix (*na:-* in Hupa). In most languages, the iterative normally precedes the disjunct plural prefix (state 0), but the opposite order is found in Hare, Dakelh (Carrier), and Dëne Sųłiné (Chipewyan). Relevant examples were not located for some languages (Galice, Gwich'in, Minto).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | ? | Ahtna | 0 | Gwich'in | ? | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 1 | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | ? | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

6. Robust/productive noun incorporation. Many Northern Athabaskan languages have robust noun incorporation (state 1), but elsewhere in the family it is absent or vestigial (state 0). If noun incorporation is reconstructible to Proto-Athabaskan, this character may need a more nuanced coding.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 1 | Gwich'in | ? | Chiricahua | 0 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | 1 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

7. Order of incorporated noun and disjunct plural. In many languages with incorporation, the incorporated noun follows the disjunct plural prefix (state 1). In Dena'ina, Tsuut'ina (Sarcee), and Beaver the opposite order is found. Languages that do not have robust noun incorporation were coded as missing for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | ? | Ahtna | 1 | Gwich'in | ? | Chiricahua | ? |
| Hupa | ? | Beaver | 0 | Hare | 1 | Lipan | ? |
| Kato | ? | Dakelh | 1 | Koyukon | 1 | Jicarilla | ? |
| Mattole | ? | Dena'ina | 0 | Minto | 1 | Navajo | ? |
| Tolowa | ? | Dëne Sųłiné | 1 | Tsuut'ina | 0 | San Carlos | ? |
| Tututni | ? | | | | | | |
| Wailaki | ? | | | | | | |

8. First person singular perfective allomorph **-i:-*. This allomorph is retained in many languages (state 0), but lost in many others (state 1). Hoijer (1971:127) noted its absence in Galice. Golla (1976) reported that it is also missing in Tututni, but only in some verbs, and it is absent from Sapir's (1914) description of Chasta Costa; it appears to occur occasionally in some of the verb paradigms in Bommelyn (1989). Tututni was coded as state 0 for this character, Tolowa as state 1, but it is likely that both share partial retention of this allomorph and should instead be coded uniformly.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 1 | Gwich'in | ? | Chiricahua | 0 |
| Hupa | 0 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | 0 | Dakelh | 1 | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 0 | Tsuut'ina | 1 | San Carlos | 0 |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

9. Third person plural subject prefix. Many Athabaskan languages have a third person plural prefix (dual in Tututni) in the verb template, also found in possessed nouns in some languages (state 1); others do not (state 0). There is likely more than one prefix etymologically: although Hoijer (1971) reconstructs **x̣ə-* in Apachean and several Northern languages, Cook (1996) reconstructs **ɢe-* for the Northern languages and maintains that there is no cognate for this prefix in Apachean. This character can therefore perhaps be taken as presence vs. absence of the morphological category rather than its exponence.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 1 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | 1 | Koyukon | 1 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | 0 | | | | | | |

10. Syncretism of third person plural subject and third person plural object prefixes. In most languages with a third person plural subject prefix, the same prefix can be used with third person plural objects as well (state 1). In a few (Beaver, Hare, Ahtna) it cannot (state 0) and a distinct prefix is used with third person plural objects. In Tututni, the dual *xə-* is "apparently restricted to forms with third-person subjects" (Golla 1976:223), which was coded as state 0; data are lacking in Galice and Tolowa, which were coded as missing. In some languages the third person plural object prefix is a combination of the third person plural subject plus an element *bə-/be-*, which is the usual third person object prefix; such languages were coded as state 1 on the assumption that the plural object prefix is complex and analyzable. Languages lacking a third person plural subject prefix were coded as missing for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | ? | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | ? | Beaver | 0 | Hare | 0 | Lipan | 1 |
| Kato | ? | Dakelh | ? | Koyukon | 1 | Jicarilla | 1 |
| Mattole | ? | Dena'ina | 1 | Minto | 1 | Navajo | ? |
| Tolowa | ? | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | ? | | | | | | |

11. Function of *yə- prefix. In most Athabaskan languages this prefix is used as an obviative marker (third person subject acting on third person object) (state 0). This is true in most of PCA as well, but in Hupa and Kato it marks less animate third person subjects (Golla 2011:82), occurring with first and second person objects and in intransitive contexts. In Tolowa it has been reinterpreted as a transitivity marker (Givón and Bommelyn 2000). All three languages were coded as state 1, but the development in Tolowa is probably independent and a recoding may be warranted.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | 0 | Chiricahua | 0 |
| Hupa | 1 | Beaver | 0 | Hare | 0 | Lipan | 0 |
| Kato | 1 | Dakelh | ? | Koyukon | 0 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | 0 | Navajo | 0 |
| Tolowa | 1 | Dëne Sųłiné | 0 | Tsuut'ina | 0 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

12. Co-occurrence of *yə- prefix and noun phrases in canonical position. This morphosyntactic property of the obviative prefix figures in discussions of the status of Athabaskan verb prefixes as either agreement or pronominal elements. In most of the Northern languages, the prefix does not co-occur with NPs in canonical position (state 1); Denai'ina is the exception (Thompson 1996). Elsewhere it can co-occur with NPs (state 0).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 1 | Gwich'in | 1 | Chiricahua | 0 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 0 |
| Kato | 0 | Dakelh | ? | Koyukon | 1 | Jicarilla | 0 |
| Mattole | 0 | Dena'ina | 0 | Minto | ? | Navajo | 0 |
| Tolowa | 0 | Dëne Sųłiné | 1 | Tsuut'ina | 1 | San Carlos | 0 |
| Tututni | 0 | | | | | | |
| Wailaki | 0 | | | | | | |

The last two characters relate to permissible person reference in the optative, Proto-Athabaskan *-*u*-. In all languages third person reference is possible; in some, first person reference is also permitted, and in still others second person reference is allowed as well (with the optative functioning as an imperative). Tsuut'ina, which has "no trace" of the optative according to Cook (1985:212), was coded as missing.

13. First person reference in the optative. Present (state 1) or absent (state 0).

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 1 | Ahtna | 1 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | 0 | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | 0 | Dakelh | ? | Koyukon | 1 | Jicarilla | 1 |
| Mattole | 0 | Dena'ina | 1 | Minto | 1 | Navajo | 1 |
| Tolowa | ? | Dëne Sųłiné | 1 | Tsuut'ina | ? | San Carlos | 1 |
| Tututni | 1 | | | | | | |
| Wailaki | ? | | | | | | |

14. Second person reference in the optative. Present (state 1) or absent (state 0). Since languages disallowing first person reference with the optative also disallow second person reference, languages coded with state 0 for #13 were coded as missing for this character.

| Pacific Coast | | Northern | | | | Apachean | |
|---|---|---|---|---|---|---|---|
| Galice | 0 | Ahtna | 0 | Gwich'in | ? | Chiricahua | 1 |
| Hupa | ? | Beaver | 1 | Hare | 1 | Lipan | 1 |
| Kato | ? | Dakelh | ? | Koyukon | 1 | Jicarilla | 1 |
| Mattole | ? | Dena'ina | 1 | Minto | 1 | Navajo | 1 |
| Tolowa | ? | Dëne Sųłiné | 1 | Tsuut'ina | ? | San Carlos | 1 |
| Tututni | 0 | | | | | | |
| Wailaki | ? | | | | | | |

## Appendix C: The Status of Upper Umpqua

As outlined in Golla (2011:70-72), the earliest documentation of any Pacific Coast Athabaskan language is found in Upper Umpqua wordlists collected by Tolmie (published in Scouler 1841) and Hale (1846). Subsequent documentation of Upper Umpqua is sparse, however: there are additional 19th century wordlists from Milhau (1856), Barnhardt (1859), and Gatschet (1877), and a modest amount of material collected by Melville and Elizabeth Jacobs (1930s) and by J.P. Harrington in 1940 (1981, reel 19). On the basis of this documentation and reports that Upper Umpqua was not mutually intelligible even with adjacent Athabaskan varieties, Golla (2011:70-72) classifies Upper Umpqua as a distinct language within Oregon Athabaskan, coordinate with Chetco-Tolowa and the Rogue River dialect network (comprised of Tututni and the other Lower Rogue dialects, Upper Coquille, and Galice) (cf. Figure 1.1 in chapter 1). However, Golla notes that Upper Umpqua "possessed a number of distinctive lexical and phonological features that set it apart from the rest of Oregon Athabaskan" and "tends to preserve older forms and thus tends to bear a superficially closer resemblance to California Athabaskan languages than do the Rogue River dialects or Chetco-Tolowa" (2011:71). Golla has compiled the available documentation into a set of file slips, a preliminary analysis of which is offered in this appendix, considering in turn lexical and phonological evidence.

### C.1. Lexical Evidence

Golla (p.c.), in a preliminary lexicostatistical analysis of the Upper Umpqua data, identifies 145 glosses with forms attested in each of Upper Umpqua, at least one other Oregon Athabaskan language, Hupa, and at least one other California Athabaskan language. Just over half of these glosses have the same cognate throughout PCA (i.e., they provide no evidence for subgrouping), but 73 of them have a form attested in Upper Umpqua that is not cognate with the form(s) attested in at least one other PCA language. 26 of these are unique to Upper Umpqua, but another 45 have a cognate elsewhere in the Oregon group to the exclusion of the California Athabaskan languages. For only two of the glosses is there an Upper Umpqua form that is cognate with California Athabaskan to the exclusion of the other Oregon languages. While these figures do not distinguish shared retentions from shared innovations, they clearly suggest that Upper Umpqua has a close affinity with the Oregon Athabaskan group. This scenario makes geographic sense, since Upper Umpqua is the northernmost of the Athabaskan languages once spoken in southwestern Oregon and hence the farthest removed from the California languages (cf. Map 2 in chapter 1).

The present study assembled a Swadesh-100 list for Upper Umpqua in which 85 items were found and 15 were missing. These data were added to the lexical dataset used in the computational study reported in this chapter. Bayesian analysis was performed using a multi-state coding that retained polymorphisms and a non-clock evolutionary model (analogous to the

configuration that produced Figure 2.1 in §5.1). This produced the following tree, rooted with Tlingit as the outgroup:

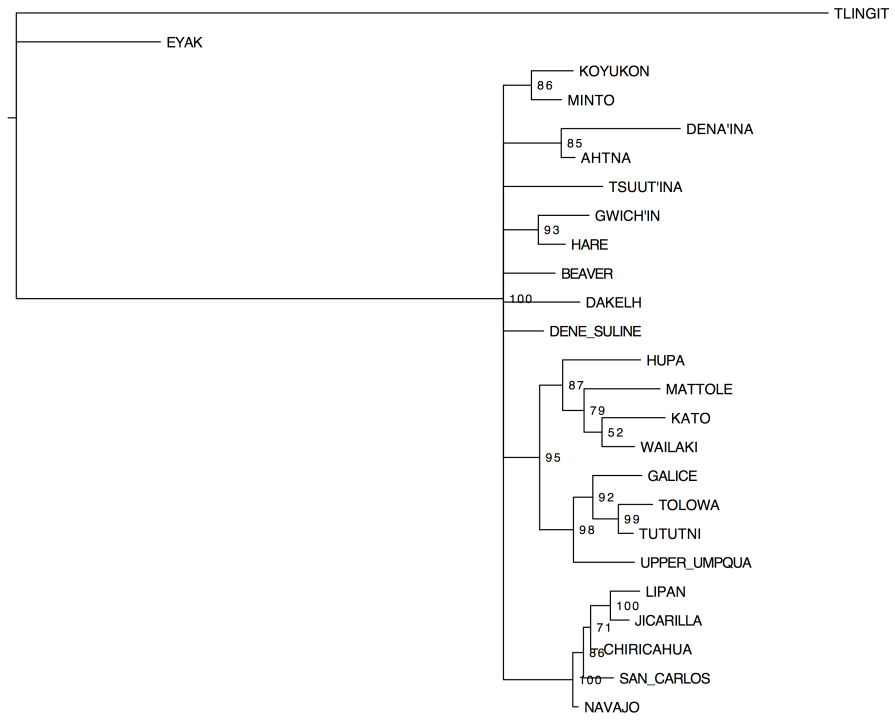



Figure 2.8: Majority-rule consensus tree for the lexical dataset including Upper Umpqua coded as multi-state characters with all polymorphisms included (non-clock evolutionary model)

When Upper Umpqua is added to the lexical dataset, a well-supported (95%) PCA subgroup still emerges. Moreover, Upper Umpqua appears in a subgroup together with the other Oregon Athabaskan languages (Galice, Tututni, and Tolowa) and is the earliest offshoot of the group. This is consistent with Golla's view that Upper Umpqua is lexically divergent relative to the other Oregon Athabaskan languages.[35]

Of the two PCA-specific Swadesh-100 items identified in the NeighborNet analysis in §5.4 above, Upper Umpqua has the root *-da?* 'mouth', but as noted this may be a shared Proto-Athabaskan retention in PCA. Forms glossed as 'cloud' are found in the Tolmie and Gatschet wordlists, who transcribed <eesh teugh uk> and <íshtchigak> respectively; Gatschet also recorded <to$^{l}$-ishtchiga> 'there are not clouds', where the first syllable <to> is undoubtedly the negative proclitic *do:* found in many Athabaskan languages. In the computational analysis, these forms were coded as non-cognate with other PCA languages because of their global dissimilarity with the bare root *?ah* found in California and *?ak* found in Oregon. However, it is possible that the last syllables <uk> (Tolmie) and <ak> (Gatschet) are reflexes of this same

[35] Unexpected in Figure 2.7 is the fact that adding Upper Umpqua to the dataset leads to 87% support for the California Athabaskan subgroup, supported only moderately at 69% in Figure 2.1 (which was produced with analogous data coding and evolutionary model). At present there is no explanation for this difference, but it underscores the importance of Upper Umpqua to understanding the historical development of PCA.

Proto-PCA root.[36]

## C.2. Phonological Evidence

The phonological data also make it clear that Upper Umpqua is a member of the PCA subgroup identified in this chapter, and while it is similar to the other Oregon languages in certain key respects it is also divergent from them in others. Like other PCA languages, Upper Umpqua appears to lack lexical tone, or at any rate retains coda laryngeal segments that develop into tone elsewhere in Athabaskan.[37] Since as already noted this is a shared retention from Proto-Athabaskan it is not positive evidence for subgrouping on its own, but at least is not inconsistent with including Upper Umpqua under a PCA subgroup. More importantly, Upper Umpqua participated in some of the telltale PCA consonant mergers: *dl with *l (#3 in Appendix A, attested only in 'mouse'), *q with *x̣ as /x/ (#6-7 in Appendix A), and *z with *s as /s/ (#10 in Appendix A) (Hoijer 1960). These mergers are illustrated in Table 2.3, which compares reconstructed Proto-Athabaskan vocabulary with normalized Upper Umpqua forms and their Tututni equivalents:[38]

| Proto-Athabaskan | Upper Umpqua | Tututni | gloss |
|---|---|---|---|
| *dlǫ̃ʷ'ə ~ *dlu̯:n'ə | ɬũ'- | ɬəm'e | 'mouse' |
| *qʊn' | xu$^{n}$' | x$^{w}$ən' | 'fire' |
| *-qeʔ | -xe' | -xe' | 'foot' |
| *-zǫ̃ts' | -ɬ-səs | -sə's | 'skin' |
| *ỹę̃:z | nes | nes | 'long' |
| *-zel | -səɬ ~ -zəɬ | -səɬ | 'hot' |
| *q$^{w}$e-nę̃:z-ya:ỹ | x$^{w}$eneza ~ x$^{w}$enezə$^{n}$ | xwese | 'ten' |

Table 2.3: Pacific Coast Athabaskan consonant mergers in Upper Umpqua

As the last two forms in Table 2.3 illustrate, the merger of *z and *s may have been incomplete in Upper Umpqua: exceptionally within Pacific Coast Athabaskan, *z appears to have been retained in some cases (perhaps retaining old Athabaskan stem voicing alternations in 'hot').

Within the Oregon group, as noted by Golla (2011:71) Upper Umpqua is somewhat more conservative phonologically than other Oregon languages. Diagnostic sound changes are the de-affrication and merger of coronal affricates: *ts and *dz as /s/ (also merging with *s and *z, item #9 in Appendix A), and *tš and *dž (and their labialized counterparts) as the retroflex /ṣ/ (item (h) in §2.1). Relevant data are given in Table 2.4:

---

[36] The fact that Tolmie separates <uk> from preceding syllables with a space suggests that it has a glottal stop onset. Gatschet's <to$^{l}$-ishtchiga> may also include a variant of the same stem if final <a> is *ʔah*. Stem alternations involving final -k and -h are common across Athabaskan and are likely to be the source of the variable realization of the stem in California and Oregon Athabaskan.

[37] As elsewhere in Oregon Athabaskan, some of the documentation (especially comments in Harrington's field notes) suggests that there may be incipient quasi-tonal phenomena in Upper Umpqua, but this remains to be explored in more detail.

[38] Proto-Athabaskan forms in this section are from Krauss (2005), Krauss and Leer (1981), and Leer (2011); Tututni vocabulary is from Golla (2008) and Dorsey (1884i). Superscript <$^{n}$> indicates that nasality is variably transcribed as a full segment or vowel nasalization, or omitted entirely.

| Proto-Athabaskan | Upper Umpqua | Tututni | gloss |
|---|---|---|---|
| *tse: | se | se | ‘rock’ |
| *tsə̰̃ʔ | səⁿʾ | sən’ | ‘meat’ |
| *-tšəỹ | -š(ʷ)ən | ṣən | ‘bad’ |
| *tša’ | ša | ṣa’kuɬ | ‘beaver’ |
| *-yədzə’ | -idzi | -ise | ‘penis’ |
| *-džʷe:yə’ | -dži’ | ṣe’ | ‘heart’ |
| *-džəɣə’ | -džəɣe’ | ṣəɣe’ | ‘ear’ |
| *džʷe:nə | džine | ṣe-ṣ | ‘day’ |

Table 2.4: Oregon Athabaskan consonant changes in Upper Umpqua

The first four forms in Table 2.4 show that the Upper Umpqua reflexes of aspirated *ts and *tš are /s/ and /š/ respectively; there is no strong evidence for retroflex /ṣ/ in Upper Umpqua, which is probably similar to Galice in this regard (Hoijer 1966). The last four forms, however, demonstrate that the unaspirated affricates *dz and *dž are retained as such in Upper Umpqua (the latter noted by Hoijer 1960). It thus appears that the mergers affecting coronal affricates in most Oregon Athabaskan varieties occurred in two stages, first affecting the aspirated affricates then affecting the unaspirated counterparts, with Upper Umpqua participating in the former but not the latter sound change.

Another sound change distinguishing the Oregon languages from the California languages within PCA is the treatment of coda nasals following a full vowel (#23-24 in Appendix A). These are retained as full segments in the California group, but they are lost in Oregon either entirely (e.g., Tututni/Lower Rogue, Upper Coquille) or retained as nasalization (Galice, Chetco-Tolowa, Pistol River). Upper Umpqua data pertaining to coda nasals are given in Table 2.5; as noted above, superscript <ⁿ> indicates that some researchers transcribed a full nasal segment, others vowel nasalization, or occasionally no nasality at all:

| Proto-Athabaskan | Upper Umpqua | Tututni | gloss |
|---|---|---|---|
| *-ʾe:n | -ʾi | -ʾi | ‘see’ |
| *-na:ỹ(ʷ) | -na | -na | ‘drink’ |
| *-ha:ỹ | -ya | -ya | ‘eat’ |
| *-ye:n | di-nyi | di-mi | ‘sharp’ |
| *-tʾa̰:nʾ ~ -tʾa̰:nəʾ | -tʾaⁿʾ | -tʾa’ | ‘leaf’ |
| *qʾṵ:nʾ | kʾõʾ | kʾu’ | ‘fish eggs’ |
| *ɢṵ̃:ts ‘rhubarb’ | guⁿs | gus | ‘camass’ |
| - | hĩsle | tesle | ‘steelhead salmon’ |

Table 2.5: Coda nasals and nasalization with full vowels in Upper Umpqua

Several verb roots documented for Upper Umpqua, such as the first four forms in Table 2.5, have lost their nasal codas entirely following full vowels. In this respect, Upper Umpqua has an affinity with Tututni and the Lower Rogue dialects in losing nasality entirely, unlike Galice, Chetco-Tolowa, and Pistol River, which retain nasal codas as nasalization on the preceding full vowel.[39] Upper Umpqua, however, sometimes retains nasality following long vowels. This

[39] This distribution of complete loss of nasality is reminiscent of Krauss’ (1973) conception of the spread of innovations within a dialect network leading to cross-cutting isoglosses and no clear linguistic boundaries.

appears to happen when the nasal segment in question is glottalized, as in 'leaf' and 'fish eggs'. The last two forms in Table 2.5 are cases where Upper Umpqua retains nasalization on a full vowel even where no reconstruction to a full nasal segment is available. The Proto-Athabaskan root that is the source for 'camass' is reconstructed with a nasalized vowel by Krauss (2005), and while no reconstruction for 'steelhead salmon' is available, the nasality is presumably archaic, whether inherent to the vowel or due originally to a full segment. In both of these cases the syllable with nasalization is closed. Since nasal codas that underwent full deletion in Oregon Athabaskan presumably passed through a stage of nasalization on the preceding full vowel (**Vn > *$\tilde{V}$ > V), the generalization is that nasalization is retained in Upper Umpqua in closed syllables. Here again, Upper Umpqua participates in typically Oregon Athabaskan innovations, but only partially.

Although conservative in some key respects, Upper Umpqua is phonologically innovative in other ways. Three of the most noticeable innovations with respect to other Oregon languages are mentioned here, based on Golla's notes. First, most glottalized obstruents (other than /t'/) debuccalize word-initially, as shown in the first four forms in Table 2.6. The last three forms demonstrate that this only occurs in polysyllabic words:

| Upper Umpqua | Tututni | gloss |
|---|---|---|
| 'exe | tṣ'axe | 'woman' |
| 'itɬ'a' | tš'ətɬ'a' | 'spear' |
| 'axəs | k'axəs | 'arrow' |
| tɬ'əɣəš ~ 'əɣəš | tɬ'əɣəš | 'snake' |
| k'õ' | k'u' | 'fish eggs' |
| k'əh ~ k$^{w}$'əh | k$^{w}$'ah | 'fat' |
| tɬ'oh | tɬ'uh | 'grass' |

Table 2.6: Debuccalization of word-initial glottalized stops in Upper Umpqua

Debuccalization seems to occur consistently with Proto-Oregon initial *ts', *tš', and *k', but variably with *tɬ': 'snake' is transcribed with debuccalized onset by Hale (1846), variably with and without debuccalization in Milhau (1856), and with an affricate by Barnhardt (1859), Gatschet (1877), Jacobs (1930s), and Harrington (1981). This trajectory suggests that the debuccalization of word-initial /tɬ'/ might have occurred much later than it did for other glottalized stops, with residual variation in the 19th century resolved in favor of the archaic affricate.[40]

This tendency to lenite word-initial affricates apparently is also instantiated in Upper Umpqua's treatment of front-velar *k̲̂. As noted, in most Oregon Athabaskan varieties Proto-Athabaskan *tš is realized as the retroflex fricative /ṣ/, as /š/ in Upper Umpqua and Galice (where it merges with *š). A palatal affricate /tš/ does occur in Oregon Athabaskan, but it is a palatalized reflex of Proto-Athabaskan front-velar *k̲̂ and *ĝ. This implies that de-affrication and retroflexion of *tš precedes palatalization, since otherwise *k̲̂ should be found as /ṣ/ or /š/ as well. But merger of *k̲̂ and *š is precisely what is sometimes found in Upper Umpqua:

[40] Golla's file slips include a note to the effect that Upper Umpqua *'əɣəšxe* 'eel', which occurs only in Jacobs (1930s), may be cognate with Hupa *tɬ'iʍxa:n* (< *tɬ'iwiʍxa:n). If so, this would be an example of debuccalization of *tɬ' that was successful.

| Proto-Athabaskan | Upper Umpqua | Tututni | gloss |
|---|---|---|---|
| *-k̲̂es | -tšis | -tšəs | ‘to strike’ |
| *-k̲̂ʊx̲̂$^{w}$ ‘big’ | -tšo | -tšuh | ‘augmentative’ |
| *-k̲̂e’ | -še’ ~ -šela’ | -tšila’ | ‘tail’ |
| *k̲̂ən ‘stick’ | šin-tšo | tšən-tšuh | ‘tree, wood’ |
| *-k̲̂ə̲̂tɬ’ə | -šitɬ’e | -tšəl’e | ‘younger brother’ |

Table 2.7: De-affrication of word-initial /tš/ (< *k̲̂) in Upper Umpqua

The first two forms in Table 2.7 show the same outcome /tš/ from *k̲̂ in Upper Umpqua as found in Tututni, implying that de-affrication of *tš also preceded palatalization in Upper Umpqua. The subsequent merger **k̲̂ > *tš > š in the last three forms in Table 2.7 is therefore most likely a separate development, the conditioning environment for which is word-initial position. The verb root *-tšis* ‘strike’ and the augmentative enclitic *-tšo* would not occur word-initially: the enclitic by definition, and the verb root because minimality constraints in Athabaskan generally disallow monosyllabic verbs (cf. Hargus and Tuttle 1997). Although ‘tail’ and ‘younger brother’ are bound noun roots and would sometimes occur with possessive prefixes, in Oregon Athabaskan the third person possessor *m- is lost in most environments, retained only with vowel-initial roots and as labialization on roots beginning with a velar obstruent or glottal stop (Golla 1976). Such nominal roots, then, would often have been word initial. Reduction of *tš (from **k̲̂) restricted to word-initial position, together with paradigm leveling for the noun roots ‘tail’ and ‘younger brother’, can thus account for the difference between /tš/ and /š/ as reflexes of *k̲̂. This scenario is rendered more plausible by considering it part of a general tendency in Upper Umpqua to lenite word-initial obstruents, one that is not found elsewhere in Oregon Athabaskan.[41]

Two other Upper Umpqua sound changes are mentioned here for completeness. First, there are a number of cases where /h/ in Upper Umpqua corresponds to aspirated /t/ elsewhere in PCA, illustrated in the first four forms in Table 2.8. This appears to occur only before a high front vowel, as the last three forms in the table demonstrate:

[41] Note, however, that unaspirated /dž/ does not undergo this word-initial reduction (e.g. in ‘day’ in Table 2.4), so this account would have to restrict the structural description of word-initial de-affrication to target aspirated /tš/ alone.

| Upper Umpqua | PCA | gloss |
|---|---|---|
| ’əɬhĩ’ | Coq. ts’əɬti’<br>Hupa ts’iɬtiŋ’ | ‘bow’ |
| hilama | Tut. ti’la<br>Hupa tehla:n | ‘whale’ |
| hintah-dəne ‘wolf’<br>lit. ‘in the woods - person’ | Tut. tita-<br>Hupa tintah | ‘in the woods’ |
| hĩsle | Tut. tesle | ‘steelhead salmon’ |
| təɬ | Tut. təɬ | ‘basket’ |
| təne | Tut. təne<br>(< P-A *təỹə) | ‘trail, road’ |
| -taɬ | Tut. -taɬ | ‘kick’ |

Table 2.8: *t > h in Upper Umpqua

Another sound change frequently encountered in Upper Umpqua is the shift of coda *d to /g/. This is seen in Table 2.9:

| Upper Umpqua | PCA | gloss |
|---|---|---|
| -’ag | Tut. -’ad<br>Hupa -’ad | ‘wife’ |
| ɬəg | Tut. ɬəd<br>Hupa ɬid | ‘smoke’ |
| -nəg | Tut. -nad<br>Hupa -na:d/-na:t’ | ‘lick’ |
| -tš’əg | Tut. tṣ’ide’ ‘sickness’<br>Hupa -tš’a:d | ‘sick’ |
| -t’og | Tut. -t’u’<br>Hupa -t’od/-t’ot’ | ‘suck’ |

Table 2.9: Coda *d > g in Upper Umpqua

In summary, the documentation of Upper Umpqua, while sparse overall, is robust enough to reveal an Oregon Athabaskan language that is somewhat divergent from the rest of the subgroup. The lexical affinity between Upper Umpqua and the other Oregon languages is supported both in gross lexicostatistical similarity and with an explicit computational model of Athabaskan linguistic evolution. Phonologically, Upper Umpqua is uniquely innovative in some regards, but conservative in others: although the language participated to some extent in sound changes typical of Oregon Athabaskan and of Pacific Coast Athabaskan more generally, in some cases it did so only partially, especially in retaining the Proto-Athabaskan affricates *dz and *dž as such and in sometimes maintaining the voicing contrast between /s/ and /z/. These results are taken as support for Golla’s (2011:70) classification of Upper Umpqua within Pacific Coast Athabaskan.

# Chapter 3: Variation and Dialect Contact in Hupa

## 1. Introduction

As discussed in chapter 1, Hupa was traditionally spoken in Hoopa Valley along the lower Trinity River in Humboldt County, California. A devastating influx of miners and other foreigners during and after the California gold rush led to the establishment in 1864 of the Hoopa Valley Reservation on large portions of traditional Hupa territory. Although the overall Hupa population was drastically reduced from its pre-contact level, many Hupa people remained more or less in situ in their ancestral homeland.[1] In the same period, refugees from surrounding areas migrated into Hoopa Valley, some coerced more directly than others. These newcomers included Athabaskans from Redwood Creek and the Trinity's South Fork who spoke dialects that are reported to be similar to Hupa (Golla 2011:76-78). While these groups had been in contact with the Hupa prior to resettlement, the nature and frequency of their interactions were suddenly altered. In the space of less than two decades, the entire region underwent a radical transformation of its physical and social landscapes, even for those whose traditional settlements fell within the boundaries of the reservation (Golla 1996a:365).

This chapter explores the dynamics of linguistic variation and dialect contact in Hoopa Valley, addressing the question of whether or not contact between the Hupa of the lower Trinity and their neighbors led to koineization as found in better-studied cases of dialect contact (cf. Kerswill and Trudgill 2005 and discussion in §3 of chapter 1). After an overview of the Hupa dialect groups (§2) and a brief history of their resettlement in Hoopa Valley (§3), two parameters of variation are traced through the documentation. The first (§4) is a regional dialect difference related to the merger of two Proto-Athabaskan back-velar obstruents as an aspirated stop /k/ versus a fricative /x/ (character #7 in Appendix A of chapter 2). The Whilkut dialect of Hupa exhibits variation between [x] and [k] for this merged category, which became the target for rapid dialect leveling in the new community (Goddard 1914b). The second (§5) is a lexically-conditioned difference in word-initial *ni-* vs. *'i-*. It is less clear that this was originally a dialect difference and may instead reflect pre-existing sociolinguistic variation that was already part of Hupa proper when other groups arrived in Hoopa Valley. It shows an interesting distribution in the 20th century, with some speakers apparently converging on the *'i-* variant across the board, while others did so only for particular items; across the population, however, this variation appears to have been largely maintained.

Discussion of the factors influencing the diachronic trajectories of these two linguistic variables under koineization is framed in terms of the demographic versus indexical modes of explanation for the resolution of dialect variation outlined in §3 of chapter 1. With respect to the direction of leveling of [k] ~ [x] variation when koineization occurred, to some extent both approaches make the same prediction, that [x] found in Hupa proper of the lower Trinity would have an advantage over other dialects. The lower Trinity Hupa were more numerous than other groups on the reservation (consistent with demographic explanations), and they are likely to have had stronger primary claims on resources than incoming groups did since the reservation was established in their aboriginal territory (consistent with Hill 2001). There were also differences in the prestige of Hupa proper relative to other dialects, some of which might have been extensions a pre-contact status quo (Powers 1877, Goddard 1914a); these might have made

[1] According to Kroeber (1925:131), the relatively high mortality rate among men in this period is evidence that "[b]ullets, not disease, killed in these first years." Cf. Goddard (1903:8-11, 1914a:267-270).

South Fork and Redwood Creek groups likely to assimilate to Hupa in order to avoid the use of stigmatized sociolinguistic indices. The failure of the *ni-* vs. *'i-* parameter of variation to undergo complete leveling is likewise attributed to a confluence of factors, including its relative structural complexity.

## 2. Hupa Divisions

Four principal divisions of Hupa-speaking people are generally identified at least since Powers (1877) and Goddard (1903, 1905, 1914a,b). These are the Hupa proper, who lived aboriginally along the lower Trinity River in Hoopa Valley; the South Fork Hupa, known as "Kelta" to Powers (1877:89-91) and Goddard (1903:7) and "Tsanungwha" to Merriam (1850-1974), at the confluence of Trinity's northern and southern branches, south of Hoopa Valley; the Chilula, who occupied the lower drainage of Redwood Creek and adjacent territory to the west of Hoopa Valley; and the Whilkut, whose territory comprised upper Redwood Creek (Merriam's "Kloki Whilkut") and portions of the Mad River southwest of Hoopa Valley near present-day Blue Lake (Merriam 9/80; cf. Baumhoff 1958:203). In addition, a group of Hupa-Chimariko bilinguals were reported to live at New River, the Chimalakwe of Powers (1877:91-95; cf. Golla 2011:77).[2] Golla (1970:10, 1996a:364) has called into question the distinction between Chilula vs. Whilkut, pointing out that both groups were known by the ethnonym *xwiłq'it-xwe:* 'people of Redwood Creek (*xwiłq'it-ding*)'. However, he elsewhere notes that the Whilkut, but not the Chilula, were excluded from Hupa ceremonies (Sapir and Golla 2001:55, 989-990), indicating that there was some social differentiation among the groups living along Redwood Creek.[3]

It is often suggested in the literature that these divisions of Hupa-speaking people corresponded in some way to dialect divisions as well, although there is sometimes disagreement over the extent (or perhaps importance) of these linguistic correlates. Goddard noted with respect to the Whilkut that "their speech is a fairly well marked dialect as compared with the Hupa tongue as regards vocabulary and pronunciation", and that "their language is closer to the other Athapascan dialects than is the Hupa itself" (1903:7).[4] Goddard apparently considered Chilula to be closer to Hupa than Whilkut was, stating that "[w]ith the Hupa the Chilula are very intimately connected. There is only a slight difference in dialect" (1914a:267). While acknowledging that the Chilula shared some distinctive cultural traits with the Whilkut, he

[2] Powers describes the situation at New River thus: "A Mr. White, a pioneer well acquainted with the Chi-mal'-a-kwe, who once had an entirely distinct tongue, told me that before they became extinct they scarcely employed a verb which was not Hupâ" (1877:72).

[3] As Conathan (2004:95-96) points out, the significance of these high-level groupings is uncertain, since it is generally held that there was only loose political organization above the level of the village according to Kroeber's classic "tribelet" analysis in the region (Kroeber 1925; cf. Golla 2011:3). For example, in his ethnography of the Hupa, Goddard (1903:7) observed that "[s]trictly speaking there are no tribes on the coast of California. The divisions are natural and of varying degrees, rather than political and well-marked." Of the Chilula specifically Goddard (1914a:271) stated, "On the whole there appears to have been a surprising lack of political coherence between various villages. It is difficult to determine what constituted the larger units."

[4] There is uncertainty as to whether Whilkut or Chilula is meant here, since Goddard describes the territory of this group as being "along the middle portion of Redwood Creek and the Bald Hills between that stream and the Klamath River," which would seem to describe the territory of the Chilula more accurately than that of the Whilkut.

concluded that "[m]ainly, however, they seem to have been one with the Hupa in language, culture, and political feeling" (1914a:266).[5] This lack of Chilula linguistic distinctiveness is further implied in Goddard (1911:93), where he recognized but two dialects, Hupa proper and Whilkut: "In the middle of this Pacific coast division [of Athabaskan] are two dialects very closely connected. One of them was formerly spoken on upper Redwood Creek and middle Mad river in Humboldt county, California; and the other, the Hupa of which this paper treats, on the lower (northern) portion of the Trinity river."

By contrast with Goddard's reports, Merriam did not find separate Chilula and Whilkut dialects of any significance: "The ʾHwil-kut belong to the Hooˡ-pah linguistic division. The language is uniform throughout Redwood Creek Valley except for one or two slight differences of pronunciation" (9/81). Golla (2011:76-77) has also suggested that there were few differences between Hupa proper, South Fork Hupa, and the Redwood Creek dialects. In Reichard's (1922) fieldwork with Whilkut speakers at Blue Lake, her consultants commented that they considered Hupa and Redwood Creek to be dialects of the same language, yet they also maintained that "their dialect is like Hupa but not so 'fancy', they consider it more original, more 'right'." While the adjective 'fancy' may not have been intended in a positive light, it does suggest an underlying assumption that Hupa was associated with an elevated status (even if that status was contested). These comments show community-internal awareness of the affinity between Whilkut and Hupa as well as definite metalinguistic evaluations of differences distinguishing them.[6]

The documentation of South Fork Hupa is extremely sparse, confined to a smattering of unpublished material collected by Goddard in the first decade of the 20th century and Merriam's research on toponyms. A South Fork Hupa text transcribed by Goddard (n.d.) has not been analyzed in detail for this study, but superficial examination of the content suggests that it is indistinguishable from Hupa proper. Goddard himself maintained that "[t]he language spoken at Southfork differs but slightly from that used in Hupa valley" (1903:7). Apparently by the first decade of the 20th century significant differences between Hupa proper and South Fork Hupa had largely disappeared. That such differences might once have existed at all is found in the report of Powers (1877:89), who claimed that "[t]hey formerly had a distinct language, but the Hupâ so encroached upon it that it now amounts to nothing." This is also attested in the memory of people Merriam worked with: "The language of the Tsănungwhă differs only slightly from that of the Hoopah, but in early days, old Indians say, it was very different"

[5] Goddard later suggested that linguistic divisions between the Hupa and Chilula on the one hand, and the Whilkut on the other, corresponded to deeper divisions in cultural practice and political sentiment. He noted that "[t]he Athapascans of upper Redwood Creek and Mad River [i.e., the Whilkut] had a culture dissimilar from the Hupa in many points" (1914a:266). That he did not regard Chilula and Whilkut as interchangeable is further emphasized in his statement that "[t]oward the Whilkut the Chilula seem to have entertained a feeling of distrust and condescension" (1914a:267). The Chilula, moreover, were invited to observe Hupa ceremonies, whereas the Whilkut were excluded from such activities (1914a:267, 272). Goddard equivocated on this point, however, in noting that the Chilula vs. Whilkut distinction "is made rather in accordance with the attitude of the Hupa than from any definitely expressed feeling on the part of the Chilula" (1914a:272).

[6] Reichard's notes also include information related to multilingualism and an asymmetric use of Whilkut with respect to Wiyot, a non-Athabaskan language spoken in the area: "Found there the Indians listed, but they talk 'Redwood' an Athapascan dialect. Jim & Dan Brock can talk Wiyot but talk mostly Redwood because Jim's wife, Molly, talks it."

(9/58). These statements can be taken as evidence that the South Fork Hupa once spoke a dialect distinct from Hupa on the lower Trinity, although precisely what these differences were are unknown.[7]

## 3. Resettlement and Koineization

Whatever linguistic differences existed among different Hupa-speaking groups in pre-contact times, there is agreement in the literature that they underwent leveling when the Chilula, Whilkut, and South Fork groups were resettled on the Hoopa Valley Reservation. That is, the outcome of contact was koineization, as expected by the three-stage model of dialect contact presented in chapter 1 (Kerswill and Trudgill 2005). This is clear enough in the case of South Fork, where as noted the leveling seems to have been so rapid that no differences remained by the turn of the 20th century. Moreover, Goddard (1914b:291) suggested that at least some differences between the Redwood Creek dialects and Hupa proper had been lost as a result of resettlement in Hoopa Valley: "The larger number of the surviving natives of that region were with the Hupa when young and use the Hupa pronunciation" (probably referring to velar obstruents discussed in §4). This description suggests an asymmetric contact whereby Redwood Creek speakers assimilated to the Hupa, not vice-versa.

This directionality is expected given the status of Hupa in Hoopa Valley described by Powers:

> "In the Hupâ reservation, in the summer of 1871, the Hupâ constituted not much more than half of the occupants, yet the Hupâ was not only the French of the reservation, the idiom of diplomacy and of intercourse between tribes, but it was also in general use within each rancheria. I tried in vain to get the numerals of certain obscure remnants of tribes; they persisted in giving me the Hupâ, and in fact they seemed to know no other." (Powers 1877:72-73)

This appears to have been part of a broader reduction in cultural distinctiveness: Powers lamented of the Chilula that "[v]ery little can be positively stated of their customs, for all that remain of them have been removed to the reservation where the process of absorption into the Hupâ has been completed", a situation he believed to obtain for the Whilkut as well (1877:87-89). This is almost certainly an overstatement, since the reservation had been established a scant half-dozen years before Powers' visit, but it does suggest a situation in which the Hupa, more numerous and living more or less in situ in their traditional homeland, exerted a strong influence (linguistic and otherwise) on groups that had recently migrated into Hoopa Valley.

It should be noted that Powers' assessment of the relative size of the populations in Hoopa Valley is probably in error, with the Hupa comprising significantly more than half of the total: Kroeber (1925:131), citing a reservation census conducted in 1870, gives a figure of 641 Hupa and 233 others, 106 of whom were from Redwood Creek (see also Goddard 1903:10). It is known that Redwood Creek people were still arriving in Hoopa Valley as late as 1888 (Sapir and Golla 2001:28), and examination of the reservation census rolls shows ongoing migration into Hoopa Valley into the 20th century (Bureau of Indian Affairs 1967). However, at present it is unclear in precisely what proportions the two groups were found after 1870 (and as Kroeber

[7] Note, however, that both statements could instead be interpreted as indicating that there had been a recent shift from some other language to Hupa, as apparently was occurring at New River at the time of contact.

points out, even the 1870 figures might be misleading). Certainly, however, the Hupa proper constituted a plurality on the reservation.

## 4. [k] ~ [x] Variation

This section explores in detail one parameter of Hupa dialect variation, virtually the only specific discussion in the literature of differences between Hupa proper and the Redwood Creek dialects: the merger of the Proto-Athabaskan voiceless velar fricative *x̣ and aspirated velar stop *q as either /k/ (IPA /$k^h$/) or /x/ (Goddard 1914b; #7 in Appendix A of chapter 2). The distribution and diachronic trajectory of these two variants in Whilkut, Chilula, and Hupa proper are considered in turn.

### 4.1. Whilkut

Goddard worked with four Redwood Creek consultants in 1906-1907, recording a sizeable collection of texts from a woman known only as the wife of a man named Molasses (henceforth "Wife of Molasses"). She was an elderly speaker of the "middle to upper Redwood Creek" variety (i.e., Whilkut) and still resided in that region. Her husband was "a native of Mad River", most likely one of the Mad River Whilkut divisions identified by Merriam. Goddard maintained the couple had "lived for many years at Hoopa before resettling on Redwood Creek" (1914a:266), but this estimate is probably inflated. They appear in the reservation census conducted in Hoopa Valley only in the year 1889 (Bureau of Indian Affairs 1967) among a group identified as "Redwood of Hoopa." According to the report of the reservation administrator for that year, this group had been living on Redwood Creek "outside the reservation" before being moved into Hoopa Valley (Office of Indian Affairs 1889:123). The couple were already adults when they arrived: their ages in the census are given as 40 and 39 respectively, so they were born around 1840, a decade before Euro-American encroachments in the region began in earnest. Moreover, the fact that the couple appears on the census in just the single year suggests that they returned to their home territory soon after arrival. These considerations make it unlikely that features of their speech were due to what was probably a brief exposure to dialect mixing during their residency in Hoopa Valley.

Goddard observed that Wife of Molasses "used kʿ [aspirated $k^h$] where Hupa has x, agreeing in this respect with the Athapascans in the region immediately south" (Goddard 1914b:291). Examples of [k] in her texts include the following:[8]

| Whilkut form | gloss | Hupa equiv. |
|---|---|---|
| a. <-kʿûn> | 'handle filled container' | -xa:n/-xun |
| b. <yis kʿûn> | 'tomorrow' | yisxa:n/yisxun |
| c. <nakʿ> | two | nahx |
| d. <kī kʿak> | A-frame net | k'ixa:q' |

Table 3.1: Examples of [k] in Whilkut texts by Wife of Molasses

By the "Athapascans in the region immediately south", Goddard meant the other California

[8] Here and elsewhere segments of interest are underlined. Goddard's original orthography is retained, and Hupa equivalents are rendered in the practical orthography used in the Hupa Language Dictionary (Golla 1996b), where <u> is [ə], equivalent to Goddard's <û>. Goddard's raised inverted comma indicates aspiration.

Athabaskan languages: Wailaki and the other Eel River dialects, Mattole, and Kato. As noted in chapter 2, in these languages Proto-Athabaskan velars *q and *x̣ merge as /k/. In Hupa they merge in the opposite direction, as /x/ (Hoijer 1960:965).[9]

Goddard found in his work with the elderly Wife of Molasses that it was "impossible to record a complete text directly. The interpreter, O'Haniel Bailey, a Whilkut, supplied many phrases and sentences. These are Hupa in their phonetics" (implying that Bailey was one of the younger generation of Redwood Creek people who "were with the Hupa when young and use the Hupa pronunciation") (1914b:291).[10] Unfortunately, Goddard does not distinguish material recorded directly from Wife of Molasses versus material produced by Bailey, so the texts are an undifferentiated admixture of their idiolects. There are many examples of [x] in these texts where [k] is expected given Goddard's description, and these are found even in fairly close proximity to one another:

| Whilkut form | gloss | Hupa equiv. | ref. |
|---|---|---|---|
| a. <kī la k̲ʻûtc> | 'boy' | kile:x̲ich[11] | 324.11 |
| b. <kī la x̲ûtc> | 'boy' | | 324.12 |
| c. <x̲ōn ta> | 'house' | x̲ontah | 324.9 |
| d. <k̲ʻōn ta> | 'house' | | 324.11 |

Table 3.2: Variation with [k] and [x] in Whilkut texts by Wife of Molasses

While most words with [k] can perhaps be ascribed to Wife of Molasses, and those with [x] to Bailey, there is evidence that there was some mixing of the two. Assuming that Goddard transcribed each word from exactly one speaker, it is significant that some words have both [x] and [k]: it indicates that at least one speaker, Wife of Molasses or Bailey, was not consistent in his or her choice of velar obstruent. The following are examples:

[9] This affects both plain and labialized /k/, /x/ and /k$^{w}$/, /x$^{w}$/, which are conflated in the discussion.

[10] Bailey's name has not been located in the reservation censuses for Hoopa Valley, and his surname is one commonly found in censuses from the Blue Lake Rancheria. The same O'Haniel Bailey was one of Merriam's consultants in 1910 and apparently resided in the Blue Lake region (Baumhoff 1958:201). Bailey may have been one of the Whilkut people whose family migrated into Hoopa Valley for a time but later returned to their home region, like Wife of Molasses.

[11] Initial orthographic <k> in this word is an aspirated front-velar stop, retained from Proto-Athabaskan *k̲ (IPA /k$^{jh}$/); the merger only affects back-velar *q and *x̣. Goddard consistently does not transcribe the initial aspiration in this word, both here in Whilkut and in his Hupa Texts (1904).

| Whilkut form | gloss | Hupa equiv. | ref. |
|---|---|---|---|
| a. <na xō wil de kʻal a kût> | 'he tracked him' | xo- '3OBJ'<br>-xiw/-xe: 'track'[12] | 324.7 |
| b. <kʻel weL xō lûn taʻ> | 'where he had camped'[13] | xo- 'areal'<br>-e:=l-wetɬ' 'to camp'<br>-xola:n 'evidently'<br>-tah 'among' | 329.15 |
| c. <xō kʻōn tau ne en> | 'his house used to be'<br>('his former house') | xo- '3POSS'<br>xontah 'house' | 326.12 |
| d. <xō kʻai ya> | 'his boy' | xo- '3POSS'<br>-xiy 'boy, son' | 335.19 |
| e. <mûx xa dje kʻō len> | '*Leptotaenia Californica*' | mixa:ch'e'-xole:n<br>'incense root' | 348.6 |

Table 3.3: Mixed [k] and [x] in Whilkut texts by Wife of Molasses

Tokens of the same morphemes are produced variably, e.g., the 3rd person animate possessor prefix is [ko] in *kʻō kʻōn tau* (328.2), contrasting with *xō kʻōn tau* in (3.3c). The words in Table 3.3 have the general character of so-called 'interdialect forms' identified in the dialect contact literature (Trudgill 1986) insofar as they have multiple values of a linguistic variable realized in a single form.[14]

While the words in Table 3.3 show that someone was mixing [x] and [k], they don't reveal precisely who it was. Evidence that both Bailey and Wife of Molasses participated in the variation is provided by another phonetic detail noted by Goddard (1914b:291): Wife of Molasses "often used a for e in many words, but this is considered a personal peculiarity" (Goddard 1914b:291).[15] The occurrence of [a:] where [e:] would be expected in Hupa can thus be used as a diagnostic for words uttered by Wife of Molasses. For example, Goddard transcribed the plural perfective verb stem *-de:tɬ'* 'go' both as <daL> and as <deL>, the former with [a:] ascribable to Wife of Molasses, the latter with [e:] presumably from O'Haniel Bailey. If so, a verb such as <kʻa na is deL> 'they came up again' (337.13; cf. Hupa *xa:-* 'up from beneath') would have been produced by Bailey and <xa si ta daL> 'they came over' (339.9, cf. Hupa *xa:-si-* 'up to the top') by Wife of Molasses, the former with initial [k] and the latter

[12] The verb stem here is probably <kʻa> plus the progressive suffix *-l*. This stem is transcribed as <xa> ~ <kʻa> in Goddard (1904, 1914b), but Goddard (1905:249) notes that the quality of the vowel perhaps should be [e] instead (cf. entry for *-xiw* in Sapir and Golla 2001:801).

[13] Cf. *xwe:lwe:tɬ'* 'he stayed there' (Sapir & Golla 2001, 7.10).

[14] However, these forms differ in an important respect from interdialect forms discussed by Trudgill, which involve intermediate values for phonemic categories realized over a continuous phonetic space and as such can be considered a sort of compromise or splitting the difference of input dialects' target values. Here, the compromise involves a choice of discrete values at different positions in the word.

[15] It is unclear what led Goddard to deem this a personal peculiarity rather than another vestigial dialect difference, whether it was his own opinion or that of the other Redwood Creek speakers he was working with.

with initial [x]. Thus, although Wife of Molasses produced [k] often enough for Goddard to consider it characteristic of her speech, and Bailey produced [x] often enough for Goddard to treat his utterances as "Hupa in their phonetics", both speakers were prone to produce both [x] and [k].[16]

That [k] was not an idiolectal feature of the speech of Wife of Molasses is further demonstrated by similar variation in the Whilkut vocabularies recorded by Merriam. Importantly, not all Redwood Creek groups were relocated to Hoopa Valley, many Whilkut people instead winding up at the Blue Lake Rancheria, established in 1908 along the Mad River some 30 miles south and west of Hoopa Valley, where Merriam's data were gathered. This provides an important point of comparison: Whilkut as spoken by people who resided in Hoopa Valley in close contact with Hupa, as reported by Goddard (1914b), versus those who resided at Blue Lake. The speakers Merriam consulted included the same O'Haniel Bailey that had interpreted texts from Wife of Molasses for Goddard:[17]

| Whilkut form | gloss | Hupa equiv. | ref. |
|---|---|---|---|
| a. <Nahˈ-tin-nukˈ-kă> | 'Hoopa of Hoopa Valley' | na:tinixwe: | 30/252 |
| b. <Ko-che-tis-teng> | 'bow' | xo'ji-ts'iɬting' | 30/192 |
| c. <Nungˈ-kahˈ-ten> | 'chief' | ningxa'te:n | 30/206 |
| d. <Ho-lahˈ> | 'hand' [3POSS] | xola' | 30/180 |
| e. <Ho-tsin-nĕˈ> | 'whole leg' [3POSS] | xots'ine' | 30/181 |
| f. <Keoˈ-hahk> | 'fish net' | k'ixa:q' | 30/195 |

Table 3.4: [k] and [x] in Whilkut from speakers at Blue Lake transcribed by Merriam

In many cases Merriam transcribed invariant [k] where Hupa has [x], as in (3.4a) through (3.4c). Merriam also recorded numerous words where Whilkut matches Hupa with [x] and no variation with [k], as in (3.4d) through (3.4f), especially the 3rd person animate possessor prefix (but cf. (3.5h) below). However, there are also many examples where the same root occurs with both [x] and [k], as in (3.5a) through (3.5k) – the last three of which show [k] and [x] occurring at different positions in the same token:

[16] Unfortunately, this test is unreliable because there are many words where both vowel qualities are found. For example, alongside <kʻel weL> '(where) he had camped' in example (3.3b) (with prefix and stem vowels both [e:] as in Hupa) are found <kʻa nal waL> 'she camped' (331.17, with both prefix and stem vowels [a:]), suggesting that the former was produced by Bailey (with initial [k]) and the latter by Wife of Molasses. However, one also finds <xwal weL kût> 'they camped' (339.11), with the prefix vowel [a:] and the stem vowel [e:], and <kʻel waL> 'he spent the night' (343.19), with the prefix vowel [e:] and the stem vowel [a:]. Perhaps the prefix vowel quality in these particular cases is irrelevant, because Goddard himself reports having difficulty distinguishing [a:] and [e:] following [x] in Hupa (1905:249), so presumably after the "strongly aspirated" [k] as well (Goddard 1914b:292).

[17] The Whilkut data were collected in September 1910 and September 1920; no speakers are named directly in the list, but John Stevens, Laura Stevens, and O'Haniel Bailey are mentioned in the finding aid for Merriam's papers at the Bancroft Library.

| Whilkut form | gloss | Hupa equiv. | ref. |
|---|---|---|---|
| a. <Hōn'-tah'> | 'house'<br>lit. 'fires-among' | xontah | 30/189 |
| b. <Hong'><br><'Kong> | 'fire' | xong' | 30/190 |
| c. <Hong'-teng><br><Kong'-teng> | 'fireplace' | xong'-ding | 30/190 |
| d. <Hŏn'-sel> | 'hot weather' | xonsiɬ | 30/215 |
| e. <Kon'-sel tes'-se-ah><br><Kō'n-sĭ$^{ch}$> | 'summer' | xonsiɬ | 30/215 |
| f. <Klik'-kon'><br><Klih$^{ch}$'hon> | 'sweet' | ɬixun | 30/203 |
| g. <Nahk'><br><Nah$^{'ch}$> | 'two' | nahx | 30/175 |
| h. <Ho'-wah><br><Koo'-wah> | 'teeth' [3POSS] | xowo' | 30/182 |
| i. <Hō'l-chā-ketch'> | 'little spotted skunk' | xolje:-xich | 52/38 |
| j. <Muk-kun'-tow-haw-len'> | 'wood rat' | mixontaw'-xole:n<br>'its houses - there are plenty' | 52/40 |
| k. <Ho$^{ch}$-'hĕ><br><Ho-kĕ'><br><Ho-krĕ-ă> | 'foot' | xoxe' | 30/181 |

Table 3.5: Variation involving [k] and [x] produced by Whilkut from speakers at Blue Lake transcribed by Merriam

The upshot is that the Merriam vocabularies exhibit much the same variation between [k] and [x] as documented by Goddard: [k] corresponding to Hupa [x] is produced frequently but not consistently by Whilkut speakers.

The other source of Whilkut documentation comes from a short wordlist recorded by Reichard (1922) in Blue Lake with a Whilkut speaker named Molly. Unlike Goddard and Merriam, Reichard transcribed only [k], never [x], as in the following examples:

| Whilkut form | | gloss | Hupa equiv. |
|---|---|---|---|
| a. | <ku wa'> | 'teeth' [3POSS] | xowo' |
| b. | <ko kɛ'> | 'foot' [3POSS] | xoxe' |
| c. | <ko kạ'> | 'husband' [3POSS] | xoxung' |
| d. | <kọ'> | 'fire' | xong' |
| e. | <kostá'n> | 'six' | xosta:n |
| f. | <kilä kɑtc> | 'little boy' | kile:xich |

Table 3.6: [k] in Whilkut at Blue Lake transcribed by Reichard (1922)

Interestingly, Reichard's consultant also produced words that were explicitly offered as examples of differences between her dialect and Hupa as spoken in Hoopa Valley. This included two body parts with the 3rd person animate possessor prefix, Hupa *xo-*, which she pronounced with Whilkut [k]:[18]

| Molly's Whilkut | Molly's Hupa | gloss | Hupa equiv. |
|---|---|---|---|
| a. <ku ká'nɛ'> | <ku Tca'ŋ ai'> | 'arm' [3POSS] | xoky'a:ng'ay |
| b. <ko tcé ye'> | <ko keṇ tsan> | 'heart' [3POSS] | xokyun-sa'a:n |

Table 3.7: [k] for [x] in Hupa words as produced by Whilkut speaker Molly, transcribed by Reichard (1922)

Both of these are cases where Whilkut retains a Proto-Athabaskan root in a given meaning and Hupa has innovated (*-gạnə' 'arm' vs. *-ky'a:ng'ay*, lit., 'it extends away'; *-džʷe:yə' 'heart' vs. *-kyun-sa'a:n*, lit., 'insides - round object lies there'). Thus, while Reichard's consultant was aware of these lexical differences between Hupa and her own dialect, she was not aware of, or unable to produce accurately, Hupa [x] in lieu of Whilkut [k].

## 4.2. Chilula

Goddard's (1914b) collection of Redwood Creek texts includes material from two additional speakers. Tom and Dan Hill were a father and son from lower Redwood Creek (i.e., Chilula) who moved to Hoopa Valley in 1888 (Goddard 1914a:265). Their ages in the reservation census of 1891 are given as 61 and 24 respectively, putting their dates of birth at approximately 1830 and 1867. As far as is known the two of them did not live among the Hupa until adulthood. They uniformly have the Hupa [x] and show no signs whatsoever of the [k] ~ [x] variation recorded by Goddard and Merriam with Wife of Molasses, O'Haniel Bailey, and speakers of Whilkut in the Blue Lake region. While it is possible that the Hills adapted to the Hupa pronunciation as adults upon moving into Hoopa Valley in the late 1880s, this seems unlikely due to differences in adult versus child adaptation to new dialects they come into contact with (Kerswill and Trudgill 2005, Labov 2007). It seems more likely that Chilula simply never had the [k] variant (or had already lost it by the mid-19th century). This would restrict the [k] variant to the upper Redwood Creek and Mad River Whilkut varieties – that is, the dialect bordering the territories of California Athabaskan languages where /k/ corresponds regularly to

[18] Item (3.7a) also shows palatalization of the root-initial front velar stop, which is a feature of Mattole, Kato, and some Eel River dialects.

Hupa /x/.

## 4.3. [k] ~ [x] in Hoopa Valley

In identifying [k] produced by Wife of Molasses as typical of the Whilkut dialect, Goddard (1914b) noted that Whilkut speakers who grew up among the Hupa instead used Hupa [x]. Adult Whilkut speakers who moved into Hoopa Valley might have retained the [k] variant, and it might be expected that this should be detectable in the documentation of Hupa proper. However, there is only marginal evidence in the earliest Hupa wordlists for a [k] variant at play in Hoopa Valley. The following table is an exhaustive listing of words with [k] for expected [x] that have been identified in wordlists collected in the 19th century (including the extensive vocabulary of Curtin 1888-1889):

| Hupa form | gloss | dictionary form (Golla 1996b) | ref. |
|---|---|---|---|
| a. <hwais-tai> | 'an Indian' | xo'osday ~ xwe'esday 'man' | Crook (1852-1861) |
| b. <quais-tai> | 'man' | xo'osday ~ xwe'esday 'man' | Crook (1852-1861) |
| c. <kin-nook-yeh-h> | 'to speak' | xiniwhye:wh 'I speak' | Azpell (1870) |
| d. <kíl-cha-kan> | 'deer' | k'iɬixun<br>lit. 'that which tastes sweet' | Powers (1870s) |
| e. <na-ki-tai-kich-lá-kan> | 'rabbit' | na:q'itah-k'iɬixun 'jackrabbit'<br>lit. 'on the gravel bar - deer' | Powers (1870s) |

Table 3.8: [k] in early Hupa wordlists

A reasonable interpretation of Crook's <qu> in (3.8b) is a labialized velar stop, suggesting that he heard the same word two different ways as early as the 1850s, once with [k$^{w}$] and once with [x] in (3.8a). The other words in the table also suggest that some speakers might have occasionally used [k].

Conspicuously, there is no trace of [k] in most of the main Hupa documentation recorded from the 1880s onwards (Curtin 1888-1889; Goddard 1903, 1904; Golla 1970; Sapir and Golla 2001). Merriam recorded [k] in <Ning$^{l}$kah-ten> 'rich' (30/103) versus [x] in <Nŭ$^{l}$-hah-ten$^{l}$> 'chief' (30/122; Hupa *ningxa't'e:n* 'leader, boss, rich person'), and [k] in <Nă-kōs-lĕ> 'lost' (30/142; Hupa *nixohsle'* 'it got lost') and <'Hlŭ-'kahn> 'sweet' (30/119; Hupa *ɬixun*). In the lexical data recorded by Mary Woodward in the 1950s, words are occasionally flagged as "Redwood Creek dialect." One such example is <xaŋk'aW> (Redwood Creek, Woodward.001.007.0059) versus <xaŋxaW> (Hupa, Woodward.001.007.0060), both glossed 'bring it (liquid) out'.[19] These forms suggest that some people living in Hoopa Valley may have

[19] The Hupa verb is derived from a classificatory stem *-xa:wh* meaning 'handle a filled container' transcribed by Woodward with a stop for Redwood Creek. Woodward transcribed a glottalized stem-initial <k'>, but there is ample evidence that she was exuberant in her marking of glottalization, and that the key difference here is the stop

occasionally produced Redwood Creek [k] as late as the 1920s, and as late as the 1950s explicitly as an imitation of that dialect. That so few tokens can be found in Hupa's robust documentary record suggests that this pronunciation was rare at best.

Whatever the source of [k] in the 20th century, its salience as a marker of Redwood Creek speech was found in Woodward's documentation from the 1950s and continues to this day for present-day Hupa speaker Verdena Parker. The following is a comment she made concerning a Redwood Creek person she knew when Hupa was still actively spoken in Hoopa Valley:

> "...like the name of a deer in Redwood Creek is *na:wilxe'*, we say *k'iɬixun*, so some words were entirely different than our words... *na:wilke'*, it was more like a k sound for the Redwood Creek people... I used to listen to Mike Moon, he spoke it very well, he'd say *na:wilke'*." (Verdena Parker, VP-230308-10)

Mrs. Parker here uses a heavily aspirated [k] to imitate Mike Moon's pronunciation. This word is not documented in any other source, but the verb root is certainly analogous to a Hupa perfective verb stem *-xe'* 'finish, reach' in verbs such as *xwe:na:nehɬxe'* 'I caught up with him' (Golla 1996b:17). The word *na:wilke'* is most likely a nominalized passive formation meaning something like 'that which is chased' (with passive prefix *wi-* and regular change of voice/valence prefix from *ɬ-* to *l-*).

Mike Moon first appears in the Hoopa Valley census in 1916, a few years after the establishment of the Blue Lake Rancheria, where the surname Moon was common. His date of birth is given as 1890, so he probably arrived in Hoopa Valley as a teenager or young adult. Mrs. Parker was born in the mid-1930s, so Moon retained his Redwood Creek [k] until his acquaintance with her in the 1940s or 1950s. So whereas some people of Redwood Creek ancestry may have shifted to the Hupa pronunciation early on, as suggested by Powers (1877) and Goddard (1914b), others retained this older difference, probably because they were exposed to it in the Blue Lake region and only arrived in Hoopa Valley as adults. The influence of such late arrivals might account for some of the sporadic tokens of [k] found in the Hupa documentation as well.

## 4.4. [k] ~ [x]: Summary and Interpretation

Analysis of texts produced by Wife of Molasses with O'Haniel Bailey shows that the Whilkut dialect had a high degree of variation with respect to the velar obstruents [x] and [k]. Goddard's (1914b) description suggests that the older speaker tended toward [k] and the younger speaker toward [x], but both speakers apparently produced both variants. Similar kinds of variation are found in Whilkut wordlists collected in the Blue Lake region by Merriam; Reichard's documentation confirms that [k] was a general feature of Whilkut at Blue Lake and indicate that there were speakers who did not use [x], even when explicitly attempting to imitate a Hupa pronunciation. The [k] variant is not found in the speech of Chilula speakers who arrived in Hoopa Valley as adults and is only sporadically attested in the documentation of Hupa after the resettlement of Redwood Creek populations in Hoopa Valley in the mid-19th century; some recorded tokens are plausibly interpreted as being due to continuing immigration of Whilkut people in the 20th century. The virtual absence of [k] in documentation collected in

---

versus fricative. Notice too that the prefix *xa:-* 'up from beneath, out of the ground' has a fricative [x] in both forms.

Hoopa Valley confirms Goddard's (1914b) observation that [k] ~ [x] dialect variation was a target for leveling in the reservation setting. This, along with the apparent leveling of other dialect differences reported to have existed (e.g., the South Fork variety), is consistent with the expectations of the three-stage model of Kerswill and Trudgill (2005) that koineization is the typical outcome of dialect contact.

As discussed in §3 of chapter 1, when koineization occurs one of the phenomena demanding explanation is the direction of concomitant structural leveling: why some variants survive in the resulting koineized dialect but others do not. In the present case, where Hupa [x] survived and Whilkut [k] was eliminated, both demographic and indexical approaches to dialect contact and variation make the same prediction. The lower Trinity Hupa were numerically superior to other groups on the reservation, so their variant might have had an advantage by sheer force of numbers, as predicted by a majority-rule demographic theory of the kind advocated by Trudgill (2004) (albeit restricted to tabula rasa scenarios, which does not obtain in the present case). They also doubtless had more secure primary claims to resources on the reservation, which was established in their aboriginal territory, than immigrants from other dialect groups did, allowing for a materialist interpretation along the lines of Hill (2001) whereby the Hupa were in a position to adopt a localist stance with respect to variation involving [k] and [x]. Moreover, since the Whilkut were apparently a socially stigmatized group even in pre-contact times, indexical accounts invoking speakers' agency in deploying linguistic resources at their disposal would also predict [k] to be at a disadvantage in a mixed dialect setting. Thus, the outcomes of contact in this case does not help discriminate between the various theories.

From the perspective of the theory of dialect contact outlined in chapter 1 (Kerswill and Trudgill 2005), it is also necessary to explain deviations from the typical timeline of koineization spanning three generations. Interestingly, the extreme variation encountered in the speech of Wife of Molasses and Whilkut people living at Blue Lake, both across speakers and lexical items, is characteristic of Kerswill and Trudgill's (2005) Stage II speakers in dialect contact situations – that is, the first generation of speakers born in new, mixed dialect communities. Such variation would probably be expected if it were found instead among people born in Hoopa Valley after the reservation was established, since even children of speakers who used the [k] variant would have had ample exposure to [x] among their peer groups and elsewhere in the community. What is surprising is the fact that Wife of Molasses, born well before the establishment of the reservation, and who may have lived in Hoopa Valley for only a brief time, also participated in this variation: in other documented cases of dialect contact, Stage I adults' speech patterns are relatively fixed and they typically don't absorb dialect features from groups they come into contact with. Her variation could be indicative of imperfect learning of Hupa [x] due to exposure to it in Hoopa Valley as an adult, but the fact that the same variation was also found at Blue Lake among Whilkut people who as far as is known never resided in Hoopa Valley makes this explanation unsatisfying.

In fact, the timeline of dialect leveling affecting velar obstruents in Hoopa Valley seems to have been accelerated by a full generation, at least for this parameter of variation: a Stage I speaker exhibits Stage II variation, and Stage II speakers – people of Redwood Creek ancestry raised in Hoopa Valley that Goddard (1914b) observed – have Stage III stability. This raises the intriguing possibility that the [k] ~ [x] variation in Whilkut predates resettlement of Redwood Creek populations in Hoopa Valley. From this perspective the variation encountered in the speech of Wife of Molasses was not a consequence of her contact with people in Hoopa Valley, but rather reflects the pre-contact status quo in Whilkut that was also found in Merri-

am's data from Blue Lake. This pre-existing variation would have favored rapid leveling, since even Whilkut adults who were resettled in Hoopa Valley would already have had some command of the [x] variant they encountered there. Leveling of velar obstruents in Hoopa Valley was not a competition between stable [x] in Hupa and stable [k] in Whilkut, but rather between stable [x] in Hupa and pre-existing [k] ~ [x] variation in Whilkut. This interpretation also raises interesting questions about the origins of the variation in Whilkut – whether it is due to assimilation to Hupa [x] from original Whilkut [k], or assimilation to [k] found in other California Athabaskan languages from original Hupa-Whilkut [x]. This issue is considered in the appendix to this chapter.

One final point worth highlighting is the continuity in the salience of the variation distinguishing Hupa proper from Redwood Creek dialects among people acquainted with Hoopa Valley: Goddard in the early 20th century, people Mary Woodward worked with in the mid-20th century, and Verdena Parker today. The [k] variant was and remains a marker (in the sense of Labov 1972) of Whilkut (today collapsed with Chilula as simply "Redwood Creek"). This is interesting because it shows that children raised in the reservation setting, even as the [k] variant became the target of leveling under koineization, acquired and faithfully transmitted at least some of the social-indexical values of the linguistic input they encountered. This point is discussed further in §6.

## 5. ni- ~ 'i- Variation

### 5.1. Overview

When asked about dialect differences in Hupa that she remembers, Verdena Parker, a present-day speaker of Hupa, usually offers a particular example: words starting with *'i-* ([ʔi-]) in her variety of Hupa from the southern end of Hoopa Valley, but *ni-* in other varieties. Examples of words she has identified in this regard, where her variety of Hupa has *'i-* versus *ni-* documented in the Hupa Language Dictionary (Golla 1996b), include the following:

| Hupa form | gloss | ref. |
|---|---|---|
| a. 'iɬtuq' ~ niɬtuq' | 'black oak' perhaps from 'between one another' | VP-230308-03 |
| b. 'iɬq'it dahsa'a:n ~ niɬq'it-dahsa'a:n | 'flint design (on baskets)' lit. 'on top of each other – it lies on top' | VP-100607-09 |
| c. 'iɬwah ~ niɬwah | 'separated from each other' | VP-040807-02 |
| d. 'iɬch'in'-who:wh ~ niɬch'in'-who:wh | 'kingsnake' lit. 'to each other – parallel' | VP-121105-8 |

Table 3.9: Examples of Hupa words with variation in word-initial *ni-* versus *'i-*

Mrs. Parker identifies the *ni-* variant as being characteristic both of Hupa as spoken "down the valley" (i.e., at its northern end) and of Redwood Creek dialects, a connection she attributes to

the number of Redwood Creek families who settled in that area.[20] This reflects a traditional division between the northern end of Hoopa Valley, centered on the village of Ta:k'imiłding, and the southern end, centered on the village of Me'dilding, that was important in many ceremonial contexts (Goddard 1903:12). For her part, Mrs. Parker never produces the *ni-* variant in texts or under elicitation (except when imitating other dialects).

As the glosses and literal translations in (3.9a)-(3.9d) show, many of the words implicated in this variation involve the reciprocal prefix *nił-*.[21] Also affected are the directional prefix *nina:-* 'rising up from lying down' and the nouns (probably synchronically unanalyzable) *nista:n* 'log', *niskin(g)* 'tall conifer, pine', and *niłtuq'* 'black oak', including compounds such as *'isking-mina:k'iwilda:l* 'hairy woodpecker' (lit. 'conifer – it keeps running around it') or *'ista:ng-q'eh-k'itiqowh* 'pine marten' (lit. 'log – along – it scampers').[22] Additional documented examples include the following:

| Hupa form | gloss | speaker | ref. |
|---|---|---|---|
| a. 'iłma'n-ê: | 'on both sides' | Emma Frank | Sapir and Golla (2001, 25.11) |
| b. niłma'n-e: | 'on both sides' | Emma Frank | Sapir and Golla (2001, 68.38) |
| c. 'ina:na'se'indil | 'again they get up' | Sam Brown | Sapir and Golla (2001, 2.66) |
| d. nina:na'se'indil | 'they get up again' | Sam Brown | Sapir and Golla (2001, 2.41) |
| e. is tan | 'log' | Wife of Molasses | Goddard (1914b, 333.17) |
| f. nis tan | 'log' | Wife of Molasses | Goddard (1914b, 312.10) |
| g. ʔɪskiŋ yá·W | 'xmas tree' | Stella Jarnaghan | Woodward.001.002.0014 |
| h. nɪskiŋ yaw | 'small fir/xmas tree' | Stella Jarnaghan | Woodward.001.002.0014 |

Table 3.10: Additional examples of Hupa words with variation in word-initial *ni-* versus *'i-*

The important thing to notice in these examples is that a given speaker might sometimes use the *ni-* variant, other times the *'i-* variant.[23]

Complicating the picture somewhat is a similar, perhaps related dimension of variation mainly affecting a word derived from the noun *nin'* 'ground' plus a postpositional element *-yiw*

[20] This historical scenario is confirmed by Goddard, who reports that "[a] number of the families living at the northern end of Hoopa Valley are of Chilula origin" (1914a:270).

[21] It is perhaps relevant that the morphophonology of this prefix is complex, the initial syllable *ni-* lost altogether in certain contexts (Golla 1970:212), as when preceding a vowel-initial prefix: *łe:na:ya'asow* 'they scrape them together' (Sapir and Golla 2001, 2.30).

[22] The relatively rare prefix *nint'a:-* 'turning around, turning back' is also affected, but there is only a single text example (Sapir and Golla 2001, 47.13) so it has been excluded from this analysis.

[23] That this ni- ~ 'i- variation is lexically conditioned is evident from the fact that many word-initial *ni-* sequences are not affected. This appears to be the case where *ni-* is a second person singular agreement prefix or a thematic conjunct prefix (i.e., one that is lexically-specified and in a relatively close association with the root), as in *nina:tse:* 'ahead of you' or *niłtung* 'damp' (Sapir and Golla 2001, 17.9 and 22.41). This is also implicit in Golla's footnotes to the texts in Sapir and Golla (2001), where particular morphemes are identified as participating in this variation but no general rule is formulated.

'under', translated as 'underground' and often found in its "light" form *ninyeh*.[24] The following forms for 'underground' are listed in the Hupa Language Dictionary (Golla 1996b) and/or the analytic lexicon accompanying the text collection in Sapir and Golla (2001): *yineh, 'ineh*, *nineh*, *ninyeh, ninyiw*. These forms have an additional variant *yi-* alongside *ni-* and *'i-*. This *yi-* variant is merely noted for the moment, but its importance will become clear in due course.

Given that the *ni-* variant is emblematic of different dialect groups for one of today's speakers, it is of interest to consider its status in earlier documentation of the language, which will be the focus of §5.2 - §5.6. Two key points emerge from the data. First, while there may be some connection between the *ni-* variant and the Chilula (lower Redwood Creek) dialect and the northern end of Hoopa Valley, the *ni-* ~ *'i-* data can instead be plausibly analyzed as stylistic variation and not due to dialect mixing per se. Second, whatever its source, this variation continued well into the 20th century and was not a consistent target of leveling. This raises the question of why the two parameters of variation - [k] ~ [x] versus *ni-* ~ *'i-* have different outcomes. It is unclear what predictions are made by theories of dialect contact with respect to variation that is not attributable to dialect divisions when a new community is established, whether or not it is expected that such variation should persist when koineization occurs. This issue is revisited in §6.

## 5.2. 19th Century Documentation

The earliest documentation of Hupa does not contain many tokens of the items implicated in the *ni-* ~ *'i-* variation, mainly because their meanings are not among the basic vocabulary items that researchers in that period were likely to inquire about. The following list of tokens is probably exhaustive:

[24] The selection of heavy vs. light stems is partially predictable in verbs based on morphological context (Sapir and Golla 2001:823-824), but the difference in other lexical categories is among of a residue of cases that are not well-understood.

| Hupa form | gloss | ref. |
|---|---|---|
| a. <nis-kingh> | 'pine tree' | Azpell (1870) |
| b. <nis-kín> | 'tree' | Powers (1870s) |
| c. <iskĭng mĭkyangai> | 'tree limb'<br>lit. 'pine tree its arm' | Curtin (1888-1889) |
| d. <nĭskǵng holen> | 'forest'<br>lit. 'pine trees plentiful' | Curtin (1888-1889) |
| e. <nisking me na kwĺltal> | 'woodchuck' | Curtin (1888-1889) |
| f. <niskimetau> → <iskimetau> | 'owl'<br>lit. 'among the trees' | Curtin (1888-1889) |
| g. <nistan> → <istan> | 'log' | Curtin (1888-1889) |
| h. <istänˈ kaiq kitûko> | [unglossed animal sp.]<br>cf. 'ista:nq'eh-k'itiqowh<br>'pine marten' | Curtin (1888-1889) |
| i. <nĭn> | 'ground' | Curtin (1888-1889) |
| j. <ineh> | 'in the ground' | Curtin (1888-1889) |

Table 3.11: *ni-* ~ *'i-* variation in 19th century Hupa sources

(3.11a) and (3.11b) are of interest because they demonstrate that the ni-variant was already found in a variety identified as 'Hupa' in one of the candidate words ('pine tree') as early as 1870. The words from Curtin's extensive vocabulary schedule are more interesting still: the same words 'pine,' 'log' and 'ground' are found with both the ni- variant and the 'i- variant (3.11c-f), (3.11g-h), and (3.11i-j). In (3.11f) and (3.11g) the ni-variant is explicitly overwritten with the 'i-variant. The latter correction is in a different color of ink and may have been made by one speaker when Curtin checked a form obtained from another, but the former may have been made by the same speaker in a single elicitation session. Overall, Table 3.11 shows that both variants were in use in Hoopa Valley in the late 1880s, perhaps by the same individuals, and they were probably subject to metalinguistic evaluation since some speakers felt the need to correct themselves and/or each other in this regard.

## 5.3. Goddard (1900-1910)

The following tables show the distribution of the ni- and 'i- variants in Hupa and Redwood Creek texts collected by Goddard in the first decade of the 20th century (Goddard 1904, 1914b), organized first by speaker (Table 3.12) and then by dialect (Table 3.13).[25] Speakers who are known to have had an affiliation with Chilula or Whilkut are marked with an asterisk:

[25] In Table 3.13, it is important to bear in mind that the "Upper Redwood Creek" row is just Wife of Molasses and/or O'Haniel Bailey, and the "Lower Redwood Creek" row is Tom and Dan Hill and McCann.

| | (n)ista:n 'log' | | (n)iɬtuq 'oak' | | (n)isking 'pine' | | (n)iɬ- 'reciprocal' | | (n)ina:- 'up' | | total | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speaker | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- |
| Emma Lewis | - | - | - | - | - | - | 3 | 2 | 2 | 5 | 5 | 7 |
| Lillie Hostler | - | - | 1 | 0 | - | - | 0 | 3 | - | - | 1 | 3 |
| Oscar Brown | - | - | - | - | 1 | 0 | - | - | 0 | 2 | 1 | 2 |
| Robinson Shoemaker | - | - | - | - | - | - | 0 | 1 | - | - | 0 | 1 |
| Wife of McCann | - | - | - | - | - | - | 1 | 0 | 0 | 1 | 1 | 1 |
| Emma Dusky (Frank) | 1 | 0 | - | - | - | - | - | - | 1 | 0 | 2 | 0 |
| Henry Hostler | - | - | 3 | 0 | - | - | - | - | - | - | 3 | 0 |
| Mary Marshall | - | - | - | - | 1 | 0 | - | - | - | - | 1 | 0 |
| *McCann[26] | 1 | 0 | - | - | - | - | 3 | 0 | - | - | 4 | 0 |
| *Tom/Dan Hill | 1 | 0 | - | - | - | - | 3 | 0 | 1 | 0 | 5 | 0 |
| *Wife of Molasses/ O'Haniel Bailey | 3 | 3 | - | - | 2 | 0 | 0 | 6 | 0 | 5 | 5 | 14 |
| *Total* | 6 | 3 | 4 | 0 | 4 | 0 | 10 | 12 | 4 | 13 | 28 | 28 |

Table 3.12: Distribution of *ni-* and *'i-* in Goddard (1904, 1914b), by speaker

| | (n)ista:n 'log' | | (n)iɬtuq 'oak' | | (n)isking 'pine' | | (n)iɬ- 'reciprocal' | | (n)ina:- 'up' | | total | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dialect | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- |
| Upper Redwood Creek | 3 | 3 | - | - | 2 | 0 | 0 | 6 | 0 | 5 | 5 | 14 |
| Lower Redwood Creek | 2 | 0 | - | - | - | - | 6 | 0 | 1 | 0 | 9 | 0 |
| Hupa | 1 | 0 | 4 | 0 | 2 | 0 | 4 | 6 | 3 | 8 | 14 | 14 |
| *Total* | 6 | 3 | 4 | 0 | 4 | 0 | 10 | 12 | 4 | 13 | 28 | 28 |

Table 3.13: Distribution of *ni-* and *'i-* in Goddard (1904, 1914b), by dialect

While there are too few tokens of most items to say much with certainty, it is clear that many speakers used both variants. This is sometimes the case even for different tokens of the same

[26] As noted above, Goddard (1914a:270) observed that many people from the northern end of Hoopa Valley had Chilula (lower Redwood Creek) origins. He specifically mentions McCann in this regard, although elsewhere he states that McCann "was born and had always lived at the northern end of the valley [Hoopa Valley]" (1904:198). McCann was born prior to colonization, suggesting that some migration of Chilula families into Hoopa Valley occurred in pre-contact times.

word. For example, Goddard recorded five tokens of Emma Lewis saying <in-na-is-dûk-ka> 'he/she got up' or similar with the *'i-* variant, and two tokens of <nin-na-is-dûk-kai> with the *ni-* variant. The aggregate distribution of the two variants throughout the text collections is even, with 28 tokens of each, although fully half of the tokens of *'i-* were produced by the Whilkut speaker Wife of Molasses or her proxy O'Haniel Bailey, with a definite preference for the *'i-* variant in the reciprocal and directional prefixes. Although the number of tokens is small, there seems to be the opposite tendency for Chilula, with *ni-* preferred in the reciprocal. Taken together these data show that (a) both *ni-* and *'i-* were commonly used both in Hupa proper and in Redwood Creek dialects in the first decade of the 20th century, (b) many speakers used both variants, sometimes in different tokens of the same word and (c) the *ni-* variant may have been more closely associated with lower Redwood Creek, the *'i-* variant with upper Redwood Creek.

## 5.4. Merriam (1910-1930)

Goddard's material can be compared with the Hupa and Whilkut vocabularies collected by Merriam in the second and third decades of the 20th century:

| Hupa form | gloss | ref. |
|---|---|---|
| a. <Is-ke-ung><br><Is$^{l}$-king> | 'Douglas fir' | 52/23 |
| b. <Is$^{l}$-king> | 'tree' | 52/32 |
| c. <Mis-king> | 'Douglas fir' | 52/23 |
| d. <Is-king$^{l}$-tsah-mah> | 'goshawk'[27] | 52/11 |
| e. <Is$^{l}$-ke met$^{l}$-to> | 'great horned owl'[28] | 52/11 |
| f. <Is-king ke-til$^{l}$-tsil ke-li-kah> | 'red-breasted sapsucker'[29] | 52/13 |
| g. <Es-tahn$^{l}$ ki-kit$^{l}$-to-ko> | 'fisher'[30] (pine marten) | 52/7 |
| h. <Il$^{s}$-chung-who><br><Is-chin-who$^{l}$> | 'king snake'[31]<br>'red-ringed snake' | 52/20 |
| i. <Is-wah-kew wit$^{l}$-li> | 'trade'[32] | 30/141 |
| j. <'Is-wa kew-wit$^{l}$-li> | 'Trade!' | 30/154 |
| k. <Ā-nahn$^{l}$-tă-ki> | 'get up!' | 30/163 |

Table 3.14: Targets of *ni-* ~ *'i-* variation in Hupa vocabularies collected by Merriam

[27] Merriam's note: "lives in spruce forest altogether"

[28] Merriam's note: "forest bird"

[29] Merriam's glosses: <Is-king> 'tree', <ke-til$^{l}$-tsil> 'dries (takes sap out)', <ke-li-kah> 'female'. Cf. *king-k'idiłtsay'* (lit. 'tree – it dries up', Golla 1996b); *nisking-mina:t-ch'iwiltong'il* (lit. 'conifer – around it – he jumps along', Golla 1996b).

[30] Cf. *'ista:ngq'eh-k'itiqowh* (Golla 1996b).

[31] Merriam's note: "ringed together." Cf. *nił̓ch'ing'-who:wh* (lit. 'to each other – parallel', Golla 1996b).

[32] Cf. *'iłwa'k'iłiwh* 'store' (lit. 'where they sell things to each other', Golla 1996b).

| Whilkut form | | gloss | ref. |
|---|---|---|---|
| a. | <Nisking> | 'tree' | 52/64 |
| b. | <Is-ching> | 'Douglas fir'[33] | 52/54 |
| c. | <Nis-king minnah kool tsil> | 'nuthatch'[34] | 52/44 |
| d. | <Is-tahn$^{l}$-ket> ~ <Is-tah$^{l}$-met-tik-kahn> | 'Bald Mountain'[35] | 30/250 |
| e. | <Es-mahn$^{l}$-chin suk-ki> | 'bald eagle' lit. 'white both ends' | 52/42 |
| f. | <Esh$^{l}$-choon-who$^{ch}$> | 'kingsnake' | 52/51 |
| g. | <Es-tŏk$^{l}$> | 'black oak' | 52/55 |

Table 3.15: Targets of *ni-* ~ *'i-* variation in Whilkut vocabularies collected by Merriam

Significant in Merriam's data is the fact that there are so few examples of the *ni-* variant: aside from the aberrant word-initial <M> for 'Douglas fir' (perhaps a copying error for *ni-*) in (3.14c), only *'i-* occurs in the Hupa list. In Whilkut, *ni-* is similarly found only in *nisking* (3.15a) and the related compound meaning 'nuthatch' in (3.15c). These data suggest a more *'i-*oriented language than was found by Goddard, in both Hupa proper and Whilkut, similar to Hupa as spoken today by Verdena Parker. It may also be significant that many of Merriam's Hupa forms were obtained not in Hoopa Valley proper, but at Burnt Ranch, a small settlement on the South Fork of the Trinity, a region adjacent to the southern end of Hoopa Valley where Mrs. Parker's family is from.

## 5.5. Sapir (1927)

The *ni-* ~ *'i-* variable has the following distribution in the Hupa texts transcribed by Sapir in 1927 (Sapir and Golla 2001), the majority of which came from speakers Sam Brown and Emma Frank:

[33] <ch> is unexpected here. This palatalization of front-velar [k̲] is characteristic of some other California Athabaskan varieties, including Nongatl, spoken directly south of Hupa, as in <tciñ> 'tree' (Hupa *kin*), <-tcō> 'big' (Hupa *-kyoh*) (Kroeber 1967:270). See also Whilkut speaker Molly's rendition of Hupa 'arm' in Table 3.7 above.

[34] Cf. *nisking-mina:k'iwiltsil* 'nuthatch' (perhaps lit. 'conifer – it moves around it squatting', Golla 1996b).

[35] Derivation uncertain, but the first element is most likely (n)ista:n. Deletion (or assimilation) of final n before m in <Is-tah$^{l}$-met-tik-kahn> is phonologically regular (Golla 1970:55).

| | (n)ista:n 'log' | | (n)iɬtuq 'oak' | | (n)isking 'pine' | | (n)iɬ- 'reciprocal' | | (n)ina:- 'up' | | total | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speaker | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- |
| Sam Brown | 2 | 1 | 2 | 0 | 3 | 0 | 41 | 0 | 3 | 29 | 51 | 30 |
| Emma Frank (née Dusky) | 0 | 5 | 0 | 4 | 0 | 7 | 4 | 7 | 0 | 2 | 4 | 25 |
| 4 others | 0 | 1 | 2 | 0 | - | - | 8 | 0 | 0 | 5 | 10 | 6 |
| Total | 2 | 7 | 4 | 4 | 3 | 7 | 53 | 7 | 3 | 36 | 65 | 61 |

Table 3.16: Distribution of *ni-* and *'i-* in Sapir and Golla (2001)

As in the Goddard corpus, the aggregate distribution of the two variants is roughly equal across speakers and items: 65 *ni-* vs. 61 *'i-*, and some speakers continue to use both variants in some words (e.g., Sam Brown's *nista:n*, *'ista:n*). This corpus contains enough tokens of some elements for interesting patterns to emerge. In particular, on a by-item basis we see that directional *(n)ina:-* 'up' and reciprocal *(n)iɬ-* have polarized: *(n)ina:-* is almost exclusively *'i-*, a trend that probably was already underway in at the turn of the 20th century (Table 3.12 above). The reciprocal is almost exclusively *ni-*, thanks mainly to Sam Brown; other items seem more evenly distributed, but there are too few tokens to be certain.

Also interesting is the comparison of data from Sam Brown and Emma (Dusky) Frank. While Brown uses *ni-* exclusively in the reciprocal, Frank uses *'i-* more frequently for the same item. Little is known about Emma Frank's background other than the fact that she was the last fully trained Indian doctor in Hoopa Valley (Sapir and Golla 2001:27) and that she lived at Redwood Creek at some point in her life (Sapir and Golla 2001:227); in his work with her in 1901-1902, Goddard had described her as "very conservative" (1904:324). Her preference for *'i-* over *ni-* in the texts transcribed by Sapir, even in contexts where other Hupa speakers gravitate exclusively towards *ni-*, makes her similar to both Wife of Molasses (with O'Haniel Bailey) in Goddard (1914b) and contemporary speaker Verdena Parker.

In fact, Emma Frank was even more innovative in this regard. Recall from the discussion above that the word glossed as 'underground', derived from the freestanding noun *nin'* 'ground', has three documented variants in the first syllable: *yi-*, *'i-*, and *ni-*. In general the variation does not extend to other words derived from *nin'*. However, unique among all documented speakers, including Mrs. Parker, Emma Frank sometimes uses the *'i-* variant with the word *ninis'a:n* 'world', which Golla (1996b:31) suggests this might be derived from *nin'-sa'a:n* 'ground which lies'. In the Sapir corpus, she has seven tokens with *ni-*, matching other speakers, and eight tokens with *'i-*. Even more interestingly, Emma Frank, then known as Emma Dusky, used this same word fourteen times in the two short texts she provided to Goddard (1904), only one fewer than found in the Sapir corpus. In all of these, she has *ni-*, not *'i-*. In the Goddard corpus she also has only *ni-* in the single tokens of *nista:n* 'log' and *nina:-* 'up' (<nin-na-is-dûk-kai-hit>, 'when he got up', Goddard 1914:346.18) and in three tokens of <nin-mū-win-na-kis-ten> 'ground around lies' (the name of a mythical being that became the

earth), also built on the *nin'* root. Taken as a whole, it appears that Emma Frank, in the quarter century between her interviews with Goddard and Sapir, had become a leader in using the *'i-* variant, extending it even to a word where no other speakers had done so.[36] She apparently did this very late in life, since she was already 40 years old when she worked with Goddard.

## 5.6. 1950s-1960s

The final point of comparison for variation involving *'i-* and *ni-* is the mid-20th century documentation of Bright (1950-52), Woodward (1953), and Golla (1963, 1984). Tokens scattered around various sources are few enough that they are collapsed in Table 3.17; particular forms of interest are given in Table 3.18:

| | (n)ista:n 'log' | | (n)iɬtuq 'oak' | | (n)isking 'pine' | | (n)iɬ- 'reciprocal' | | (n)ina:- 'up' | | total | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speaker | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- | ni- | 'i- |
| (all) | 5 | 1 | 3 | 1 | 5 | 2 | 3 | 6 | 0 | 4 | 16 | 14 |

Table 3.17: Distribution of *ni-* and *'i-* in mid-20th century sources

| form | gloss | speaker & ref. |
|---|---|---|
| a. ɪɬ.dɛʔᵋsɪndɪl | '(they) meet' (recip.) | Stella Jarnaghan<br>Woodward.002.001, IV.3 |
| b. nɪɬčiŋ' | 'to gather' (recip.) | Stella Jarnaghan<br>Woodward.001.005.0030 |
| c. ʔɪskiŋ ~ nɪskiŋ | 'pine tree' | Stella Jarnaghan<br>Woodward.001.001.0048 |
| d. nɪɬmáˈntˈšɪnˀ méˈdaʔʌy →<br>ʔɪɬmáˈntˈšɪnˀ méˈdaʔʌy | 'snake sp.'<br>lit. 'head on both sides' | Ned Jackson<br>Golla (1963:1.53b) |

Table 3.18: Variation involving *ni-* and *'i-* in mid-20th century sources

The data in Table 3.17 show that there remained variation involving *ni-* and *'i-* in the Hupa-speaking community in the 1950s and 1960s. Nearly all of the target items occur with at least one token of *ni-* and one of *'i-*. The exception is the directional prefix *(n)ina:-* 'up', which occurs only with the *'i-* variant; while there is only a small number of tokens, this appears to continue the trend already found in Sapir's documentation from the 1920s (summarized in Table 3.16). There also seems to be a trend towards *ni-* in the nouns 'log', 'oak', and 'pine',

[36] Golla (p.c.) has suggested that Goddard might have normalized his text material, masking variation in this regard. This is certainly possible, but if he did normalize he did so inconsistently across the entire collection, as the variation in Table 3.12 above demonstrates.

but again there are too few tokens to say much with certainty. Table 3.18 shows that intra-speaker variation was still found as well: Stella Jarnaghan produced both variants for the reciprocal (3.18a-b). She did this for 'pine' as well (3.18c), and explicitly accepted both variants according to a note on Woodward's file slip: "corrected to niskiŋ; infmt says can use either." In (3.18d) *'i-* overwrites original *ni-* in [ʔɪɬmáˑnt'šɪnʔ méˑdaʔʌy] 'snake sp.', the same direction of correction encountered in Curtin's materials in the 1880s.[37]

## 5.7. ni- ~ 'i-: Summary

Unlike variation involving [k] vs. [x] discussed in §4, there is no clear evidence that *ni-* ~ *'i-* variation encountered in the documentation of Hupa proper was grounded in a well-defined dialect difference. The *'i-* variant seems to have been preferred among Whilkut speakers from upper Redwood Creek and Blue Lake (i.e., the southern end of Redwood Creek and around Mad River), and in Hupa as spoken in the vicinity of the southern end of Hoopa Valley (as documented by Merriam and in the speech of Verdena Parker). In the small number of tokens obtained from Chilula speakers associated with lower Redwood Creek, only the *ni-* variant is found. However, the *ni-* variant appears in the earliest documentation of Hupa proper, and both variants, *ni-* and *'i-*, were already very much at play in Hoopa Valley in the late 19th century. The fact that the variation occurs in the first documentation robust enough to contain tokens of the items implicated in the variation suggests (but of course does not require) that it might simply have been pre-existing variation and not due to dialect mixing that occurred with the establishment of the reservation in Hoopa Valley.

This variation continued well into the mid-20th century, although by the 1920s all speakers had begun to converge on *'i-* for the directional prefix *(n)ina:-*. At least one speaker, Emma Frank, went further and converged on *'i-* in nearly all other target lexical items, and moreover extended this variant to new vocabulary late in life. Contemporary speaker Verdena Parker is also one of the speakers favoring *'i-* : although Mrs. Parker does not extend *'i-* to new words like *ninis'a:n* 'world' as Frank did, she is more consistent than Frank in not using the *ni-* variant at all.[38] Other speakers, however, clearly retained *ni-*, especially in nouns denoting 'conifer/pine', 'black oak', and 'log' and related compounds, sometimes variably alongside *'i-* for a given speaker. Thus, while it is possible to discern a uniform diachronic trajectory among some speakers, among other speakers, and for particular lexical items, the outcomes are more diverse.

The data presented in this section also offer clues about how the variation may have been perceived in the Hupa-speaking community. The direction of correction in field notes is from *ni-* to *'i-* in both the 1880s (twice) and the 1960s. One speaker in the 1950s made the correction in the opposite direction, from *'i-* to *ni-*, but in so doing explicitly accepted both variants. The status of *'i-* as the normatively correct way to speak Hupa is implicit in contemporary speaker Verdena Parker's estimation of *ni-* as belonging to other dialect groups: this probably is based on historical reality insofar as the *ni-* variant appears to have been preferred by Chilula speak-

[37] Bright recorded the same word from Sam Brown as *niɬmanc'iŋ metaʔy* 'snake with head on each end' (BrightW.003.001.0021); *ni-* in the reciprocal is consistent with Sapir's documentation of the same speaker.

[38] Unique among all documented speakers, Mrs. Parker has for the directional *(n)ina:-* not *'i-* but *yi-*. The addition of the *yi-* variant resembles the variation found in *yineh/'ineh/nineh* 'underground'.

ers, although the data from that dialect are scanty.[39] In changing from *ni-* to *'i-* between 1902 and 1927, Emma Frank might also have been gravitating toward a normative standard. These data suggest that there was a measure of transgenerational continuity not only in retaining this as a parameter of variation, but also in the metalinguistic evaluation associated with it – even among the last generation of people to acquire native fluency in Hupa.

## 6. Discussion

This chapter has traced in detail the diachronic trajectory of two linguistic variables in Hupa and related dialects from the mid-19th to the mid-20th century. Variation involving the velar obstruents [k] and [x] observed in the Whilkut dialect was the target of rapid leveling in the reservation setting. Variation involving word-initial *ni-* versus *'i-* persisted well into the 20th century, although some speakers showed a tendency to converge on the *'i-* variant. The obvious question is what factors influenced the different outcomes in the two cases. The possibility that the *ni-* ~ *'i-* variation, within Hupa proper at least, was stylistic or otherwise not an unambiguous index of particular reference groups is intriguing in this regard, and points out that it is not entirely clear what outcome theories of dialect contact predict for variation in mixed dialect communities that is not strictly dialectal. Since the locus of koineization in studies such as Kerswill and Williams (2000, 2005) and Trudgill (2008a,b) is placed squarely among Stage II and Stage III children who are exposed to a wide range of variation in their environment, the question can be put thus: how did children raised in the reservation community know that one source of variation they encountered could be eliminated, but the other could be largely maintained?

A number of factors can be invoked to explain the different outcomes for the two linguistic variables. First, as discussed in §4, it is likely that variation involving velar obstruents in Whilkut was not due to dialect contact in the reservation setting, but rather was already present in pre-contact times. This would have accelerated leveling, since even adult Stage I Whilkut speakers would have arrived in Hoopa Valley with [x] already in their linguistic repertoire (in the sense of Matras 2009). Second, if the relative frequency of variants in texts can be taken as an indication of their proportional distribution across a population, the totals for *ni-* versus *'i-* in Tables 3.13, 3.16, and 3.17 suggest that neither was in the majority. Whilkut [k], by contrast, was a minority variant in Hoopa Valley. According to theories of majority-rule demographic determinism such as Trudgill (2004, 2008a), [k] would be vulnerable as a target for leveling, whereas there is no expectation that there would be leveling of *ni-* and *'i-* since neither variant had a numerical advantage.

Third, differences in structural complexity might also have played some role. While the correspondence between [k] and [x] appears to have applied throughout the lexicon and in all phonological positions, the *ni-* ~ *'i-* variation was lexically restricted: each generation of learners had to memorize which words participated on an item-by-item basis. This complexity may have opened the door to speakers generalizing one variant or another to different degrees and in different directions. Some, like Sam Brown, seem to have gravitated towards *ni-* in the

[39] There could instead be an interesting indexical inversion at play: as some speakers gravitated toward one of the variants (*'i-*), the other variant (*ni-*) came to be associated with historically stigmatized groups living in Hoopa Valley. The typical progression in dialect contact situations is in the opposite direction, from regionalism to stigmatized variant under so-called "reallocation" whereby erstwhile dialect variation in a mixed dialect community takes on new sociolinguistic functions (Trudgill 1986).

reciprocal prefix but *'i-* in the directional prefix; others, like Emma Frank, apparently generalized the *'i-* variant in most candidate items and even extended it to new ones. Structural complexity, then, is likely to have been a factor in the fact that variation involving *ni-* and *'i-* persisted into the 20th century. But note that one of the properties of internally-motivated linguistic changes – changes within a speech community that are not due to contact with other dialect or language groups – is the fact that this kind of structural complexity can be highly stable diachronically (Labov 2007:353-356). From this perspective, the maintenance of *ni-* and *'i-* as a linguistic variable is part of normal transgenerational transmission of linguistic knowledge.

Less clear in the present context is the relevance of the indexical status of competing values associated with [k] versus [x] on the one hand, with *ni-* versus *'i-* on the other. As noted, the metalinguistic evaluation of particular variants appears to have been transmitted even to the last generation of first-language speakers of Hupa, including Verdena Parker, with [k] as a marker of Redwood Creek speech and *'i-* as normatively preferred (and also an index of regional identity, whatever its historical origins). Indexical accounts of variation in dialect contact situations could be developed that exploit the fact that whereas *ni-* ~ *'i-* variation might not have been strongly tied to any particular reference group in pre-contact times, [k] ~ [x] certainly was: variants with strong indexical associations of this sort might be more susceptible to leveling pressure than are non-indexical ones (even if there is still a normative judgment associated with them). Nonetheless, while the data presented here are not incompatible with indexical accounts, alternative explanations for the difference are equally plausible.

A final point concerns the analysis presented in §4.4 of [k] ~ [x] variation in Whilkut as having origins in pre-contact times. The appendix to this chapter considers the chronology of this variation, whether Whilkut originally had [k] and acquired [x] due to diffusion of an innovation from Hupa proper, or whether Whilkut originally had [x] and acquired [k] due to contact with the neighboring Eel River dialect Nongatl. Arguments are presented that the latter interpretation may be more nearly correct, but either scenario is problematic for a theory of linguistic diffusion that relies only on population size and density as a predictor for the direction of changes, e.g., Trudgill (1983) for urban-to-rural diffusion in industrial societies or Trudgill (2004) for tabula rasa contact.

The reason for this is that the Whilkut were among the most populous groups in the region according to some estimates. Pre-contact population estimates for the various Hupa dialect groups vary widely. The most conservative is that of Kroeber (1925), also adopted by Wallace (1978), which puts the number of Hupa in Hoopa Valley at roughly 1000, with 500 each of Chilula and Whilkut. Cook (1956) estimated the Hupa proper to be twice as numerous, at 2000, and the Chilula-Whilkut combined at roughly 2100. Baumhoff (1958:223) reached an intermediate figure for the Hupa proper of nearly 1500, and an even greater number of 2600 for the Redwood Creek groups, approximately two-thirds of which (roughly 1700) are the groups identified as Whilkut in the present study. As Baumhoff notes, these discrepancies are based on different assumptions about the extent of each group's territory, the number of permanent settlements at the time of contact, the number of households per settlement, the number of people per household, and in the case of Baumhoff (1958) an estimate of the ecological carrying capacity of each group's territory.

Baumhoff's figures for the Redwood Creek groups may be somewhat inflated, but even if a 10% margin of error is allowed the Whilkut groups of Upper Redwood Creek and Mad River

were probably at least as numerous as the Hupa proper.[40] The population density of the Hupa proper and Redwood Creek groups were moreover estimated by Baumhoff to be 3.48 and 5.61 people per square mile respectively. Again, even allowing for some reasonable margin of error in Baumhoff's study, it is unlikely that the population density of the Hupa was greater than the population density of the Redwood Creek groups. If the chronology of variation in Whilkut involves a recent adoption of Hupa [x], a theory of diffusion based on relative population density either makes no prediction (if the densities were approximately equal), or the wrong prediction (if the Whilkut density was as high as estimated by Baumhoff).

If the true chronology of variation in Whilkut instead involves cross-linguistic diffusion of [k] from Nongatl, the Eel River dialect whose territory was adjacent to the Whilkut, a demographic theory of the direction of the change fares no better. The Nongatl-speaking groups are estimated by Baumhoff to have had a pre-contact population of 2,325 people in an area of 855 square miles, or 2.72 people per square mile. This is the lowest population density of any of the California Athabaskan groups that Baumhoff studied and much lower than his estimate for the Redwood Creek groups. Here again, a strictly demographic account of the direction of diffusion seems to make the wrong prediction: if anything, Redwood Creek groups should have been the source of innovations leading to variation in adjacent Athabaskan varieties rather than vice-versa. A more nuanced consideration of Baumhoff's population estimates, especially one that does a better job of teasing apart the Whilkut and Chilula populations that Baumhoff conflated, is needed in order for there to be certainty on this point, but taking his data at face value suggests that something more than asymmetries in absolute population and population densities needs to be said in order explain the observed patterns of variation.

## Appendix: Chronology of [k] ~ [x] Variation in Whilkut

As noted in §4.4, dialect leveling targeting Whilkut [k] in favor of Hupa [x] was achieved very rapidly, essentially in just one generation. It was suggested that this was because variation in Whilkut as a precursor to leveling in Hoopa Valley reflected a pre-contact state of affairs. This invites speculation about the origins of this variation, whether Whilkut uniformly had [k] and later assimilated to Hupa [x], or whether it uniformly had [x] and later assimilated to [k], as found in Eel River dialects, Mattole, and Kato. This question is crucial, since these scenarios imply two very different historical developments: in one, it involves the incomplete spread of an innovation within a single language; in the other, it involves the diffusion of a phonological category across a sharper linguistic boundary. This, in turn, implies very different sociolinguistic relationships immediately prior to colonization.

The first view, that [k] in Whilkut is archaic and [x] is a recent innovation due to Hupa influence is supported to some extent by the report of Powers (1877), who suggested that many groups in the region, apparently in the pre-contact past, "lost the distinctiveness of their respective languages and customs" (1877:87) at the hands of the Hupa, whom he considered the politically dominant group in the region:

[40] Baumhoff's calculation of carrying capacity is based on land area and number of miles of watercourses with salmon. He estimated that Redwood Creek had roughly twice as many miles of fishable watercourses as the portions of the lower Trinity occupied by the Hupa. Andrew Garrett (p.c.) has pointed out, however, that Redwood Creek is a much narrower watercourse with fewer accessible fishing points than the Trinity, so Baumhoff's estimate for the population supported by Redwood Creek may be inflated.

> "...they [the Hupa] compel all their tributaries to this day, to the number of about a half-dozen, to speak Hupâ in communication with them. Although they originally occupied only about twenty miles of the Lower Trinity, their authority was eventually acknowledged about sixty miles along that stream, on South Fork, on New River, on Redwood Creek, on a good portion of Mad River and Van Deusen's Fork; and there is good reason to believe that their name was scarcely less dreaded on the Lower Eel River, if they did not actually saddle the tribes of that valley with their idiom." (Powers 1877:72)[41]

Powers' description must be used with caution, since he was given to the journalistic excesses typical of the period, and Goddard (1903:3) considered him to be "on the whole misleading." Goddard (1903:7) specifically disagreed with Powers' assessment that the Whilkut dialect had been discarded in favor of Hupa, a judgment based on what Goddard perceived to be Whilkut's enduring affinity with the adjacent California Athabaskan languages to the south among speakers still living in his time. Powers' assessment of the influence of the Hupa is also at odds with the opinion of C. Hart Merriam, who considered the Redwood Creek groups collectively to be "one of the dominant tribes of Humboldt County in northwestern California" (cited in Baumhoff 1958:202).

Some linguistic arguments can be mustered in support of the view that [k]-variant in Whilkut is relatively recent compared with the [x]-variant. Bearing in mind that the categorical merger of Proto-Athabaskan *q and *x̣ is found throughout California Athabaskan, the existence of the outcome /x/ in Hupa and /k/ elsewhere suggests one of the following scenarios:

1. the merger occurred once prior to the differentiation of Proto-California Athabaskan, either
    a. as /k/, with subsequent shift of /k/ to /x/ in Hupa (i.e., the outcome /x/ is the most recent innovation)
    b. as /x/, with subsequent shift of /x/ to /k/ everywhere except Hupa (i.e., the outcome /k/ is the most recent innovation)
2. the merger occurred independently in Hupa (as /x/) and elsewhere (as /k/) (i.e., there is no connection between the merger in Hupa and elsewhere)

Scenario (2) can be ruled out on the grounds that an analysis in which the merger is an independent phenomenon is dispreferred, especially in light of the evidence reported in chapter 2 that it is clade-defining not only in California Athabaskan but in Pacific Coast Athabaskan as a whole. It is therefore necessary to distinguish scenarios (1a) and (1b). In scenario (1a), it is possible that the stop found in Redwood Creek is conservative: an innovative sound change of Proto-Athabaskan **q, **x̣ > Proto-California *k > Hupa /x/ may have spread incompletely through the Hupa-speaking region, affecting Hupa proper and Chilula; variation in Whilkut may be due to the fact that the change was still in progress at the time of colonization. In scenario (1b), Proto-Athabaskan **q, **x̣ > Proto-California *x > /k/ in all California lan-

[41] Powers continues: "Although most of their petty tributaries had their own tongues originally, so vigorously were they put to school in the language of their masters that most of their vocabularies were sapped and reduced to bald categories of names. They had the dry bones of substantives, but the flesh and blood of verbs were sucked out of them by the Hupâ" (Powers 1877:72).

guages except Hupa; variation in Whilkut may be due to the relatively recent diffusion of the [k]-variant, perhaps from Nongatl,the Eel River dialect adjacent to Whilkut territory.

Golla (2011:81) accepts scenario (1b) in his implicit reconstruction of *x for Proto-California Athabaskan. Evidence that this view is correct comes from two arguments. The first invokes structural parallels found in other sound changes that took place in California Athabaskan. First, the merger of Proto-Athabaskan *q, *x̣ is not an isolated sound change. The key correspondences are as follows:[42]

| | *q | *x̣ | *ɢ | *ɣ | *w |
|---|---|---|---|---|---|
| Hupa | x | x | g | w | m |
| Other CA Athabaskan | k | k | g | g | b |

Table 3.19: Sound correspondences for select velar consonants in California Athabaskan

Crucially, all back-velar continuants shift to non-sonorant stops in languages other than Hupa. Under scenario (1b), but not (1a), this outcome can be treated as part of a single historical process of fortition of velar continuants. Assuming an early merger of Proto-Athabaskan **q and **x̣ as Proto-California *x, the changes *x > k, *ɣ > g, and *w > b receive a unified treatment as having occurred subsequent to the separation of Hupa from the rest of the California Athabaskan group. If instead there was an early merger of **q and **x̣ as *k in Proto-California Athabaskan, fortition affecting **ɣ and **w would have to be treated as unrelated changes that occurred after Hupa had split off from the rest of the group.[43]

A second argument in favor of an early merger of *q and *x̣ as /x/ comes from C. Hart Merriam's documentation of the Eel River dialects, specifically the variety he called Set-ten-bi-den (better known as Lassik). Merriam reports that the aboriginal territory of people speaking this dialect stretched from the South Fork of the Eel River to the South Fork of the Trinity River in Humboldt and Trinity Counties (31/005). Their aboriginal territory was therefore not adjacent to that of the Hupa or Whilkut. Merriam worked with two consultants named Lucy Young and Yellowjacket.[44] Merriam explicitly noted that Lucy Young had initial [k] where Yellowjacket had [x] (Merriam's <h>). If there was an early merger of **q and **x̣ as a fricative in all of California Athabaskan, then Yellowjacket's speech can be considered conservative and indicative of a dialect where the subsequent shift *x > k occurred across a continuous area but was incomplete in some areas. The opposite view, that there was an early merger **q, **x̣ > *k, would require positing an independent shift *k > x in Hupa and in Set-ten-bi-den as spoken by Yellowjacket. The latter scenario is not impossible but requires positing parallel development in lineages that had already been separated for hundreds of years.

These two arguments suggest that the [x]-variant in Whilkut is archaic. If so, [k] in Whilkut was most likely due to contact with Athabaskans speaking other languages immediately to the south. Variation involving [x] and [k] in Whilkut, including among people living in the Blue

[42] Cf. Li (1930a:32-35) for complications in stems vs. prefixes, and Goddard (1912:11) for a description of variable phonetic realizations of velar stops in Kato and their correspondences with Hupa.

[43] This argument is weakened by the fact that *w also undergoes partial fortition to /m/ in Hupa, but even there it remains a sonorant.

[44] According to Golla (2011:80), Yellowjacket was also known as Jack French. Golla reports that he was a Pitch Wailaki (Merriam's Che-teg-ge-kah), whose aboriginal territory was along the North Fork of the Eel River. The argument given here holds either way, since neither territory is adjacent to Hoopa Valley.

Lake region who as far as is known had spent no significant amount of time living in Hoopa Valley, can therefore be interpreted as innovative in a direction that made Whilkut less rather than more similar to Hupa on the lower Trinity. The fact that the [x]-variant was to some extent retained in Whilkut can still be considered a factor in its advantage over the [k]-variant in rapid leveling of this parameter of variation under koineization in Hoopa Valley. However, the existence of [k] implies a pre-contact scenario in which the Whilkut had a connection with their neighbors to the south that was strong enough to foster the partial adoption of [k]. It is unclear whether this was a structural borrowing per se or perhaps a substrate effect under language shift (in the sense of Thomason and Kaufman 1988). It does, however, imply that the linguistic hegemony of the Hupa on neighboring groups in pre-contact times may not have been as pervasive as Powers suggested.

# Chapter 4: Variation and Dialect Contact in Oregon Athabaskan

## 1. Introduction

The Oregon Athabaskan languages were once spoken in an area stretching from the upper reaches of the Umpqua River in present-day Douglas County in Oregon to Del Norte County in northern California. In the wake of the Rogue River Indian Wars of 1855-1856 (Miller and Seaburg 1990, Schwartz 1997), most indigenous people of the region were dispossessed and consolidated on the Siletz and Grand Ronde reservations in the northwest part of the state, far from their ancestral homelands. Traditional food sources were scarce, and chronic undernutrition and disease took a heavy toll (Kent 1973, Bureau of Indian Affairs 1979). Those who managed to survive found themselves thrust elbow-to-elbow in a small number of reservation communities organized according to coarse-grained ethnolinguistic similarities. As in Hoopa Valley and elsewhere in North America, the ultimate result of the colonial encounter was obsolescence of the Athabaskan languages once spoken in the region. However, there were several decades during which Athabaskan languages continued to be spoken at Siletz and Grand Ronde alongside non-Athabaskan indigenous languages, Chinook Wawa, and English. This nexus of multilingual and multi-dialectal contact, specifically contact between different Athabaskan varieties, is the focus of the present chapter.

As discussed in §3 of chapter 1, a common outcome of dialect contact is koineization, which can be understood as the leveling of dialect differences over successive generations of speakers raised in close proximity to each other. Apparently this was the result of the resettlement of Athabaskan-speaking people at Siletz. Reporting on his work there in the mid-1960s, Golla (1976:218) noted that most of the differences between the Lower Rogue River dialects, by then collectively known as "Tututni," had all but disappeared: "The phonology and grammar of Mrs. Bensell's [Euchre Creek] dialect appear to hold for all Tututni. The only major differences among the dialects that survived at Siletz at the time of my work were lexical."[1] However, most details of what these differences might have been remain unexplored. As Golla (2011:70) has pointed out, much of what is known about Oregon Athabaskan is based on interviews conducted in the mid-20th century with speakers who had spent most of their lives in the reservation setting, presumably after any contact-induced changes had already run their course. It is only by examining documentation from earlier periods, closer to the time when resettlement occurred, that erstwhile differences between varieties can be detected and the effects of linguistic contact revealed.

These observations lead to the following questions of interest: which similarities between closely-related varieties of Oregon Athabaskan found in 20th century documentation are extensions of a pre-reservation status quo, and which are the result of processes set in motion by intensive contact after resettlement? Just as importantly, which differences between varieties were maintained despite the leveling that took place in the same period? Working with documentation of the Oregon Athabaskan languages from the mid-19th century onwards, the present chapter addresses these questions in light of the theoretical issues raised in chapter 1. Evidence is presented in §2 that the new reservation community underwent a degree of focusing (in the

[1] Golla's report matches that of Pierce and Ryherd (1964:142): "The amount of leveling that has taken place since these tribes were uprooted in 1856 and transported to reservations at Siletz and Grand Ronde, Oregon might tend to lead us to assume that the early nineteenth century relationships between these dialects and languages were closer than they actually were."

sense of Le Page and Tabouret-Keller 1985) in this period, leading to the expectation in the framework developed by Kerswill and Trudgill (2005) that koineization should have occurred.[2] Evidence presented in §3 tends to confirm this prediction, with dialect leveling targeting variation in Tututni related to the merger of Proto-Athabaskan *q and *x̣ (§3.1) (the same one discussed in Hupa in §4 of chapter 3) and distinctive properties of the Chetco variety (§3.2).

Nonetheless, despite convergent pressures in this period favoring the elimination of dialect variation as koineization proceeded, some dialect idiosyncrasies survived, the Galice Creek variety in particular maintaining several unique traits into the 20th century (§4.1). Even closely related dialects formerly spoken on the lower Rogue River avoided complete erasure of differences, Mikwanutni apparently maintaining a tendency to affricate onset coronal fricatives that set it apart from Tututni and other Lower Rogue dialects (§4.2). A final parameter of linguistic variation involving the phonetic realization of Oregon Athabaskan /s/ as alveolar versus dental or interdental is also considered (§5). This is interpreted as a functionally-motivated innovation (or perhaps reversion to an archaic variant) that spread within the Siletz reservation community in the historical period, suggesting that to some extent linguistic changes not attributable to language or dialect contact proceeded even as the languages approached the brink of oblivion.

Beyond its contribution to the understanding of the post-contact development of the Oregon Athabaskan languages in particular, this chapter sheds light on the general mechanisms by which dialects and closely related languages influence each other when they come into contact. In particular, it addresses the question raised in chapter 1 of the extent to which koineization proceeds essentially deterministically, with majority variants displacing minority ones, or is instead mediated by speakers' indexical use of linguistic resources at their disposal. The data presented in this chapter bear crucially on this issue insofar as the resettlement of Oregon Athabaskan people has a sociolinguistic profile similar to Trudgill's (2004, 2008a) tabula rasa dialect contact, the special circumstance that, according to Trudgill, led to similarities between colonial varieties of English that emerged in the southern hemisphere. Unlike the situation in Hoopa Valley discussed in chapter 3, where a reservation was established on one group's aboriginal territory, the Siletz reservation was over 100 miles from the nearest Athabaskan homeland, much farther for many groups, and no single Athabaskan group had a pre-existing claim to legitimacy in the locus of contact.[3]

Despite the sociolinguistic similarity with the tabula rasa cases considered in Trudgill (2004), in §6 the outcomes of contact are shown not to be what Trudgill's demographic determinism would predict: there is both too much and not enough retention of the idiosyncratic features of the pre-contact Oregon Athabaskan varieties, and the outcomes do not appear to be based straightforwardly on the relative proportions of speakers of each variety. The case study presented here can thus be taken as a counterexample to Trudgill's (2008b) strong claim that

---

[2] This is not to suggest that the languages had been previously isolated from each other, as there is ample historical and ethnographic evidence that the Athabaskan groups of Oregon had deep-rooted connections with each other and surrounding groups well before contact with Euro-American society (cf. Miller and Seaburg 1990, Hall 1992). Rather, the social and geographic dislocation in this period changed the nature and frequency of their interactions.

[3] The focus of this chapter is Siletz, where most of the Oregon Athabaskan groups wound up, but reference is sometimes made to the adjacent Grand Ronde reservation where similar socio-historical processes were taking place. The Tolowa, whose traditional territory was on the California side of the border, did not undergo this removal and instead were consolidated within their aboriginal territory; the Upper Umpqua wound up at Grand Ronde (Golla 2011:70-75).

there is no known case of koineization failing to occur in tabula rasa contact situations. It is suggested that indexical approaches to variation – including variation introduced through dialect contact – are better-suited to explaining what happened linguistically among Oregon Athabaskans living at Siletz.

## 2. Resettlement and Focusing

As outlined in §1, the Rogue River Indian War of 1855-1856 and subsequent resettlement on the Siletz and Grand Ronde reservations were watershed events in the history of contact between indigenous people of southwestern Oregon and Euro-American colonizers (Schwartz 1997). Most Oregon Athabaskan groups wound up at Siletz and were ultimately consolidated in a small number of communities that were partially segregated from the Takelma and other non-Athabaskan indigenous groups on the reservation.

Resettlement had far-ranging social consequences, among which was a partial erasure of distinctive pre-contact micro-regional identities. For example, Dorsey (1889, 1890) observed that most village groups at Siletz still maintained separate burial grounds, an extension of pre-reservation practice, but noted that all of the Chetco-speaking people had come to share just one, even though they had once lived in nine villages distinct enough that they could intermarry (village exogamy having been the norm in pre-contact times).[4] The intermingling of populations is also reflected in the following statement from the report of the Office of Indian Affairs' agent at Siletz in 1887: "It would be impossible to give the exact number of each tribe, on account of so much intermarrying among them" (Office of Indian Affairs 1887:189). This accords with the perception of Athabaskan people who were raised at Siletz in this period: Drucker's (1934) field notes include one woman saying that "people down on Lower Farm were all mixed up – Chetco, Rogue R., Euchre Cr., Sixes Cr., etc." At the same time, however, when Dorsey visited Siletz in the mid-1880s he found no common ethnonym in use for the Athabaskans with origins along the Rogue River (Dorsey 1890). The enduring importance of people's ancestral villages is evident moreover in the fact that people interviewed by Barnett (1934) and Drucker (1934) in the 20th century knew with a high degree of precision the names (if not the exact locations) of their parents' and grandparents' former village affiliations, even when in some cases they had never visited these places themselves (Drucker 1937:269, fn. 24).

There is very little published information about the status of Athabaskan and other indigenous languages spoken at Siletz and Grand Ronde for nearly thirty years following resettlement, except indirectly in the annual reports of the U.S. government's agents. The earliest of these indicate increased use of contact languages – English to some extent, but especially Chinook Wawa. J.H. Huffer, the schoolteacher at Grand Ronde in 1864, complained, "The

---

[4] In a similar vein, Everette (1882) reported that

> "they are very clannish in respect to 'Burying their Dead'. Each tribe greatly preferring to be by itself; they sometimes bury, in a 'Common church' or 'Burying Ground,' but this occurs very rarely: this, of course, only refers to 'Full Blood's' of each tribe; as the 'Half Blood's,' (i.e. Indians, whose Parents and Grandparents are all of Different tribes) generally bury their Dead, at the nearest and handiest place they can find." (Everette 1882:236)

Everette further observed of the Tututni that "these Indians are so mixed by Intermarriage, with other tribes, that they Literally have no tribal Laws or Relations of their own" (1882:236) – surely a distortion, but in keeping with the general picture of the dissipation of some cultural practices associated with pre-contact social organization.

progress of the Indian in school must necessarily be slow, because we have the difficulty of their language to overcome... I have found considerable difficulty in getting the scholars to speak the English language properly or at all, but I am now beginning to overcome that difficulty by giving premiums to those that speak the most and the best English at the end of each month" (Office of Indian Affairs 1865:94). Two years later J.B. Clark, another teacher at Grand Ronde, reported, "I find no difficulty in making them understand the English language, and I endeavor to make them communicate their ideas in the same language, but they will use that barbarous jargon, the Chenook" (Office of Indian Affairs 1866:82). This last statement shows clearly that only eight years after resettlement, Chinook Wawa had become a regular language of communication among children, in certain contexts at least (cf. Zenk 1988): the seeds of language obsolescence were already sown.

A similar ascendency of Chinook Wawa at Siletz is highlighted in comments by Dorsey (1889:55): "As more than twenty tribes, each having its own dialect, have been consolidated on this reservation, they are obliged to use a common language. So all speak Chinook jargon, and many are learning English." In providing population figures for Siletz, Everette (1882) counted the various Oregon Athabaskan groups separately, but noted "all these Indians talk the same language and call it (tu = tu = tĕne wá wá)" (that is, 'Tututni' plus 'language' in Chinook Wawa). Everette's comments, like Dorsey's, suggest a growing role for Chinook Wawa on the reservation, and also point to an emergent pan-Athabaskan ethnolinguistic identity, with all varieties of Oregon Athabaskan collapsed as 'Tututni'.[5] Ultimately, many people at Siletz also acquired some proficiency in English, and in 1889 Indian Agent Beal Gaither was able to report of people living at Siletz that "[m]ost of them speak and understand English well enough for ordinary intercourse" (Office of Indian Affairs 1889:274).

Nonetheless, the enduring strength of indigenous languages at Siletz and elsewhere in Oregon is found in the 1870 report of A.B. Meacham, Superintendent of Indian Affairs in the state, who called for the creation of boarding schools separating children from their families in order to better meet the government's assimiliationist aims. Meacham argued for this on the grounds that "so long as Indian children remain with their parents, spending all their leisure hours at home, *where they use their native language only*, they forget what is learned through the day" (Office of Indian Affairs 1870:53, emphasis added). A few years later, the continued influence of Athabaskan languages at Siletz was highlighted in another comment by Everette. Expressing gratitude to Tututni cultural informant William Strong, he wrote, "though Educated in the 'English Language,' he still 'thinks' in the Indian Language and therefore cannot talk 'Grammatically'" (1882:90).

The picture that emerges is that by the 1880s the Oregon Athabaskan languages, although still widely spoken some twenty-five years after removal to Siletz, were coming to be identified as varieties of a single language and were encountering new pressure from lingua francas, especially in public domains associated with Euro-American society. This shift went hand-in-

[5] This usage is already found in one of the earliest summaries of the Athabaskans in the region, the 1854 report of Indian Agent J.L. Parrish, who wrote, "I have found the natives all speaking one language, and from similarity of appearance, habits, and pursuits, consider them as being one nation or people, who from their language may be denominated To-to-tin, or To-to-tut-na; the latter appellation being applied to them by their early visitors" (Office of Indian Affairs 1854:286). There is no evidence that Parrish's opinion reflects a pre-contact sentiment shared among the Athabaskans themselves, but raises the possibility that community focusing found in the 1880s had antecedents in pre-reservation times.

hand with the dissipation of some salient cultural practices associated with pre-reservation village affiliations, especially among younger people of mixed ancestry. This accords with Leavelle's (1998:436) assessment of the situation on the adjacent Grand Ronde reservation, where "shared cultural practices and attitudes and the common reservation experience helped the Indians of Grand Ronde overcome the challenges of this diversity to forge an Indian identity rooted in the place and the history of their valley home."[6] The importance of this social and sociolinguistic convergence lies in the link between koineization and a focused social identity. As Kerswill (2002) observes, "for a koine to form, the speakers must waive their previous allegiances and social divisions to show mutual solidarity." The fact that some of the regional divisions separating Athabaskan groups in pre-reservation times were dissipating in the 1880s suggests that the community was becoming focused in the relevant sense and renders it more likely that leveling of erstwhile dialect differences might have taken place in the same period.

## 3. Koineization

The expectation of the three-stage model of dialect contact presented in chapter 1 (Kerswill and Trudgill 2005) is that dialect differences in the newly-focused speech community that emerged at Siletz should have undergone leveling and a koineized Oregon Athabaskan should have developed. To some extent, this is what the documentary record shows; two specific cases are considered in this section.

### 3.1. [k] ~ [x] Variation in Tututni

One case of leveling involves reflexes of Proto-Athabaskan *q (an aspirated back velar stop) and *x̣ (a back velar fricative), which merge as a fricative /x/ everywhere in Oregon Athabaskan except Galice, where aspirated /k/ is found instead.[7] This is illustrated normalized forms in Tables 4.1 and 4.2.[8]

---

[6] While the injustices perpetrated against indigenous people of the region in this period are undeniable, Leavelle (1998) maintains that indigenous groups were not mere passive victims of circumstance, emphasizing instead their agency, extraordinary resilience, and creativity in adapting to their new circumstances.

[7] Most 20th century researchers describe the Oregon Athabaskan reflexes of these consonants simply as "velar" and represent them phonemically as /g/, /k/, /k'/, /x/, and /ɣ/. The narrower phonetic transcriptions of Sapir (1914), Harrington (1981), and Landar (1977) suggest that a back (post-velar or uvular) articulation was probably maintained in some varieties. While it may not be an entirely trivial point, the precise place of articulation of the back velars is not the main concern here, and I adopt the usual phonemic transcription in the Oregon Athabaskan forms.

[8] Labialization of the initial consonants in Table 4.1 indicates a third person possessor, historically derived from a prefix *_m_- (Golla 1976, Collins 1985). According to Golla (1976), in most cases Oregon Athabaskan labialized velars are not reflexes of Proto-Athabaskan labialized velars. Instead, Proto-Athabaskan labialized velars merged with their non-labialized counterparts, with contrastive labialization re-emerging subsequently in the vicinity of round vowels and labial consonants (such as the possessed nouns shown here). Thus, while the merger of velar stops and fricatives also affected *q$^{w}$ and *x̣$^{w}$, as in chapter 3 I will abstract away from details of labialization in the discussion since they are orthogonal to the issues at hand.

| form | dialect & source |
|---|---|
| xʷeʔ | Coq (Harrington 1981) |
| xʷeʔ | Euc (Golla 2008) |
| xʷeʔ | Tut (Landar 1977) |
| xʷeː | Mik (Landar 1977) |
| xʷeː | ChC (Harrington 1981) |
| kʷeʔ | Gal (Landar 1977) |
| xʷe | PiR (Dorsey 1884m) |
| xʷeʔ | Tol (Bright 1964) |

Table 4.1: 'foot' in Oregon Athabaskan (P-A *qeʔ)

| form | dialect & source |
|---|---|
| xənnəs̬ | Coq (Harrington 1981) |
| xənəs | Euc (Golla 2008) |
| xənnəs | Tut (Dorsey 1884i) |
| xənnəθ | Mik (Landar 1977) |
| xənəθ | ChC (Harrington 1981) |
| kədəθ | Gal (Harrington 1981) |
| xenəs | Tol (Bommelyn 1989) |

Table 4.2: 'canoe' in Oregon Athabaskan (P-A *x̣əỹg̱ts' 'raft')

This is considered a largely exceptionless categorical merger in published descriptions of Oregon Athabaskan (e.g., Sapir 1914, Bright 1964, Hoijer 1966, Golla 1976; cf. #7 in Appendix A of chapter 2 and §4 of chapter 3).

Despite the ubiquity of this merger, a close look at early documentation of the Oregon Athabaskan languages reveals that aspirated [k] was sometimes found outside of Galice as late as the 1880s. Consider first the following selection of entries from a wordlist collected by Hubbard in the mid-1850s, one of the earliest attestations of Tututni (published in Taylor 1860). The first two columns of Table 4.3 show Hubbard's original transcriptions and normalized interpretations of them; the corresponding Euchre Creek forms transcribed by Golla (2008) in the 20th century are provided to illustrate the basis for this interpretation:[9]

[9] Golla's consultant, Ida Bensell, was associated with the Euchre Creek dialect, but as mentioned above all of the Lower Rogue dialects were known collectively as "Tututni" by the mid-20th century, and Mrs. Bensell's speech can be taken as representative of the Siletz koine. As noted in §5.4 of chapter 1, the normalized forms presented here do not attempt to correct likely omissions from the original transcriptions (especially glottalization).

| Tututni (Taylor 1860) | | Euchre Creek | gloss |
|---|---|---|---|
| <hwoquarne> | xʷogʷane | xogʷaneʔ | ‘arm (3POSS)’ |
| <hwoquot> | xʷogʷod | -gʷəd | ‘knee (3POSS)’ |
| <tsach> | tsax | tṣ’a:xe | ‘woman’ |
| <tlchut> | təlxəd | təlxəd | ‘water’ |
| <scharchah> | sxaxa | ṣxexe | ‘child’ |
| <hwun> | xʷən | xwənʔ | ‘fire’ |
| <narke> | nake | naxi | ‘two’ |
| <squallah> | skʷəla | sxʷəlaʔ | ‘five’ |
| <quiesse> | kʷiəse | xʷese | ‘ten’ |

Table 4.3: Selected vocabulary from Hubbard’s Tututni wordlist (published as Taylor 1860)

As might be expected for an adaptation of English orthographic conventions to a new language, <qu> in Hubbard's list represents a labialized velar stop (voiceless unaspirated in this case), as shown by the entries for ‘arm’ and ‘knee’. The sequence <hwo> in the same entries is an indefinite/generic third person possessor prefix *x(ʷ)o-*, probably etymologically derived from the Proto-Athabaskan areal prefix *qo-, used as an “‘indefinite human object/possessor” in some Athabaskan languages (Leer 2005:292).[10] Hubbard’s <ch> in ‘woman’, ‘water’, and ‘child’ represents [x], and his <narke> is probably [na:ke], <ar> perhaps a representation of vowel length by a speaker of a non-rhotic dialect of English (cf. Hinton 1979:7).[11]

Hubbard’s entry for ‘fire’ in Table 4.3, from Proto-Athabaskan *qu̯nʔ, shows the expected fricative in Tututni, as does in the indefinite possessor prefix (from *q) found with body part vocabulary. This is confirmed in another wordlist collected in the same period. Kautz (1855) transcribed a continuant in <whun> (= [xʷən]) ‘fire’ and several words for body parts, such as <whus-see> (= [xʷəsi:]) ‘head’, where <whu-> is the indefinite possessor prefix. In the remaining forms in Table 4.3, Hubbard recorded a stop rather than a fricative in ‘two’, ‘five’, and ‘ten’.[12] This is again partly confirmed by data from Kautz (1855), who transcribed <nah-ke> (= [na:ke]) ‘two’ and <qui-es-sai> (plausibly [kwəse] or [kwese]) ‘ten’ with velar stops, but <swo-lah> (= [swola]) ‘five’ with a continuant [w]. Nothing is known about

[10] This prefix is not included in Golla’s (1976) description of Tututni but is pervasive in the documentation of Everette (1882) and Dorsey (1884i). Its use as a generic is highlighted in minimal contrasts documented by Dorsey like <qwĭc> (= [xʷiš]) ‘any nose’ versus <mĭc> (= [miš]) ‘his nose’ (1884i). It may be the same prefix as the indefinite human possessor *мə-* in Galice identified by Hoijer (1966), but /ʍ/ from *q in Galice is unexpected. Interestingly, in Dorsey’s (1884e) documentation of Chetco he notes that “kwŭ can be prefixed to any noun (S.3) to form the generic,” offering examples such as <kwŭ-ṭa> ‘mouth, a mouth’ – notice that in Chetco the prefix starts with a stop instead of a fricative. One speaker used both variants, explicitly identifying the one with the fricative as Tututni and the one with the stop as Chetco.

[11] Garrett (p.c.) has suggested that this coda <r> might instead indicate some difference in vowel quality rather than length.

[12] Leer (2005) reconstructs Proto-Athabaskan *nąː(-qe:) for ‘two’ (in the phrase *nąː(-qe:) dəne: ‘two persons’); Krauss and Leer (1981) reconstruct *qʷe-nẹ:z-ya:ỹ for ‘ten’. Leer (2011) reconstructs a velar stop in ‘five’.

the speakers Hubbard and Kautz worked with, but these data show that in the mid-19th century a velar stop was found instead of [x], for some lexical items at least, in a variety of Oregon Athabaskan identified as Tututni spoken by people raised prior to resettlement at Siletz.

A generation after resettlement, both Everette (1882) and Dorsey (1884i) transcribed [x] in Tututni <na$^{l}$-qe> (= [naxe]) 'two' and <qwe$^{l}$-e-sĕ> (= [x$^{w}$ese]) 'ten'. Documentation of most varieties of Oregon Athabaskan in the 20th century have a fricative in 'five', contrasting with Hubbard's stop, but Everette (1882) recorded [k$^{w}$] in Tututni, matching Hubbard, where Dorsey (1884i) transcribed [w], as did Kautz:[13]

| form | dialect & source |
|---|---|
| ṣx$^{w}$əlləʔ | Coq (Harrington 1981) |
| swala | Six (Dorsey 1884l) |
| sx$^{w}$əlaʔ | Euc (Golla 2008) |
| sk$^{w}$əɣla | Tut (Everette 1882) |
| swala | Tut (Dorsey 1884i) |
| sx$^{w}$əla: | Mik (Landar 1977) |
| sxola: | ChC (Sapir 1914) |
| ṣxolla: ~ ṣwəlla: | ChC (Harrington 1981) |
| sko:loʔ | Gal (Landar 1977) |
| swəla | PiR (Dorsey 1884m) |
| ṣweloh-ne | Tol (Bommelyn 1989) 'five people' |

Table 4.4: 'five' in Oregon Athabaskan

In the 1880s, then, some speakers of Tututni still produced a velar stop rather than a fricative in 'five', apparently a continuation of variation that was also present in the 1850s.

Another item of interest in Hubbard's wordlist is an entry *kowlawo*, glossed 'tooth'. The syllable <la> is mysterious, but the root is probably <wo> 'tooth' (lenited [ɣ], transcribed as such by Harrington and others) and <ko> a possessive prefix. We might expect initial <hwo> here for the indefinite possessor, matching the other entries for body parts in his list (e.g., the first two forms in Table 4.3). This form is anomalous until we compare it with Dorsey's and Everette's documentation. Dorsey consistently transcribed a fricative for this prefix in Tututni; Everette frequently did so as well, but there are numerous examples (many in a list of body parts) where he transcribed a stop instead. A sample of these forms is given in Table 4.5.[14] Notice that Everette transcribes 'heart' twice, once with a fricative [x$^{w}$] in the prefix (in a list of human body parts) and once with [k$^{w}$] (in a list of animal body parts).

[13] Galice has /k/, as expected, and some other varieties show /w/ instead of /x$^{w}$/. The labialization here is likely due to an original round vowel following /k/ (Golla 1976), as in the Chasta Costa forms.

[14] Everette sometimes wrote <k'> or <k'w> for this prefix, the apostrophe diacritic indicating that the consonant is "exploded" (Everette 1882:232); he appears to have used it for any sort of laryngeal noise, both glottalization and aspiration.

| Everette (1882) | | Dorsey (1884i) | | gloss |
|---|---|---|---|---|
| <qwŏˈni> | xʷoni | <qoni> | xoni | ‘face’ |
| <qoˈ-la> | xola | <qwa-laˈ> | xʷala | ‘hand’ |
| <kwûsaˈ-qeˈ> | kʷəsaxe | <qwo-sĕˈ-qĕ> | xʷosexe | ‘saliva’ |
| <k’wûc> | k’ʷəš | <qwic> | xʷiš | ‘nose’ |
| <hwûˈseˈ> | xʷəse | <qwû-sĕˈ> | xʷəse | ‘heart’ |
| <k’wûˈsä> | kʷəse | - | - | ‘heart’ |

Table 4.5: Indefinite possessor prefix in Tututni (1880s)

Similar variation is found in Everette’s transcription of words formed on the roots glossed ‘foot’ (cf. Table 4.1 above) and ‘fire’:

| Everette (1882) | | Dorsey (1884i) | | gloss |
|---|---|---|---|---|
| <ck’eˈ> | šk’e | <qo-qwĕˈ> | xo-xʷe | ‘foot’ (1SG.POSS / 3POSS) |
| <tcăc-k’we> | tšəš kʷe | <tc’ac qwĕˈ> | tš’aš xʷe | ‘bird’s tracks’ |
| <k‘wûˈg-qe> | kʷəkxe | - | - | ‘hoof’ |
| <qoˈ-qeˈ> | xoxe | <qo-qwĕˈ> | xoxʷe | ‘footprint’ (3POSS) |
| <qeˈ-năʟ-meˈ> | xe nəɬmeʔ | - | - | ‘(bird’s) spur’ (lit. ‘feet knives’) |

Table 4.6: ‘foot’ (< P-A *qeʔ) and related vocabulary in Tututni (1880s)

| Everette (1882) | | Dorsey (1884i) | | gloss |
|---|---|---|---|---|
| <qhwûn> | xʷən | <qwûn> | xʷən | ‘fire’ |
| <qhwûn-sĭt> | xʷənsid | <kwûn-sĭtˈ> | kʷənsid | ‘charcoal’ |
| <qhwûn-tseˈ> | xʷəntse | <qwûn-seˈ> | xʷənse | ‘ashes’ |
| <qûn’-taˈ> | xənʔda | <qwûn-t̰aˈ> | xʷənda | ‘fireplace’ |
| <qûn-taˈ-Let> | xənda ɬed | <qwûn-t̰aˈ> | xʷənda | ‘smoke-hole’ |
| <k’wûn-taˈ> | kʷənda | - | - | ‘smoke-hole’ |

Table 4.7: ‘fire’ (< P-A *qu̲nʔ) and related vocabulary in Tututni (1880s)

Overall, Everette’s documentation suggests ongoing variation involving [k] and [x] in different tokens of some words and the indefinite possessor prefix.

Dorsey’s notes, collected only two years later, show much less evidence for [k], not only in Tututni but in the other Lower Rogue dialects he documented as well (Dorsey 1884c,f-m). However, Dorsey did transcribe [kʷ] in the compound ‘charcoal’ built on the root ‘fire’ in Table 4.7, where Everette had recorded [xʷ]. Moreover, for at least one lexical item Dorsey’s Tututni consultants showed considerable variation. Recall the form *scharchah* (normalized as [sxaxa]) ‘child’ from the Hubbard wordlist in Table 4.3. Both Dorsey and Everette transcribed this word in various ways, sometimes with [k], sometimes with [x], as the normalized forms in Table 4.8 show:

| form | gloss | dialect & source |
|---|---|---|
| sxexe, tsxexe | ‘child’ | Six (Dorsey 1884l) |
| tsxeɣt’i | ‘she had a child’ | |
| | (*-t’i* ‘proprietive’) | |
| skeke yasle | ‘boy beginning to walk’ | Tut (Dorsey 1884i) |
| tsxexe | ‘young man’ | |
| skexe ~ sxexe | ‘boy’ | Tut (Everette 1882) |
| ts’axe sxexe | ‘girl baby’ | Mik (Dorsey 1884f) |
| ts’axe skeke yasli | ‘female infant’ | ChC (Dorsey 1884c) |
| sxexe | ‘boy’ | |

Table 4.8: Normalized transcriptions for ‘child’ and related vocabulary in Lower Rogue dialects (1880s)

Strikingly, both of Dorsey's Tututni forms in Table 4.8 were produced by a single speaker, Shem Lafayette, and both of his Chasta Costa forms were elicited from Jake Orton. Notice also the intermediate form [skexe] transcribed by Everette, with one velar stop and one velar fricative, perhaps a hybrid interdialect form commonly found in dialect contact situations (cf. Trudgill 1986:62-65 and discussion of similar phenomena in Hupa in §4.1 of chapter 3). Table 4.8 gives the impression of speakers using sometimes one, sometimes another of several variants of the same word in circulation at Siletz in the 1880s.[15]

To summarize the data presented thus far, variation involving [k] and [x] is found in 19th century documentation of Tututni. In the 1880s especially, variants of some words and the indefinite possessor prefix are found with [k] instead of [x]. In some cases it is clear that this variation involved different tokens of the same word, sometimes as produced by a single speaker. The timing of Everette’s and Dorsey’s visits to Siletz becomes significant: one generation after the Oregon Athabaskans were consolidated on the reservation, the documentation shows precisely the inter- and intra-speaker variation predicted by Kerswill and Trudgill (2005) for Stage II speakers. The interpretation of these data is that speakers of Oregon Athabaskan who grew up at Siletz from the 1850s onwards were exposed to a wide variety of dialects, some with [k], others with [x], and some (like Tututni) most likely with both. By the 1880s, [k] was losing ground to [x] as leveling of the closely related Rogue River dialects got underway in the development of the Siletz Athabaskan koine.

20th century documentation of Oregon Athabaskan presents a much less complicated picture, with variation associated with [k] ~ [x] in the 1880s reduced in the last generations of first-language speakers. Although aspirated velar stops are rare in 20th century sources, there are enough tokens recorded to warrant consideration. Some normalized examples are given in

[15] Comparison with other Athabaskan languages suggests that the fricative is innovative in this word: Babine-Witsuwit’en, which retains Proto-Athabaskan *q as such (Krauss and Golla 1981), has a word *sqaq* ‘child’ that appears to be cognate with the Oregon Athabaskan *sxexe* (Hargus and Tuttle 1997:203-204). Initial [ts] vs. [s] in Table 4.8 is discussed in §4.2.

Table 4.9:

| Tut<br>(Dorsey 1884i) | ChC<br>Sapir (1914) | Coq<br>(Harrington 1981) | Euc<br>Golla (2008) | gloss |
|---|---|---|---|---|
| sakətɬ | - | ṣa:kkʷəɬ | ṣaʔkuɬ | ‘beaver’ |
| θikədled | - | θi:kkəhd | sika:d | ‘top of head’ |
| kʷət-tɬ’i | - | xʷəttɬ’iʔ | kʷətɬ’iʔ | ‘clay, mud’ |
| kʷəstane | kʷəsta:ne | gʷəθta:nne | gʷəsta:ne: | ‘six’ |
| dəmmilke | dəmelʔge | dəmme:lkeʔ | - | ‘pelican, crane’ |
| katɬ’a | - | ka:tɬ’əʔ | - | ‘crab’ |
| kəllixe | - | kəlli:xxe: | - | ‘moccasin’ |

Table 4.9: Normalized transcriptions of words with aspirated [k] in Rogue River dialects (19th c. – 20th c.)

Sapir (1914:284) had found a single exception to the velar stop~fricative merger in Chasta Costa, aspirated back velar [kʷ] occurring in <kʻwʌs/tʻā́/ne> [kʷəsta:ne] ‘six’. Later, Harrington (1981) transcribed <kʻ>, an aspirated velar stop, in a handful of words in Upper Coquille, Chasta Costa, and Tututni (Upper Coquille examples given in Table 4.9), consistently enough that it is unlikely to be a mis-transcription; he may, however, have intended a front velar rather than back velar articulation in some cases.[16] Similar considerations apply to a handful of words transcribed with [k] by Golla (2008). These data show that aspirated [k] was found outside of Galice in some lexical items in the 20th century, in some cases in the same items where [k] was transcribed in the 19th century. Many of these same words were transcribed with velar stops most plausibly interpreted as aspirated by Dorsey and/or Everette, indicating that for some lexical items an aspirated velar stop found in the 1880s was transmitted as such transgenerationally throughout the reservation period.

Variation involving [k] and [x] in particular lexical items is found only rarely in 20th century documentation: most words have [x] and a few may have [k], but for any given word speakers seem to have converged on one or the other. Direct comparison with the 19th century Tututni data presented above is impossible in many cases due to the small number of forms recorded in the 20th century with speakers identified specifically with the Tututni-Joshua dialect. However, the normalized forms in Table 4.10 were transcribed by Harrington with [x] in both Tututni (as spoken by Lucy Smith) and Upper Coquille (as spoken by Coquelle Thompson), where either Everette (1882) or Dorsey (1884i) had recorded [k] in Tututni.

[16] Many of Harrington’s aspirated tokens are transcribed with lowercase <k>, as opposed to the much more frequent (when unaspirated) small capital <ᴋ>.

| Everette (1882) | Dorsey (1884i) | Harrington (1981) | gloss |
|---|---|---|---|
| kʷəskəs | hʷəɬ haɬ | xʷəɬxa:ɬ | ‘flea’ |
| kxə̃ne ~ kʷəxẽ(ʔ)e | - | xənʔe:- | ‘river’[17] |
| skexe ~ sxexe | tsxexe ~ skeke | sxe:xxe | ‘child’ |

Table 4.10: Normalized transcriptions with variable [x] ~ [k] in 19th and 20th century Rogue River dialects

In each of these cases, Harrington transcribed [x] in both Tututni and Upper Coquille, suggesting that speakers of Tututni in the mid-20th century tended to converge on [x] for words that varied between [k] and [x] in the late 19th century documentation of Dorsey and especially Everette.

Compare also 20th century tokens of words containing the root ‘fire’ in Table 4.11 (normalized) with the 19th century forms in Table 4.7 above:

| form | gloss | dialect & source |
|---|---|---|
| xʷənʔ | ‘fire’ | Coq (Harrington 1981) |
| xʷənʔṣeʔ | ‘ashes’ | |
| kʷənʔθid | ‘charcoal’ | |
| xʷənʔ | ‘fire’ | Euc (Golla 2008) |
| xʷənʔsid ~ kʷənʔsid | ‘charcoal’ | |
| xʷən | ‘fire’ | Tut (Harrington 1981) |
| xʷənʔθid | ‘charcoal’ | |
| xʷɔn | ‘fire’ | Mik (Landar 1977) |
| xʷən | ‘fire’ | ChC (Harrington 1981) |
| xʷən xənəs | ‘steamboat’ (lit. ‘fire canoe’) | |
| kʷɔnʔ | ‘fire’ | Gal (Landar 1977) |
| kʷɔnʔθid | ‘live coals’ | |

Table 4.11: ‘fire’ in 20th century Rogue River dialects (P-A *qu̯nʔ ‘fire’)

Notice that Tututni *xʷənʔθid* ‘charcoal’ as transcribed by Harrington (1981) has a fricative rather than a stop. In other dialects, some residual variation between [x] and [k] is found for this item, as demonstrated by the two forms meaning ‘charcoal’ recorded in both Euchre Creek and Upper Coquille.

[17] Transcribed by Harrington in Tututni [xənʔe:ttɬ’əɣəš] ‘river snake’ (*tɬ’əɣəš* ‘snake’), Upper Coquille [xənʔe:kk’wət] ‘on the river’ (-k’wət ‘on’).

The upshot is that even in the 20th century, Stage III speakers (the second generation raised in the mixed dialect reservation community) retained some variation involving [x] and [k], but much less pervasively than was found in the 1880s. With few exceptions (such as 'charcoal'), most lexical items became uniformly [k] or [x], both for particular speakers and across the community as a whole. The stop variant may have been retained sporadically, but there was an overall tendency for dialect leveling to favor the fricative. What had been a linguistic variable in the Siletz reservation community in the 1880s and probably earlier largely ceased to be so in the 20th century. The exception was Galice, which maintained /k/ throughout, a point that will be considered in §4.1.

## 3.2. Chetco

As discussed in §5.3 of chapter 1, Chetco-Tolowa is generally considered distinct enough from Tututni and the other Rogue River dialects for them to be treated as separate languages, albeit with some degree of mutual intelligibility. The earliest documentation of Chetco is found in Dorsey (1884e), where at least three features distinguish it from his documentation of Tututni (1884i). The first is nasalization: Proto-Athabaskan coda nasals are lost following full vowels in all of Oregon Athabaskan (#23 in Appendix A of chapter 2), but they are maintained as vowel nasalization in Chetco-Tolowa (Bright 1964, Golla 2011), transcribed by Dorsey with superscript <n>. The others are coda [tɬ] (Dorsey's <tçl>) in Chetco versus [l] in Tututni, and two words beginning with [dž] (Dorsey's <t̨ç>) in Chetco but [d] (Dorsey's <t̨>) in Tututni: 'man/male' and 'elk'. Examples of these Chetco features as compared with Tututni are shown in Table 4.12:[18]

| Chetco (Dorsey 1884e) | | Tututni (Dorsey 1884i) | | gloss |
|---|---|---|---|---|
| <ts'a$^{n}$-qe> | ts'ãxe | <tsa$^{ˈ}$-qe> | tsaxe | 'woman' |
| <qwa$^{n}$-k'qi> | x$^{w}$ãk'i | <qwa$^{ˈ}$-k'qĕ> | x$^{w}$ak'e | 'rib' |
| <t̨i$^{n}$-tcu> | dĩtšu | <di$^{ˈ}$-i-tcu> | ditšu | 'mountain lion' |
| <r$^{x}$i-tūtçl-ta$^{ˈ}$-k̨wĕ> | ɣitutɬtag$^{w}$e | <qwa-tŭl$^{ˈ}$-tu-kwĕ$^{ˈ}$> | x$^{w}$ətəltək$^{w}$e | 'calf (3POSS)'[19] |
| <r$^{x}$i-mûtçl$^{ˈ}$-kûl-lĕ> | ɣimətɬkəle | <qo-mûl-kûl lĕ> | xoməlkəle | 'bladder (3POSS)' |
| <natçl-mĕ> | natɬme | <nal$^{ˈ}$-mĕ> | nalme | 'knife/iron' |
| <t̨ç̨ûs$^{ˈ}$-snĕ> | džəsne | <t̨i-snĕ> | disne | 'man/male' |
| <t̨ç̨is$^{ˈ}$-tcu> | džistšu | <t̨ĕs$^{ˈ}$-tcŭ‘> | destšuh | 'elk' |

Table 4.12: Features of Chetco vs. Tututni documented by Dorsey (1884e)

[18] Tolowa is described as having allophonic [tɬ] for coda /ɬ/ (Bright 1964), so the first of these differences might involve phonemic /ɬ/ vs. /l/ with the affrication a regular allophonic rule in Chetco. However, there are many other cases where Dorsey transcribes coda <çl> (= [ɬ]) in both Chetco and Tututni, so it is unclear precisely what conditions the difference. Dorsey (1884g) noted variation in Tututni involving [ǯ] and [d] in proximal deictics such as <t̨çi> ~ <t̨i> 'this', which can also be interpreted as due to dialect mixing in this period.

[19] The possessed nouns 'calf' and 'bladder' differ in the definiteness of the pronominal element: *ɣi-* (Dorsey's <r$^{x}$i->) in the Chetco words is a proclitic demonstrative glossed as 'that' by Golla (1976); *xo-* or *x$^{w}$ə-* (Dorsey's <qo-> and <qwa->) in the Tututni words is the indefinite possessor prefix discussed in §3.1.

It is important to note that even in the 1880s, there was variation along these lines: while two of the Chetco speakers Dorsey worked with on the whole agree in the features given in Table 4.12, a third offered forms like <ṯisne> 'man' (with initial [d] rather than [dž]) and <tsaˡ-qe> 'woman' (with the first vowel oral instead of nasalized).[20]

In the 20th century, Chetco people living at Siletz appear to have adopted the Tututni values for these features. The main source of information comes from Billy Metcalf, who was interviewed by both Homer Barnett and Elizabeth Jacobs in the 1930s. Metcalf's mother was Chetco and his father was from Rogue River; he was born in Tolowa country in California and lived at Smith River until the age of nine before moving to Siletz (Barnett 1934). E. Jacobs (1968) estimated him to be in his mid-fifties in 1935, so he was born around 1880, squarely in the generation of speakers that would be identified with stage III of Kerswill and Trudgill's (2005) theory of koineization outlined in chapter 1.

In the published edition of a text dictated by Metcalf, Jacobs noted that he identified himself as a speaker of Chetco even though his speech lacked the vowel nasalization typical of Chetco as documented by other researchers.[21] She suggested that "[s]ince Mr. Metcalf had resided at Siletz most of his life, these phonological differences may reflect an assimilation to the predominant Tututni dialect continuum spoken there" (E. Jacobs 1977). Moreover, it is evident in the texts transcribed by Jacobs (1968, 1977) that Metcalf had coda [l] rather than [tɬ] and initial [d] in 'man', both features associated with Tututni rather than Chetco. Other direct comparisons with Dorsey's (1884e) documentation of Chetco are hard to find because the texts transcribed by Jacobs happen not to include vocabulary found in Dorsey's wordlists. However, some points of comparison are shown in Table 4.13, and also in Table 4.14 if the Pistol River dialect documented in Dorsey (1884m), where coda [tɬ] is pervasive, can be used as a proxy for Chetco:[22]

---

[20] The same speaker offered body part vocabulary that was distinct enough from the other Chetco speakers that Dorsey (1884e) noted it "may belong to some other dialect." There were also evidently differences among speakers in their treatment of possessed noun paradigms. This is further evidence for a high degree of dialect variation at Siletz in this period, as expected in the theory of dialect contact developed by Kerswill and Trudgill (2005).

[21] Jacobs further noted that Metcalf's speech lacked the prominent retroflex consonants documented in Chetco by other researchers, reflexes of a Proto-Athabaskan palatal series (cf. Bright 1964, Golla 1976, 2011). The retroflexes in Chetco-Tolowa are described by Golla (1976:219) as "strongly r-colored, lip-rounded palatal sibilants." In most other Oregon Athabaskan varieties the distinction between plain versus retroflex consonants "is reduced to a subtle flat versus plain contrast, difficult even for a linguist with a trained ear to hear consistently in the speech of surviving speakers." Retroflexion is mostly absent in Dorsey's Chetco notes, but he did transcribe it in the few Tolowa words he recorded and probably was able to hear it accurately in its extreme realization in that dialect. Hence Tolowa <t'ʳaⁿˡ-qe> 'woman' (Dorsey 1884n) but Chetco <ts'aⁿˡ-qe> (Dorsey 1884e), superscript <r> doubtless indicating retroflexion (cf. Tolowa [tʂ'ɑ̃ɑ̃xeh] in Bright 1964:102). Interviews with Chetco speakers conducted by Golla and Bright took place at Smith River (Golla p.c.), so the salient retroflexion detected in their speech might be a contact-induced assimilation by people who lived in close proximity to the Tolowa, or might indicate a sub-dialectal division within Chetco.

[22] The Pistol River is located between the Chetco and Rogue Rivers along the Oregon coast. The dialect from this region, as documented by Dorsey (1884m), also matches Chetco with respect to vowel nasalization and the pronunciation of 'man/male' and 'elk', suggesting that it may have been linguistically intermediate between Chetco-Tolowa and the Lower Rogue dialects (contra Golla 2011:73).

| Chetco (Dorsey 1884e) | | Tututni (Dorsey 1884i) | | Billy Metcalf | gloss |
|---|---|---|---|---|---|
| <ts'a$^{n|}$-qe> | ts'ăxe | <tsa$^{|}$-qĕ> | tsaxe | ts'axe | 'woman' |
| <qa$^{|}$-$^{r}$xûtçl-kĕt> | xaɣətɬked | <ha-ɻal$^{|}$-kĕt> | haɣalked | haɣəlged | '(animal) skin' |
| <ʈ̵ç̵ûs$^{|}$-sne> | džəsne | <ʈ̵isnĕ$^{l}$> | disne | disne | 'man' |

Table 4.13: Tututni features in Billy Metcalf's Chetco (20th c.)

| Pistol River (Dorsey 1884m) | | Tututni (Dorsey 1884i) | | Billy Metcalf | gloss |
|---|---|---|---|---|---|
| <tûtçl-qût> | təɬxəd | <tûl$^{|}$-qût> | təlxəd | təlxəd | 'water' |
| <ne-ʈ̵etçl$^{|}$-yu> | nedetɬyu | <nas-tĭl$^{|}$-yu> | nasdilyu | naɣasdəlyu | 'beads' |

Table 4.14: Coda [l] in Billy Metcalf's Chetco (20th c.)

However, not all Chetco features are lacking from Metcalf's speech. Thus, while he has the Lower Rogue initial [d] in 'man', Barnett (1934) transcribed [dž] in two separate tokens of 'elk' produced by Metcalf.[23]

Although Billy Metcalf identified his language as Chetco, presumably due to a connection to his mother's family's home territory, he was in fact speaking a variety of Oregon Athabaskan that was virtually indistinguishable from one known as Tututni. What this suggests is that in the development of an Oregon Athabaskan koine at Siletz, there was a partial decoupling of linguistic structure and ethnolinguistic identity. As noted in §2, it is clear that knowing the place of origin of one's parents and grandparents remained in the 20th century an important part of people's family histories, since nearly everyone interviewed by Drucker (1934) was able to provide that information. Some people identified their language as "Tututni" regardless of whether or not their ancestors happened to come from the mouth of the Rogue River; others called their language "Chetco" according to their ancestral place of origin. The evidence considered here suggests that the structural details distinguishing these varieties from one another were minimal, the result of dialect leveling among Stage III speakers raised in the mixed dialect community at Siletz.[24]

Another point of interest is that the Chetco-Tututni leveling is directional: "Chetco" of 20th century Siletz resembles the Tututni of the late 19th century rather than vice-versa. Whether this is always the case or whether some Chetco features can be found in 20th century Tututni remains to be explored, but this is probably what Trudgill's (2004) demographic determinism in tabula rasa dialect contact would predict. Although the Chetco were a relatively large group at Siletz (see Table 4.20 in §6 below), they were a minority overall, so Chetco-specific linguistic features would have been at a disadvantage for selection as an Athabaskan koine developed.

---

[23] Metcalf's speech as transcribed by Barnett also lacks nasalized vowels, and this is also the case for other Chetco speakers documented by Philip Drucker (1934) at Siletz. However, it unclear whether or not Barnett and Drucker were reliable transcribers in this regard (see discussion for Galice in §4.1). Drucker also transcribed what might be coda [tɬ] in four placenames provided by a speaker identified as Chetco, but the interpretation of his orthography, superscript capital L with underdot (<$^{ɻ}$>) is uncertain at present.

[24] According to Golla (1976), residual lexical differences separated dialects well into the 20th century. Indeed, comments by people at Siletz in the notes of Barnett (1934), Drucker (1934), and Harrington (1981) show that there remained awareness within the community of certain words being associated with particular varieties.

## 4. Maintenance

While Chetco became largely indistinguishable from Tututni, other varieties of Oregon Athabaskan spoken at Siletz maintained their idiosyncrasies more or less intact into the 20th century. Two cases are considered: Galice, which was relatively divergent compared to other varieties in the Rogue River dialect network, and Mikwanutni, a dialect that was otherwise thoroughly Lower Rogue in character.

### 4.1. Galice

As discussed in §5.3 of chapter 1, the Galice Creek variety of Oregon Athabaskan is classified by Golla as a member of the Rogue River dialect network, coordinate with Upper Coquille and the Lower Rogue dialects (including Tututni). Some authors disagree with this assessment (e.g., Sapir 1914), and computational results presented in chapter 2 suggest that Chetco-Tolowa and Tututni may have been more closely related to one another than Galice and Tututni were. Nonetheless, Galice is sufficiently similar to Tututni for this to be a matter of debate.

Galice, as documented in the 20th century, maintained several phonological properties that set it apart from the emergent Oregon Athabaskan koine that developed at Siletz: retention of nasalized vowels, merger of Proto-Athabaskan *q and *x̣ as /k/ instead of /x/, denasalization of nasal stops before oral vowels (Krauss and Leer 1981), and lenition of intervocalic /ɣ/. These features are all evident in the 19th century documentation of Galice collected by Dorsey (1884b) and in 20th century sources such as Sapir (1914), Barnett (1934), Hoijer (1966, 1973), M. Jacobs (1968), and Landar (1977). Most of the 20th century documentation came from speaker Hoxie Simmons (b. 1870s), and his retention of Galice idiosyncrasies is shown in Table 4.15:[25]

| Galice (Dorsey 1884b) | Galice (Hoijer 1973) | Euchre Creek (Golla 2008) | gloss |
|---|---|---|---|
| <kwûn> | kʷənʔ | xʷənʔ | 'fire' |
| <kwe'> | -keʔ | -xeʔ | 'foot' |
| <t̨s̨añ-kĕ> | ts'ã:ke: | t̨s̨'axe | 'woman' |
| <tçli$^{l}$-i$^{n}$-tcŭ> | ɬĩʔ-tšoh | ɬiʔ-tšuh | 'horse' |
| <tû$^{l}$-dĕ> | dəde: | dəne | 'person' |
| <da$^{l}$-ĕ> | -da:i | -nəɣeʔ | 'eye' |
| <-ts'a-dĕ$^{l}$> | -ts'ədeʔ | -ts'əneʔ | 'leg' |

Table 4.15: Dialect features retained in Galice (20th c.)

The point is simply that the Galice dialect retained the most salient features setting it apart from other members of the Rogue River group even among Stage III speakers raised in the

[25] Barnett (1934) didn't transcribe nasalized vowels with Simmons or with Nettie West, the other Galice speaker he worked with, but his transcriptions are most likely in error since they disagree with all other researchers who worked with Simmons. West, whose speech has the same features as Simmons', didn't move to Siletz until the age of ten, so she might have been unlikely to acquire the Oregon Athabaskan koine in any case, given that people exposed to new dialects in late childhood are less likely to acquire them accurately, if at all (cf. Labov 2007:349-350). Note, however, that Chetco speaker Billy Metcalf moved to Siletz in late childhood as well but apparently did acquire competence in the reservation koine.

mixed dialect community at Siletz. Moreover, this does not seem to be due to a lack of awareness of differences or of competence in Tututni (i.e., the Oregon Athabaskan koine known by that name), as comments by Simmons paraphrased in Harrington's field notes suggest:

> "It is funny – these coast Inds. can't talk our Gal. lang, but they can understand what we say. Hox can carry on a conversation in the coast language & can give the equivalent of almost every Gal. word in the coast language, but the coast Indians rarely indulge in trying to say a few Gal. words and then say them poorly." (Harrington 1981, reel 28, frame 42)

This statement indicates that Simmons had awareness and command of the language of the 'Coast Indians' (Tututni), and moreover that there was an asymmetry in accommodation, with the speaker of Galice competent in Tututni but not vice-versa. Nonetheless, this does not seem to have had a significant impact on the structural features of Galice.[26]

## 4.2. Mikwanutni

As noted in §1, Golla (1976) found no significant non-lexical differences between the Lower Rogue dialects spoken at Siletz in the early 1960s. There is, however, evidence for one feature distinguishing the Mikwanutni dialect from the emergent Siletz Athabaskan koine in the 20th century, one that may be a continuation of an older dialect difference. In pre-contact times, Mikwanutni was spoken in a village just a few miles up the Rogue River from the village where Tututni originated, and documentation suggests that these dialects were for the most part indistinguishable from one another. However, Dorsey (1884f) transcribed <ç> in a number of words before onset <s>, <c> (= [š]), and <ts'> in Mikwanutni but not in Tututni. This is shown in the first six forms in Table 4.16; the remaining forms in the table show that this was not the case for all onset coronal fricatives, even 'child' (where Dorsey had transcribed both initial [s] and [ts] in Lower Rogue dialects in the 1880s – cf. Table 8 above). Given that Dorsey transcribed only a few dozen Mikwanutni words, the density of this exceptional <ç> before <s> is striking; it is found almost nowhere else in his documentation of Oregon Athabaskan.

[26] That this was not always the direction of accommodation is seen in a comment by Mikwanutni speaker Miller Collins, who was married to the daughter of Galice speaker Nettie West: "Miller says he had to talk so much Gallice Ck. lang. with Silvy & Nettie that now he is talking with Mrs. Ada Collins he has to learn his mí·kʊnú· language" (Harrington 1981, reel 27, frame 554).

| Mikwanutni (Dorsey 1884f) | | Tututni (Dorsey 1884i) | | gloss |
|---|---|---|---|---|
| <çsi> | <ç>si | <çi’> | θi’ | ‘head’ |
| <çsŭ-gă$^{l}$> | <ç>səgə | <çŭ$^{l}$-ga> | θəga | ‘hair’ |
| <çsû-rxĕ$^{l}$> | <ç>səɣe | <sû-$^{r}$xĕ$^{l}$> | səɣe | ‘ear’ |
| <çsê$^{l}$-rxĕ> | <ç>seɣe | <qo-sĕ$^{l}$qĕ> | xosexe | ‘saliva’ |
| <ts’a$^{l}$-çci> | ts’a-<ç>ši | <ts’a$^{l}$-sĭ> | ts’asi | ‘widow’ |
| <çts’ĕ> | <ç>ts’e | <qwa-ts’ĕ$^{l}$> | x$^{w}$ats’e | ‘(his) navel’ |
| <sa$^{l}$-çlu> | sału | <sa$^{l}$-çlu> | sału | ‘tongue’ |
| <sqê$^{l}$-qĕ> | sxexe | <ske$^{l}$-ke> | skeke | ‘child’ |
| | | <tsqĕ$^{l}$-qĕ> | tsxexe | ‘child’ |

Table 4.16: Exceptional <ç> transcribed by Dorsey in Mikwanutni (1880s)

The phonetic interpretation of Dorsey’s <ç> here is uncertain. Its usual value is [θ] in the Bureau of American Ethnology alphabet Dorsey was using, as in the Tututni forms for ‘head’ and ‘hair’ in Table 4.16. However, Dorsey also used <ç> in ad hoc digraphs he invented to represent sounds not accounted for in the BAE alphabet, such as <çl> for [ɬ]. That Dorsey might have intended some sort of quasi-affricated quality is suggested by its appearance in his transcriptions of ‘navel’: Mik <çts’ĕ> (Dorsey 1884f), PiR <çs‘ĕ> (Dorsey 1884m, practically the only example of <çs> outside of Mikwanutni) alongside Tut. <-ts’ĕ> (Dorsey 1884i), cf. Euchre Creek *-ts’eʔ* in Golla (2008); the appeal of this interpretation will be apparent shortly.[27] While Dorsey’s precise intent for this letter is uncertain, it is clear enough that in the 1880s Dorsey sometimes detected a difference between coronal fricatives in Mikwanutni versus Tututni and other Lower Rogue dialects that warranted a special representation.

Mikwanutni as spoken by Miller Collins (b. 1880s) in the 20th century – tape-recorded by Morris Swadesh in the 1950s, transcribed and published by Landar (1977) – has a number of words with initial affricates where other dialects have a plain fricative. These are shown in Table 4.17. Merger of Proto-Athabaskan *ts, *dz, *s, and *z as /s/ (/θ/ in some varieties) is one of the hallmarks of Oregon Athabaskan (Sapir 1914, Golla 1976; #8-9 in Appendix A to chapter 2). Remarkably, Collins produced [ts] or [tθ] in nearly all of the words for which Proto-Athabaskan reconstructions with *ts are available, the main exception being [sən] ‘star’ (reconstructed by Krauss 2005 variably with or without an initial affricate). Collins also tended to produce [ts] in ‘child’, a word that exhibited a great deal of variation at Siletz in the 1880s (cf. Table 4.8). In many cases Collins’ affricate is found alongside [s] or [θ] as a variant, e.g., the root meaning ‘head/hair’ in the compound [θi:dogeʔ] ‘grey-haired’, or the possessed noun [tš’əθən] ‘some animal’s meat’. The last four forms in Table 4.17, ‘star’, ‘skin’, ‘liver’, and ‘ear’ show that Collins did not always produce onset coronal fricatives with affrication, in one

[27] Dorsey’s Mikwanutni vocabulary also includes <nû-rxût$^{l}$-sûs> [nəɣətsəs] ‘upper eyelid’ alongside <nû-rxĕ$^{l}$ ya-rxĕ$^{l}$-ăn sûs$^{l}$> [nəɣe yaɣe’ən səs] ‘lower eyelid’. <nû-rxû-> ~ <nû-rxĕ-> here is equivalent to Tutuni *naɣe* ‘eye’, and <sûs> is *səʔs* ‘skin’ (< P-A *zƛ’s’). The <t> in <nû-rxût$^{l}$-sûs> ‘upper eyelid’, then, can be considered another affricated onset in Mikwanutni (cf. Tut. <qo$^{l}$-na-ɻĕ$^{l}$ sus> [xonaɣe səs] ‘someone’s upper eyelid’, Dorsey 1884i). But notice that <t> is not transcribed in ‘lower eyelid’, which has plain <sûs>.

case (‘ear’) where Dorsey had transcribed <çs> with speakers in the 1880s.[28]

| Mikwanutni (Landar 1977) | Euchre Ck. (Golla 2008) | gloss | Proto-Athabaskan |
|---|---|---|---|
| tse:(ʔ) ~ tθe: ~ tsi:(ʔ) | siʔ | ‘head, hair’ | *tsiʔ |
| θi:dogeʔ | si-təlge | ‘grey-haired’ | |
| tse:ɣoɬ | siɣʷəs | ‘brain’ | *tsi:ɣa̱:ỹʔ |
| tθən ~ tsən | sənʔ | ‘meat’ | *tsəỹ’ |
| -θən in *tṣ̌’əθən* | | ‘some animal's meat’ | |
| tse: ~ θe: | se | ‘stone’ | *tse: |
| tsa:ɬo: ~ tθa:ɬo: | saɬuʔ | ‘tongue’ | *tsu̱:(ləʔ) |
| tsxei ~ tsei ~ sxei | ṣxe | ‘child’ | - |
| sən | sin | ‘star’ | *tsə̱w̃, *sə̱w̃ |
| səθ | səʔs | ‘skin’ | *-zə̱ts’ |
| -θəd in *tṣ̌’əθθəd* | - | ‘its liver’ | *-zə̱ť’ |
| səɣe: | ṣəɣeʔ | ‘ear’ | *džəɣəʔ |

Table 4.17: Onset coronal fricatives in Miller Collins’ Mikwanutni (1950s)

Although the Euchre Creek equivalents in Table 4.17 are on the whole typical of other Rogue River dialects, affricated [ts] and [tθ] were occasionally produced by speakers of varieties other than Mikwanutni as well. Landar (1977), for example, transcribed in Hoxie Simmons’ Galice both [tθe:] and [ṣe:] ‘stone’, [tsi:ʔ] and [ṣi:h] ‘hair’, and [wastsə̃ʔ] ‘its meat’. Sapir (1914) and Harrington (1981) transcribed (normalized) [tsxa:xe:] and [tsxe:xe:] respectively for ‘child’ in Wolverton Orton’s Chasta Costa. Thus, affricated coronal fricatives were found outside of Mikwanutni on occasion, although Mikwanutni appears to have had them most consistently. The variation associated with [ts] and [s] in the 20th century is reminiscent of that found with velar obstruents in Tututni in the 1880s: there is a lack of consistency both across varieties and in the speech of individual speakers. The significant fact is that whereas the [k] ~ [x] variation had largely leveled out by the mid-20th century, the [ts]~[s] variation seems to have remained in a state of flux, speakers not having converged on a single value.[29] Moreover, although it remains unclear whether the affrication of Miller Collins in the 20th century can be fully equated with the exceptional <ç> of Dorsey (1884f), such an interpretation is plausible.

[28] Landar (1977) also transcribed Collins producing [ɬtθe:] and [ɬtse:] ‘dry’ (P-A *tsa:y) and [ɬθo:] ‘yellow’ (P-A *tsʊxʷ), the former with an affricated onset and the latter without. The affrication in ‘dry’ could be due to an independent process, identified by Harrington (1981) in Upper Coquille, of affrication separating clusters of fricatives like /ɬs/ created at morpheme boundaries (e.g., [ɬtθuwwi] ‘yellow’, [ɬtθe:] ‘dry’, both beginning with a prefix *ɬ-*). The process applies more often than not in Harrington’s notes, but apparently not exceptionlessly, as would be the case here with Collins’ ‘yellow’.

[29] See Li (1930a) and §2.1 in chapter 2 for discussion of similar variation involving coronal fricatives and affricates in Mattole.

If correct, Mikwanutni appears to have maintained from the 1880s into the mid-20th century a stronger tendency to affricate onset coronal fricatives than other dialects did.

## 5. Fronting of /s/

In his description of Chasta Costa, Sapir (1914) observed that Proto-Athabaskan *ts and *s had, with some exceptions, merged as a single fricative (#9 in Appendix A to chapter 2). Sapir identified this fricative as [θ], noting that [s] was found instead in a few cases.[30] Hoijer (1960) considered Chasta Costa [θ] to be a "unique development"; in other varieties, according to the data Hoijer had at his disposal, the outcome of the merger was [s] alone. However, consideration of published and unpublished data (some of it unavailable to Hoijer) reveals a considerably more complicated picture. This is illustrated in the following normalized transcriptions for 'stone' in various sources:

| form | dialect & source |
|---|---|
| θe: | Coq (Harrington 1981) |
| se: | Euc (Golla 2008) |
| θe | Tut (Dorsey 1884i) |
| ðe | Tut (Everette 1882) |
| se-ət'-tša 'stone pipe' | Tut (Everette 1882) |
| se: | ChC (Sapir 1914) |
| ṣe:, θe: | ChC (Harrington 1981) |
| tse:, θe: | Mik (Landar 1977) |
| tθei, sei, θe: | Gal (Landar 1977) |
| se | PiR (Dorsey 1884m) |
| se: | Tol (Bright 1964) |

Table 4.18: 'stone' in Oregon Athabaskan (P-A *tse:)

Table 4.18 shows that a dental, possibly interdental, fricative was transcribed by many researchers in both the 19th and 20th centuries, often alongside [s]. Individual speakers were prone to produce both variants: all of the Mikwanutni, Galice, and Chasta Costa transcriptions in table 4.18 come from Miller Collins, Hoxie Simmons, and Wolverton Orton respectively. This same variation is found with many other lexical items as well. Like the variation involving initial affrication in Mikwanutni just considered in §4.2, there seems to be a wide range of intra- and inter-speaker variation.

The question, then, is whether the Oregon Athabaskan category usually represented as /s/ in phonemic analyses (and in related affricates such as /ts'/) was typically realized as alveolar [s] or as fronted (dental/interdental) [θ], as in Sapir's documentation of Chasta Costa. Published

[30] It is unclear whether Sapir intended his <θ> as dental or fully interdental – he describes the glottalized affricate written <tθ!> as "dental," but illustrates the plain fricative <θ> with "English *thin*" (1914:283). As his analysis and choice of symbols implies, he clearly considered the Proto-Athabaskan segments *s and *ts to be something other than dental. However, some scholars have analyzed these segments as being dental (not interdental) in the proto-language, e.g. Krauss and Golla (1981:71) and Leer (1996, cited in Hargus 2010).

descriptions of Oregon Athabaskan languages as spoken in the 20th century vary in this regard. Hoijer (1966:320) describes Galice [s] simply as "alveolar," and M. Jacobs (1968) and E. Jacobs (1968, 1977) transcribe only [s], never [θ], in the short Galice and (Tututnized) Chetco texts they published. Harrington's Tututni wordlist published in Landar (1977) similarly contains no tokens of [θ]. E. Bright (1964) described Tolowa's /s/ and /ts'/ as "blade-alveolar," the same description that W. Bright (1952) used for English /s/; the Tolowa Language Dictionary (Bommelyn 1989) does not include the equivalent of [θ] in the Unifon alphabet it uses.

However, most other sources give some indication that there was variation involving dental and alveolar realizations of /s/. Golla (1976) describes Ida Bensall's Euchre Creek /s/ and /ts'/ as "apical and tend[ing] toward a forward (dental or interdental, lisped) articulation" (1976:218); in his recent overview he implies that this articulation was widespread in Oregon Athabaskan (Golla 2011:75). In transcribing Hoijer's own recordings of elicitation sessions with Galice speaker Hoxie Simmons, Landar wrote [θ] in several words (including 'stone' in Table 4.18) and stated that "Mr. Simmons used a dentalized t alone, glottalized or in clusters ([t̪], [t̪ʔ], [t̪θ]). *His plain s is also dentalized*" (1977:294, emphasis added); Landar also frequently transcribed [θ] in Miller Collins' Mikwanutni, as seen in Table 4.18 above. Harrington's field notes for all of the Oregon Athabaskan varieties he studied (Upper Coquille, Chasta Costa, Tututni, Chetco-Tolowa, and Galice) are rife with [θ] as well, plain [s] being vanishingly rare. At one point he describes a geminate token of Coquelle Thompson's [θ] as "hard to hear - possibly s̪s̪. His θ is always in reality s̪s̪" – by which Harrington probably meant dental [s̪] (1981, reel 25, frame 896; cf. Krauss 1973:921). The bulk of the evidence, then, suggests that a fronted articulation of /s/ as dental [s̪] or interdental [θ] was common among speakers of Oregon Athabaskan languages living at Siletz in the early and mid-20th century.

Interestingly, transcriptions plausibly interpreted as fronted [θ] are rare in 19th century documentation of Oregon Athabaskan. The Upper Umpqua wordlists in Scouler (1841) and Hale (1846), the earliest documentation of any Oregon Athabaskan language, appear to be devoid of [θ]. Provision was made for this consonant in the Bureau of American Ethnography alphabet used by Everette and Dorsey, and literate speakers of English would have had orthographic <th> at their disposal. But <th> is found only once in the Hubbard wordlist (Taylor 1860) considered above, probably a typographical error.[31] Everette (1882) almost never transcribes a dental fricative in Tututni, his entry [ðe:] for 'stone' in Table 4.18 being a rare exception. Dorsey (1884a-m) also rarely transcribes [θ] (BAE<ç>) in his Oregon Athabaskan vocabularies. It occurs only twice in transcriptions of Chasta Costa words in Dorsey (1884c) (<çliˡ-çu̪taçl> [ɬi θutaɬ] 'spurs' and <tcŭt-laçˡ-nĕ> [tšət laθne] 'rattlesnake'), a striking gap given Sapir's description of the same dialect a generation later working with Wolverton Orton, a member of the same family as one of Dorsey's consultants (Jake Orton).

However, while [θ] is rare overall in Dorsey's transcriptions of Tututni, most of the tokens that do occur are found in words associated with a single speaker named Eneati (also known by the English name "Shellhead"). Examples include the following, the first three explicitly attributed to Eneati, the others from elicitation sessions where Eneati was likely to have been present:

[31] The form in question is <thun>, glossed 'free'. This is probably a two-way typo. The rest of the list is basic vocabulary, mostly nouns and numerals with only three verbs and no adjectives; 'free' seems aberrant in such a list. The target is probably 'tree', the expected form for which would be <chun>, i.e., [tšən].

| Tututni (Dorsey 1884i) | | gloss |
|---|---|---|
| <ça$^{l}$-tlu> | θatɬu | 'tongue' |
| <çe-> [32] | θe- | 'stone' |
| <t'çi-ĭlĕ> | tθ'ile | 'elbow' |
| <çi'> | θiʔ | 'head' |
| <çût$^{ʻ}$> | θət | 'liver' |
| <cçi-yu-wăs> | šθiyuwəs | 'my brain' |

Table 4.19: Examples of Eneati's [θ] in Dorsey (1884i)

In several of these words, Dorsey's transcription of <ç> overwrites an original <s>. That is, in many of these cases Dorsey first thought he heard [s], then decided [θ] was more accurate, suggesting a dental [s̪] that might have been articulated as fully interdental with a careful articulation when repeated under elicitation. In a couple of cases these [θ] forms can be compared with the same or related words produced by other speakers. Where Dorsey transcribes [θ] for Eneati other speakers have [s], e.g., Shem Lafayette's <qwû-ts'i$^{l}$-ĭ-lĕ> [x$^{w}$ətsʾile] '(someone's) elbow' and Charles Shellhead's <sa-çlu> [saɬu] 'tongue'. The latter of these is especially significant: Charles Shellhead was probably Eneati's (Shellhead's) son, so this form suggests that the younger man had a pronunciation subtly different from his father's, matching instead the wider Athabaskan-speaking community at Siletz.

In summary, nearly all of the words Dorsey transcribed with [θ] in the 1880s came from this single speaker. We know from an interview with Lucy Metcalf, Eneati's daughter, that he was already an adult before relocation to Siletz in the mid-1850s and had been the owner of one of the sweathouses in his home village of Dutuden (the same one that *Tututni* is derived from) (du Bois 1934). He was estimated to be 70 years old in the census conducted at Siletz in 1885 (Bureau of Indian Affairs 1967). That he might have been more linguistically conservative than other people Dorsey worked with is confirmed by a note appearing alongside the entry for "saliva" in Dorsey's notes: "present generation" <qwo-sa-r̥e> [x$^{w}$osaɣe], "old men" <qwo-sĕ$^{l}$qĕ> [x$^{w}$osexe]. The latter form is attributed to Eneati, then written a second time <qwo-çĕ$^{l}$-q$^{ʻ}$ĕ> with <ç> (= [θ]) instead of <s>. The impression is of an older man, probably wealthy and influential prior to relocation, already an adult when the reservation was established at Siletz and somewhat conservative in his speech. He often produced what Dorsey heard as [θ] where other men, most of them younger and having spent more of their lives on the reservation, produced a sound that Dorsey usually heard as [s].

What, then, is the explanation for the mismatch between the pervasive fronted articulation of /s/ transcribed with speakers in the 20$^{th}$ century versus its virtual absence from 19$^{th}$ century documentation (with the exception of tokens produced by Eneati)? One possibility is that there was a consistent dental articulation in both periods, and the difference merely reflects better transcriptions by trained linguists in the 20$^{th}$ century. However, it seems implausible that so many researchers in the 19$^{th}$ century would have transcribed a dental or interdental fricative as [θ] so seldom. This is especially the case for Dorsey. A native speaker of English who had previously spent many years working on Siouan languages with phonemic /θ/ (Dorsey 1891), Dorsey of all the early researchers should have been sensitive to the difference between [s] vs. [s̪] or [θ]. A more likely explanation is that [θ] was an innovative articulation of /s/ that spread

[32] In <çe-tcûn tûn-nĕ>, lit. 'people at foot of big rock'.

in the reservation period. The articulation transcribed as [θ] by most scholars in the 20th century was not widespread at Siletz in the 1880s, but became so as a result of diffusion among the Stage III speakers born at Siletz in that period.[33]

One potential explanation for why this change occurred when it did (addressing the actuation problem of Weinreich, Labov, and Herzog 1968) is that it is due to the increased prevalence of English and Chinook Wawa spoken at Siletz in the late 1800s. At the time of contact in the mid-19th century, Oregon Athabaskan had a three-way place of articulation contrast involving coronal fricatives: /s/ (from Proto-Athabaskan *s, *z, *ts, and *dz), retroflex /ṣ/ (from Proto-Athabaskan *tš and *dž), and /š/ (from Proto-Athabaskan *š, *ž, *x̂, and *ŷ). As noted in §4.2, Golla (1976) found the distinction between plain /s/ and retroflex /ṣ/ "subtle." As speakers increasingly relied on Chinook Wawa and English in their day-to-day interactions, the subtle contrast between plain /s/ vs. retroflex /ṣ/ might have become harder for people to distinguish. One strategy for maintaining contrasts in such circumstances is via so-called categorical "transfer" to a functionally similar contrast in another language (Babel 2009b). Adopting a fronted articulation of /s/ as [θ] might have been a convenient way to maintain the phonemic contrast. Such transfer effects are typically associated with incipient language shift. Under this view, the fronted articulation of Oregon Athabaskan /s/ that diffused at Siletz was due to contact with English and can be taken as a sign of impending language obsolescence.

However, this explanation fails to account for an important aspect of the data: Eneati, the person who seems to have been an innovator of the change, was an older man whose speech patterns were established well before contact with English-speaking populations. Eneati is known to have communicated with Dorsey through an interpreter, and he probably spoke little English, if any. Such a monolingual speaker is not a candidate for the maintenance of contrasts through categorical transfer effects of the kind discussed by Babel. It is therefore likely that his fronted articulation of /s/ had nothing to do with English bilingualism. It may have been a true innovation on his part, but it is also possible that his fronted /s/ was archaic: as noted above, Proto-Athabaskan *s and *ts are considered by many Athabaskan scholars to have been dental, not alveolar (Krauss and Golla 1981:71, Leer 1996), and Eneati, the first speaker documented who consistently produced the fronted articulation, was an older man whose speech was conservative in other respects.

In this scenario, the diffusion of Eneati's fronted /s/ in the reservation community can be interpreted as a reversal of a sound change that was in progress at the time of contact, perhaps retraction of original dental /s/ leading to near-merger with /ṣ/. The spread of the fronted articulation may still have been functionally motivated due to confusability of /s/ and /ṣ/, but the solution adopted to resolve the pressure was to exploit a conservative variant that was already part of the language. While this might have been reinforced by categorical transfer effects from increasing bilingualism in English, it may instead simply be a case of internally-motivated language change that happened to coincide with the Oregon Athabaskan languages ceasing to be acquired in subsequent generations.[34] Whichever scenario is correct, this seems to

[33] A similar change of /s/ to /θ/ in the same region in the historical period is reported by Conathan (2006) for Karuk, a non-Athabaskan language of northern California. The circumstances in which the Karuk innovation spread are quite different than what is found at Siletz, since it did not occur in a reservation setting. Nonetheless this is an interesting parallel with the case at considered here.

[34] There may also be a social dimension to this innovation. The situation at Siletz can be compared with Hinton's (1979, 1981) analysis of similar data in the Yuman languages of California and Arizona. Archival documentation

be a case of a minority variant diffusing among younger speakers who were raised in the mixed dialect community, a point revisited below.

## 6. Discussion

The main findings concerning koineization among the Oregon Athabaskan languages can be summarized as follows. First, certain differences between closely related varieties of Oregon Athabaskan spoken at Siletz were the targets of leveling in the reservation period, with [k] ~ [x] variation becoming more or less uniformly [x] in all but Galice in the 20th century. Second, differences between more distantly related Oregon Athabaskan varieties underwent leveling: Chetco in the 20th century had no nasalized vowels, coda [l] instead of [tɬ], and initial [d] instead of [dž] in 'man/male' (but [dž] still found in 'elk'). Third, some dialect differences resisted leveling: Galice as spoken by Hoxie Simmons and Nettie West in the 20th century retained a number of idiosyncrasies, including /k/ instead of /x/, denasalized *n before oral vowels, nasalized vowels, and lenition of intervocalic /ɣ/. Even some differences between closely related Lower Rogue dialects were maintained: Mikwanutni as spoken by Miller Collins had initial affricated [ts] instead of [s] in many words, plausibly a continuation of a dialect difference involving coronal onsets also manifest in the documentation of Dorsey (1884f). There is also evidence from the same period that a linguistic innovation (or perhaps a reversion to a conservative variant), a fronted articulation of Oregon Athabaskan /s/, spread through the reservation community.

As foreshadowed in previous sections, these data bear on issues of current interest in the study of contact-induced language change, especially the degree to which the outcomes of dialect contact are deterministic. Trudgill has argued that they are, in gross outline at least, in cases of tabula rasa contact where "there is no prior-existing population speaking the language in question, either in the location in question or nearby" (2004:26). He maintains that koineization in the development of southern hemisphere dialects of English proceeded according to a principle of majority rule: variants that were common to most dialects survived in the koine and minority variants perished, such that "the minority quite simply and mechanistically accommodated to the minority" (2004:148). There can be no doubt that the relative proportion of speakers in a population (or, more precisely, of particular variants in a mixed dialect population) is an important consideration in many, perhaps most, language contact situations: as Thomason (2001:66) puts it, "numbers count." The question is whether they always count equally, or whether entrenched social attitudes can alter the outcomes in unpredictable ways, even in the special tabula rasa circumstances that Trudgill (2004) considers.

Although Trudgill initially takes pains to limit his arguments to tabula rasa scenarios (2004:27), he situates his claims in a broader debate about the extent to which people's attitudes concerning variation matter in determining the course of language change (2004, ch. 7).

---

of those languages shows a shift of s (or s̱) > θ that occurred in separate branches of the Yuman family in the period 1870-1890. Hinton suggests that an innovative [θ] articulation started with Irataba, a highly influential Mojave leader, and subsequently spread to other Yuman-speaking communities as a result of intensive contact on the Colorado River reservation in the 1870s. A similar account can be entertained here: a shift to [θ] may have started with Eneati or some other influential individual among the Athabaskans living at Siletz (of whom Eneati may have been an early imitator), subsequently spreading to other members of the Siletz community. Such an account explicitly embraces the possibility that the spread of innovations might be affected by the relative prestige of the innovators (cf. Fagyal et al. 2010).

As discussed in chapter 1, for Trudgill and others working in this framework, long-term outcomes of dialect contact are the cumulative result of accommodation between speakers in micro-interactional settings. The extent to which people actively use their knowledge of the social world in tuning their speech to interlocutors is an area of current research (cf. Babel 2009a, 2012 for a recent example). Trudgill adopts a mechanistic view attributed to Labov (2001) whereby metalinguistic evaluations of variants (or people who employ them) play only a minor role in delimiting the diffusion of linguistic changes. Such evaluations might, of course, influence how frequently people interact with each other, and thereby indirectly affect the outcomes of language change, but under this view they have little or no direct impact on linguistic behavior when interaction occurs. This is at odds with the family of indexical approaches to language variation and change, in which, broadly speaking, linguistic choices are influenced by people's perceptions of their social landscapes and the positions they occupy – or wish to occupy – within them (cf. Eckert 2008, Thomason 2007, Fagyal et al. 2010, Hill 2001). In these approaches, locally-defined social-indexical values associated with variation within a population are crucial to understanding the diachronic trajectories of sociolinguistic variables.

As noted in §1, the Oregon Athabaskan case considered in this chapter satisfies Trudgill's definition of tabula rasa dialect contact. The Siletz reservation was established in northwestern Oregon near Corvallis, no small distance from the nearest homeland of any of the Athabaskan groups residing there. While the distance between place of origin and locus of resettlement is not as extreme as in the cases of overseas colonial resettlement studied by Trudgill, upon arrival in Siletz no Athabaskan group had any special claim to legitimacy there.

Nonetheless, the outcomes of contact between Athabaskan varieties that participated in mixing at Siletz are not predictable based strictly on the relative proportions of dialect features in the population as a whole. Table 4.20 presents the census figures for the Oregon Athabaskans living at Siletz from available government sources:[35]

[35] Data are drawn from the Office of Indian Affairs report for each of the years listed. The 1855 report gives the population of villages counted in situ, before relocation to Siletz; the number of Chasta Costa adult males is missing for that year and is here estimated at 132 (the remainder after subtracting all other groups from the total number of adult males). Population tables for all years except 1855 list simply "Coquill," so it is unclear whether Upper Coquille (Athabaskan) or Lower Coquille (Milluk Coos) is intended; comments in Everette (1882) suggest that the 1882 figure is for Milluk rather than Upper Coquille. The sharp decline in population in the 1870s is likely due to the fact that people started leaving the reservation to seek their livelihoods elsewhere (Kent 1973:23, Leavelle 1998:449).

| | | 1855 | 1858 | 1861 | 1865 | 1878 | 1882 |
|---|---|---|---|---|---|---|---|
| Athabaskan | Coquille | 105 | [313] | [221] | [142] | [84] | [114] |
| | Sixes | 143 | 242 | 218 | 320 | 74 | 53 |
| | Euchre Creek | 102 | 84 | 177 | 150 | 59 | 40 |
| | Tututni | 120 | 202 | 168 | 227 | 137 | 83 |
| | Joshua | 120 | 179 | 188 | 250 | 84 | 44 |
| | Mikwanutni | 135 | 129 | 247 | 348 | - | 40 |
| | Chasta Costa | [132] | 110 | 214 | 162 | 47 | 55 |
| | Pistol River | 51 | - | 71 | 161 | 57 | 33 |
| | Galice | - | - | - | - | 18 | 37 |
| | Chetco | 241 | 215 | 262 | 221 | 63 | 69 |
| Other | | - | 575 | 259 | 331 | 466 | 430 |
| total | | 1149 | 2049 | 2025 | 2312 | 1089 | 998 |

Table 4.20: Population of Siletz, 1850s-1880s

These numbers must be used with some caution, since it is unclear how the ethnic determinations were made, and the extreme fluctuations in some cases may be due more to administrative whim than to how the people being counted might have chosen to self-identify.[36] Nonetheless, the figures in Table 4.20 offer a rough sense of how many speakers of each variety resided at Siletz in the years after the reservation was established.

The key things to notice in Table 4.20 are that the Chetco, Tututni-Joshua, and Mikwanutni groups were among the most numerous on the reservation while the Galice, for the years in which figures are available, were a minority.[37] It has already been noted in §3.2 that the outcomes for Chetco, whose distinctive dialect features didn't survive the leveling associated with koineization, are probably as predicted by Trudgill's deterministic theory. Although the Chetco were one of the largest groups at Siletz, they were outnumbered by speakers of the Rogue River dialects that lacked nasalized vowels, coda [tɬ], and onset [dž] in the relevant vocabulary items. These features were at a disadvantage for selection in the emergent koine and for the most part were not retained, even though the ethnolinguistic label "Chetco" survived.

Trudgill's deterministic theory, does not, however, explain how idiosyncratic features were maintained in Galice and Mikwanutni. It is surprising that Galice's distinctive properties remained in the 20th century at all, given that the Galice contingent was one of the smallest at Siletz. Moreover, while it is true that the Mikwanutni were a large group, in most years for which data are available they were no more numerous than the Chetco. Nonetheless, if the interpretation of the data presented in §4.2 is correct, Mikwanutni speakers in both the 1880s

[36] See Hall (1992) for discussion of the problem of assigning people to discrete ethnic categories where identities were probably considerably more fluid.

[37] Although it is known from government reports that some Galice people were present at Siletz at least as early as the 1860s, they were not counted separately until the late 1870s, 20 years after the establishment of the reservation.

and 1950s had a tendency to affricate coronal fricatives. In both cases (Mikwanutni and Galice vis-à-vis Chetco), population asymmetries do not explain the outcomes of contact. If a group as small as Galice maintained its distinctiveness, Chetco should have too; if a group as large as Chetco succumbed to leveling pressure, it is surprising that Mikwanutni speakers didn't do so as well. One cannot simply count the number of speakers of Galice, Tututni, Chetco, and Mikwanutni and anticipate the outcomes of contact – something needs to be said about the social dynamics at play when the contact took place, including among younger generations of speakers raised in the mixed dialect community.

It is also worth pointing out that structural distance does not seem to be an adequate explanation for the outcomes of dialect contact at Siletz. As often noted in the language contact literature, structural/typological differences between varieties in contact can impede (if not prevent entirely) some kinds of contact effects (cf. Thomason and Kaufman 1988, Winford 2003). Perhaps Galice's idiosyncrasies were maintained because, as discussed in chapter 1, it was structurally divergent from other Rogue River varieties spoken at Siletz. However, this would fail to account for the fact that Chetco's distinctive traits didn't survive koineization, even though Chetco was divergent enough to be considered a language distinct from Rogue River varieties (unlike Galice in the classification of Golla 2011). Moreover, Mikwanutni is otherwise indistinguishable from other Lower Rogue dialects, providing no structural explanation for what appears to be the retention of a unique realization of onset coronal fricatives.

It becomes important that Galice speakers living at Siletz, in addition to their linguistic distinctiveness, also brought to the reservation a number of cultural practices that were significantly different from other Oregon Athabaskan groups. Their home territory on Galice Creek was adjacent to that of the non-Athabaskan Takelma. According to Drucker's "culture elements" ethnographic analysis, Galice people had in pre-contact times adopted many aspects of Takelma material and intellectual culture: "Probably never a large group, these people had become through daily contact and intermarriage almost wholly Takelman in culture ... [The Galice Creek culture] is so permeated with Takelman elements as to be scarcely distinguishable from the culture of these alien people" (Drucker 1937:283; cf. Barnett 1937, Gray 1987). The relevance of some of the key differences from the culture elements analysis may be disputed (e.g., whether the fact that the Galice people in aboriginal times did not build plank-walled houses or pierce their ears would be expected to have any bearing on linguistic behavior), but Drucker also points out that there had been a significant degree of intermarriage between the Galice and neighboring Takelman people. It may be reasonably suspected that such ties of kinship, more than superficial differences in attire, hunting practices, and so on, might have had a strong bearing on people's construction of identity in the aftermath of resettlement.

There is also evidence that the affinity between Galice and Takelma was a community-internal perception at Siletz: Dorsey (1884b), in a wordlist collected with Galice Creek Jim, paraphrased a Tututni speaker saying that "these are not real Galice Creek words, but Jim's pronunciation of Tutu words. The G.C. lang resembles the Takelma." Thus, while from a structural standpoint the Galice language is unmistakably Athabaskan, its speakers were considered more closely affiliated with Takelma-speaking groups. Their ongoing distinctiveness from other Athabaskan-speaking groups in the 20th century is suggested by Hoxie Simmons' use of the term "coast Indians" to refer collectively to the non-Galice Athabaskan groups at Siletz (quoted in §4.1), and two of Harrington's consultants suggested that Simmons most likely knew "quite a little" of the Takelma language because "he was raised by the old people" (1981, reel 27, frame 568).

This returns to Kerswill's observations about the relationship between koineization and social integration in a new community, mentioned in §2. While many Athabaskan groups at Siletz began to forge a new, focused identity whose linguistic consequence was koineization, the Galice and perhaps other groups may have maintained a quasi-exogenous status afforded to non-Athabaskans. The Galice way of speaking was an index of this status, as evidenced by the comments concerning Galice Creek Jim's language. An indexical approach to variation introduced through dialect contact invites a connection to be made between the maintenance of Galice's linguistic distinctiveness and the social and cultural differences separating them from other Athabaskan groups that obtained prior to the establishment of the reservation community and that apparently retained some relevance into the late 19th and perhaps 20th century. Whether such an explanation can be extended to account for the maintenance of Mikwanutni distinctiveness remains to be seen – the Mikwanutni homeland was located squarely among the other Lower Rogue dialect groups and there is no evidence that the Mikwanutni were culturally distinct in the same way as the Galice were. However, it seems a promising line of explanation to suppose that the linguistic outcomes of Galice's contact with Tututni and other Athabaskan varieties are due at least in part to the fact that they had to overcome a greater social distance from the outset.

However, as discussed in the context of Furbee's (1995) theory of dialect divergence in Chiwere (§3 of chapter 1), overcoming social distance is not to be equated with infrequent interaction. In the current context, Trudgill's demographic determinism could perhaps be salvaged by maintaining that Galice retained its distinctiveness because Galice speakers didn't interact regularly with other Athabaskan groups. However, while there was some degree of segregation according to ethnolinguistic divisions at Siletz, this mainly affected non-Athabaskan groups such as the Takelma. The historical record and reports from people who lived there suggest that the Athabaskan groups at Siletz were closely intermingled, sharing resources, intermarrying, and so on, making it improbable that Galice speakers would have interacted with Tututni speakers less frequently than, say, the Chetco did. Instead, Galice remained distinct because its speakers didn't simply and mechanistically accommodate to the majority variants in the sea of Athabaskan dialects engulfing them (even if they had some degree of command of those variants, as Hoxie Simmons apparently had). More generally, people's perceptions of the social context in which they use language and the linguistic resources they deploy in response to those perceptions are very much relevant to understanding the outcomes of linguistic contact, and not only in constraining the frequency of interaction between speakers – even in the tabula rasa cases considered by Trudgill.

A final point concerns the interpretation of the fronted realization of /s/ in Oregon Athabaskan. If the analysis presented in §5 is on the right track, this was a minority variant (perhaps innovative, perhaps archaic) that spread in the Siletz reservation community in the historical period. This is surprising under Trudgill's demographic determinism, since minority variants of any sort should not be able to spread in such circumstances. Perhaps the difference between this change and changes that did conform to a majority-rule principle under koineization (such as leveling that targeted Chetco's idiosyncrasies) is due to the fact that fronting of /s/ has a plausible functional motivation, undoing the effects of near merger with /ṣ/. It may be the case that dialect contact and koineization are special – that the diffusion of one variant at the expense of another under koineization in a newly-formed speech community proceeds according to different principles than diffusion due to functional pressure, and that in some cases the two types of change can proceed in tandem. But it seems preferable a priori to assume that the

principles governing both cases would be the same, and this points out perhaps the most serious limitation of Trudgill's demographic determinism. Even if his theory could be shown to be valid for tabula rasa contact, it offers little hope for understanding other phenomena in historical linguistics. If generalized, it essentially leads to a version of the Saussurean paradox: if minority variants are dispreferred, how is any linguistic change possible at all? Indexical theories of variation offer the possibility that the two phenomena can be related insofar as what matters is not the historical source of variation, but how people respond to it.

# Chapter 5: English and Hupa in Hoopa Valley

## 1. Introduction

Alongside the dynamics of dialect contact in Hoopa Valley discussed in chapter 3, the mid-19th century brought a simultaneous overlay of cross-linguistic contact between the Hupa and the English-speaking populations whose arrival in northwestern California transformed the indigenous societies they encountered. The earliest encounters between these societies were facilitated linguistically by white settlers who had acquired enough proficiency in local indigenous languages to act as interpreters (Gibbs 1853:145). However, Hupa people soon faced intense pressure to learn English in order to communicate with the agents of the military, bureaucratic, and economic apparatus of colonization that their lives were increasingly entangled with. Perhaps as late as 1850 no one born in Hoopa Valley spoke any English. By the mid-20th century, nearly all Hupa people spoke English, and children were acquiring no more than a passive familiarity with the Hupa language. Today, only a handful of fluent first-language speakers of Hupa remain, all of them elderly, although as noted in chapter 1 a growing number of learners has acquired a fair degree of second-language proficiency in their heritage language. Thus, the single greatest impact that the colonial encounter with English-speaking populations had on the Hupa language was to bring it to the brink of oblivion. In the terminology of Thomason and Kaufman (1988), the long-term profile of linguistic contact between Hupa and English is therefore one of language shift rather than maintenance.

The transition from Hupa to English in Hoopa Valley is reminiscent of "gradual death" in the typology developed in Campbell and Muntzel's (1989) now-classic paper on language endangerment. This is characterized by "an intermediate stage of bilingualism in which the dominant language comes to be employed by an ever increasing number of individuals in a growing number of contexts where the subordinate language was formerly used," often with an age-gradient of proficiency (1989:185). A generation after the establishment of the reservation in Hoopa Valley, acting administrator William Dougherty reported that "[t]hese Indians all speak English, many of them very well ..." (Office of Indian Affairs 1886:42), and in 1891 U.S. Indian Agent Isaac A. Beers stated that "the majority of them speak English to some extent, many of the younger ones quite fluently" (Office of Indian Affairs 1891:219). The age-gradient of bilingual competence noted by Beers is corroborated in the ethnographic description of Goddard (1905):

> "The older people speak their own language chiefly, having recourse to the few English words they know when communication with white people is necessary. The younger people all employ the Hupa language in their home life and when talking to each other, but have a good command of English for their intercourse with white people." (Goddard 1905:10)

Thus, by the early 20th century English had come to be the ordinary language used in an ever-widening sphere of everyday interaction in Hoopa Valley. Younger people, many of them educated in reservation schools (albeit coercively and often haphazardly) and doubtless more engaged in the local wage economy than older people, were on the leading edge of Hupa-English bilingualism in this period. Goddard's report also depicts a functional bifurcation of Hupa as the language used among Hupa people, with English reserved for communication with outsiders, suggesting a diglossic situation (in the "extended" sense of Fishman 1967). Howev-

er, Goddard (1911:151) suggests that English might have been more widely used by younger Hupa people, even in interactions with each other: "Some of the younger generation, who are nearly or quite bilingual, employ Hupa in giving directions about work to be done, or in relating events in which they wish place-relations to be plain, but English for ordinary social discourse."

By the turn of the 20th century, then, Hupa and English existed alongside one another as resources that Hupa people could draw on in their day-to-day lives. People were no doubt constantly making choices, implicitly or explicitly, about which resources to deploy in which settings. They moreover faced the task of developing linguistic conventions for communicating about the tangible and intangible trappings of an alien society that was radically different from any they had known previously. The present chapter considers how this transitional period of Hupa-English bilingualism, and functional pressure to adapt linguistically to a changing world, impacted the Hupa language.

Nowhere are these effects more evident than in the lexicon, and §2 considers in detail the strategies that Hupa speakers used to refer to the new items of material culture that suddenly suffused their lives. This discussion is framed under the rubric of lexical acculturation, understood to include the full array of strategies that Hupa speakers used to achieve nominal reference: borrowing words from English (which had some minor phonological side effects), extending the meanings of existing lexical items, and coining new words. The main empirical finding is that there are very few English loanwords in Hupa, a result that relates both to linguistic conservatism within the Athabaskan language family (Sapir 1921) and to the dynamics of lexical acculturation more generally discussed in Brown (1994, 1999). The discussion of the phonological adaptation of English loanwords in §2.1.2 also addresses a current debate concerning the role of L2 subphonemic information in the construction of L1 phonological representations.

Contact-induced changes in Hupa modeled on English in non-lexical domains – i.e., ones that are directly due to Hupa-English bilingualism – are harder to find, but §3 considers tendencies in the linear order of constituents in main clauses. Starting in the mid-20th century, Hupa speakers appear to use some ordering possibilities less frequently than earlier generations did. In particular, the use of postverbal subjects is greatly reduced in texts recorded after 1960. It is argued that although this change is probably due to contact with English, it is not necessarily the case that Hupa speakers altered their word order tendencies to better match an English pattern (i.e., this may not be bilingual interference under language shift as in Thomason and Kaufman 1988). Instead, the change is more likely to be a species of incipient language attrition whereby marked stylistic and pragmatic options come to be rarely used.

Overall, the main conclusion of the chapter, discussed in §4, is that people's proficiency in English appears to have had fairly little overall impact on how they spoke Hupa. This result is perhaps unsurprising given that there was a relatively short period of bilingualism before Hupa ceased to be widely used in Hoopa Valley, although §4 also highlights some outstanding questions and areas for future research.

## 2. Lexical Innovations

When Euro-Americans arrived in Hoopa Valley in the mid-19th century, Hupa people suddenly faced the challenge of referring to a plethora of material culture items that were previously unknown to them. Three main strategies for coping with functional pressure to achieve nominal reference are recognized in the literature on lexical acculturation (Haugen 1950; Brown 1994,

1999; Callaghan and Gamble 1996): borrowing words from another language, adapting the denotations of existing words (perhaps leading to semantic shift), and coining new words using productive grammatical resources. All three of these strategies were employed to some extent by Hupa speakers and are considered in turn in this section. The discussion is based on three main sources of documentation: a Bureau of American Ethnology vocabulary schedule with Hupa translations for 112 English glosses related to introduced items (Curtin 1888-1889:192-195), the Hupa Language Dictionary (Golla 1996b), and elicitation with contemporary Hupa speaker Verdena Parker (b. 1936), whose own prodigious command of Hupa and ability to coin words are complemented by her recollections of lexical items that were in common use when Hupa was still widely spoken in Hoopa Valley.

## 2.1. Borrowing

### 2.1.1. Loanwords

Despite the massive influx of new items introduced at the time of contact, there are only a handful of English loanwords documented in Hupa. Table 5.1 is an exhaustive list from the Hupa Language Dictionary (Golla 1996b); items marked with an asterisk are also found in the Curtin vocabulary (1888-1889). The words are rendered in the community orthography found in the dictionary, accompanied by an approximate broad phonetic transliteration:[1]

| orthographic | phonetic | gloss |
|---|---|---|
| bo:se | [bo:se:] | 'cat' (< 'pussy')* |
| da:la | [da:la:] | 'dollar' |
| 'e:bilos | [ʔe:bɪlos] | 'apples' |
| jigin | [džɪg$^{j}$ɪn][2] | 'chicken'* |
| Ka:n | [k$^{h}$a:n] | 'corn'* |
| Ka:whe | [k$^{h}$a:ʍe:] | 'coffee' |
| whilba | [ʍɪlba:] | 'wheelbarrow' |

Table 5.1: English loanwords in the Hupa Language Dictionary

Curtin (1888-1889) also transcribed the word <câl> [ša:l] for 'shawl'. Some of these loanwords ('chicken', 'corn', and 'coffee') are documented alongside neologisms using Hupa-sourced morphemes, cases discussed in §2.4.

Contemporary Hupa speaker Verdena Parker has confirmed the loanwords in Table 5.1 and

[1] Bear in mind that onset /b d g/ in the Athabaskanist transcription conventions used here are voiceless unaspirated consonants. Aspiration is explicitly represented in this section with IPA superscript [$^{h}$] in order to emphasize points of overlap with English pronunciation. It is important to point out that Curtin (1888-1889) deviated from the use of the Bureau of American Ethnology alphabet in transcribing <pussy> for 'cat' and <chickin> for 'hen' and 'rooster'. The appearance of initial <p> in <pussy> suggests Curtin might have heard aspiration, but could also just be his orthographic normalization of the Hupa pronunciation of the word (as indicated by the double <s> and final <y>). This could be the case for <ch> in <chickin> as well, as suggested by medial <ck>, but notice that Curtin did not normalize the final vowel in this word (and spells the English translation 'chicken' with final -en).

[2] This is written as <jikin> in the Hupa Language Dictionary, implying an aspirated front-velar stop /k$^{h}$/, but it is transcribed with <g> in Golla (1984), i.e., unaspirated /g$^{j}$/, which matches how the word is pronounced by Verdena Parker.

has also provided a number of other words that she remembers being in common use among Hupa speakers when she was young. Table 5.2 summarizes forms where the pronunciation offered by Mrs. Parker under elicitation differed significantly from her pronunciation in English:

| form | gloss | ref. |
|---|---|---|
| [binɪs] | 'beans' | VP-290513-05 |
| [bɪdžɪs] | 'peaches' | VP-110607-04 |
| [k$^{h}$a:nəpš] | 'turnips' | VP-230713-05 |
| [k$^{h}$ʌlɪk$^{h}$o:n] | 'calico fabric' | VP-290513-02 |
| [k$^{h}$e:na:sdɪk] | 'candle(stick)' | VP-300513-02 |
| [k$^{h}$e:ndowe'] | 'candy' | LA 256, 020_A[3] |
| [k$^{h}$ogɪs] | 'cookies' | VP-290513-05 |
| [k$^{h}$ondžɪs] | 'quince' | VP-290513-05 |
| [me:džɪs] | 'matches' | VP-300513-02 |
| [p$^{h}$lʌmɪs] | 'plums' | VP-290513-05 |
| [hi:] | 'tea' | VP-300513-05 |
| [t$^{h}$i:] | 'tea' | VP-290513-05 |

Table 5.2: English loanwords elicited from Verdena Parker

In some cases the forms given in Table 5.2 are Mrs. Parker's imitation of words as she remembers them being pronounced by older people. An example is 'peaches', where Mrs. Parker pronounces it as shown in the table when imitating old people she remembers from her youth, but uses an anglicized pronunciation – with aspirated [p$^{h}$] and tense [i] - in two tokens she produced in texts (VP-140309-06 and VP-300810-03). Mrs. Parker suggests that old people's way of pronouncing English-derived vocabulary was sometimes a source of amusement for younger people, who presumably had greater command of English phonology. Mrs. Parker has also suggested that several other English-sourced words ('cheese', 'chocolate', 'cocoa', 'figs', 'goat', 'peas', 'syrup', 'town') might have been in common use as well, but she did not offer distinctively Hupa pronunciations for them; these might be more akin to nonce borrowings or code-switching and are treated in §2.1.3 below, although it can be difficult to distinguish these phenomena from one another (Winford 2003:41).[4]

It should be noted that loanwords and related contact phenomena may well be underdocumented in Hupa, especially since many of the known loanwords were in common use alongside words composed of Hupa-sourced material. This is due in part to the emphasis in early research on salvaging as much as possible elements of Hupa language and culture as they existed prior to the colonial encounter. The grammatical description of Goddard (1905), for example, is based on specific examples from the texts in Goddard (1904), the focus of which is

[3] From the collection LA 256, *The Verdena Parker Collection of Hupa Audio Recordings*, archived with the Berkeley Language Center at UC Berkeley.

[4] It is also worth noting that there are no documented instances of Hupa borrowing English verbs. This is expected, of course, since verbs are cross-linguistically less likely to be loanwords than nouns are (Haugen 1950, Winford 2003:51). Moreover, Hupa's complex verbal morphology was arguably a typological barrier to adopting verbal loanwords (cf. Thomason and Kaufman 1988:72-74). By contrast, Hupa nominal morphology is relatively uncomplicated and would not have presented structural barriers to borrowing nouns.

almost entirely on traditional genres (myths, medicine formulae, and so on) where there is no opportunity to refer to items introduced recently. Moreover, Hill (1994) has pointed out that the context of elicitation itself might tend to foster purist linguistic stances in speakers who understand that the object of study is the pre-contact state of their language. Thus, where Hupa speakers had options available, their use of English-sourced material might have been less common under elicitation than it would have been in everyday conversation with other Hupa people.

Nonetheless, the small number of well-established loanwords in Tables 5.1 and 5.2 is striking. Brown (1994, 1999), in a comparative study of lexical acculturation across the Americas, develops two statistics to estimate the impact of lexical borrowing based on a standardized list of 77 glosses referring to items of "Occidental culture" introduced to the Americas by European colonizers. The first is a raw percentage of borrowed vocabulary: the number of loanwords divided by the number of items on the list for which data are available. The second is a convergence index calculated by subtracting the proportion of items labeled with indigenous resources from the proportion of items labeled with European loanwords. The resulting index ranges between -1 (items are labeled exclusively with indigenous-sourced material) to 1 (items are labeled exclusively with loanwords). Although Hupa was not included in Brown's study, there is documentation for 70 of the 77 items on his list. According to these metrics, Hupa has a borrowing rate of 20% and a convergence index of -0.67, both among the lowest of the indigenous languages of California included in Brown's study and well below the 63% borrowing rate that Brown reports as average for the region.

Brown (1994; 1999, ch. 6) seeks to explain regional differences in lexical borrowing rates across the Americas in terms of the incidence of bilingualism in different settings. Brown found that groups that came into contact with English- and French-speaking colonial powers tended not to use borrowing as a strategy for lexical acculturation, but rather relied on language-internal resources by adapting existing words or coining new ones. Building on similar observations by Bright (1960) based on a study of languages of California, Brown suggests that "in general speakers of English and/or French-influenced languages were significantly less bilingual than speakers of Spanish-influenced languages" and people "learned English without bilingualism as a significant intervening developmental stage" (1994:103). He maintains that the "type of European influence has played a strongly determinant role in lexical acculturation in Native American languages."

Although the small number of English loanwords in Hupa conforms to Brown's theory insofar as the Hupa were in contact with an English-speaking colonial power, his claim is essentially that low borrowing rates in indigenous languages will be correlated with a low incidence of bilingualism in colonial languages. However, the evidence presented in §1 suggests that by the early 20th century a sizeable proportion of the Hupa-speaking population was also bilingual in English. Brown's theory moreover has a corollary prediction that lexical borrowing should become more prevalent as bilingualism increases in a speech community. Yet Hupa shows no trend towards more lexical borrowing for items that were certainly introduced only after this bilingualism was achieved, modulo the possibility of an increased incidence of nonce borrowings discussed below. For example, the following 20th century inventions were all labeled with Hupa neologisms (cf. §2.3):

| form | gloss |
|---|---|
| me'-ch'ixine:wh | 'telephone' (lit. 'into it – one talks') |
| 'a:da:-nahɬ'its | 'automobile' (also 'train'; lit. 'by itself – it runs around') |
| nahx-ma:wilda:l | 'motorcycle' (lit. 'two – wheels') |
| ningxosting-me'iłchwe | 'refrigerator' (lit. 'ice – it makes') |
| 'a:da:-yixine:wh | 'radio' (lit. 'by itself – it talks') |

Table 5.3: Hupa neologisms for 20th century inventions

To this list might also be added items that were invented earlier but might not have been among the earliest technologies widely adopted in Hoopa Valley, such as *ta'na:n-xaywhiwh* 'pump' (lit. 'water – it sucks up'), *'a:da:-k'iɬ'e:n* 'typewriter' (lit. 'by itself – it writes'), and *'a:diɬ-na:k'iɬtul* 'bicycle' (lit. 'by oneself – one kicks around'). Thus, while there was clearly a high degree of Hupa-English bilingualism in the early 20th century, and still among adults in the mid-20th century, this seems to have had very little impact on the rate of lexical borrowing in Hupa, which remained minimal.

As an alternative to a theory invoking degree of bilingualism in a colonial language as an explanation for Hupa's conservatism with respect to borrowing vocabulary from English, it is tempting to instead consider this an extension of the longstanding tendency among Athabaskan speech communities to avoid lexical borrowing (Sapir 1921). This is unexpected, according to Sapir, given the extent to which Athabaskan groups modified their non-linguistic cultural practices under the influence of groups they came into contact with: "The cultural adaptability of the Athabaskan-speaking peoples is in the strangest contrast to inaccessibility to foreign influences of the languages themselves" (1921:214). This appears to be as true among the Pacific Coast Athabaskan languages as it is elsewhere in the family (although Conathan 2004 identifies some cases of pragmatic "functional convergence" between Hupa and its neighbors). As pre-Hupa Athabaskan groups arrived in northwestern California, they faced functional pressure to adapt linguistically to new circumstances as they migrated from north to south into regions with new, flora, fauna, and landscapes and as they interacted with groups whose languages, social organization, and technological adaptations to their environments were unfamiliar. These pre-Hupa speakers dealt with this situation in typical Athabaskan fashion, borrowing little and for the most part utilizing language-internal resources to label new things they encountered.

Such an explanation has two principal interpretations: one invoking structural barriers to borrowing (cf. Thomason and Kaufman 1988, ch. 2 for an assessment of such approaches), and a sociocultural one invoking metalinguistic attitudes that favor or disfavor borrowing (S. Rice 2012, Hill 2001). Brown (1994, 1999) considers genetic affiliation to be an indicator of structural similarity, an assumption that is especially justified in Athabaskan given the diachronic stability of many structural features in the family (K. Rice 2000, 2012). Certainly it is true that most Athabaskan languages were conservative in their approach to lexical acculturation in their contact with colonial powers: the Athabaskan languages included in Brown's study have an average borrowing rate of 25% and convergence index of -0.57. Brown notes, however, that there is a wide range of outcomes across the family, with borrowing rates as high as 72% for Dena'ina (which has a positive convergence index of .12) and approaching 50% in some other Alaskan Athabaskan languages. Thus, despite the overall trend in the family towards labeling introduced items with native-sourced vocabulary, Brown concludes that structural factors per se were probably not a crucial factor in borrowing rates, or at least were not insurmountable

barriers to borrowing since structurally similar members of the same family have such variable outcomes.

Brown argues that a sociocultural interpretation of Athabaskan linguistic conservatism therefore seems more likely, and in the case of Hupa vis-à-vis English this is especially true if considered in areal perspective. In pre-contact times, Hupa and its non-Athabaskan neighbors Yurok and Karuk shared many cultural affinities as focal members of the Northwestern California culture area identified by Kroeber (1920). Linguistically, this included a tendency to avoid lexical borrowing despite a fairly high degree of multilingualism in the region (Conathan 2004:80-81). All three groups came into contact with the same English-speaking colonial power at the same moment in history. All three remained more or less in situ in their aboriginal territory following contact (relative to the mass relocations that occurred elsewhere in the region, e.g. the movement of Oregon Athabaskans to Siletz) and were subject to many of the same assimilationist pressures in subsequent decades, which led to similar long-term outcomes: all three languages are today spoken as a first language primarily by a handful of older adults.

Hupa's borrowing rate (20%) and convergence index (-0.67) are slightly lower than those reported by Brown for Karuk (1994:99), which has a borrowing rate of 24% and a convergence index of -0.58. Analysis of vocabulary obtained from the online Yurok Language Project Digital Archive (2001-2013) and Andrew Garrett (p.c.) shows that Yurok has a borrowing rate of 34% and a convergence index of -0.39. The convergence index for all three languages is negative and the borrowing rate well below the average Brown cites for California as a whole, showing that there was a preference for endogenous-sourced vocabulary over loanwords in all three cases. Although the difference between Hupa and Yurok seems large at first glance, a statistical comparison of their borrowing rates reveals that the difference is not significant ($p > .1$).[5] These data invite an interpretation of Hupa linguistic conservatism with respect to English not in terms of a low incidence of bilingualism but instead as a continuation of a pre-existing areal tendency to eschew borrowing.

Hill's (2001) theory of localist versus distributed stances to adopting linguistic innovations, outlined in §3 of chapter 1, is appealing here. Linguistically conservative localist stances disfavoring the adoption of exogenous material would militate against borrowing. For Hill, these stances are cultural phenomena and presumably subject both to transgenerational transmission within a given society (hence the geographic portability of linguistic conservatism among migrating Athabaskan groups) and to diffusion and reinforcement within a particular region. Golla (2011:4-5) considers conservatism with respect to borrowing to be an important feature of California as a broader linguistic area, one he relates to a symbolic relationship between language and land, people's deeply-felt connection to their home territory reflected in a strong tendency towards conservative linguistic stances. While Golla's theory is more abstract than Hill's materialist one relating stance-taking to claims on resources, the two are not incompatible: Golla's symbolic connection between language and land could be reinterpreted as a linguistic instantiation of groups' primary claims on resources within a given locality (cf. Golla

[5] A chi-square test using the prop.test() function in R (R Core Team 2013) yielded a p-value of 0.1441, above a desired significance level of .05. Only 44 of Brown's 77 glosses were located for Yurok, versus 70 for Hupa. 41 glosses have data for both languages. If the analysis is restricted to these items only, the difference between the two languages is greater: Hupa has a borrowing rate of 17% and a convergence index of -0.76, whereas Yurok has a borrowing rate of 37% and a convergence index of -0.34. Even so, this fails to reach the threshold of significance ($p = 0.08104$).

2000:50).

From this theoretical perspective, the maintenance of pre-existing Athabaskan and areal tendencies towards conservatism imply that borrowing English words into Hupa conferred no advantage on asserting claims to resources in the post-contact world.[6] It is certainly the case that the colonial encounter was extremely disruptive for Hupa people and involved major shifts in social organization, land use, and subsistence patterns, inter alia (although some accounts, including the tribal history by Nelson 1978, emphasize the high degree of continuity between the pre-contact and post-contact worlds). Notice moreover that this linguistic conservatism did not prevent the development of Hupa-English bilingualism, which as discussed in §1 seems to have been quite robust by the turn of the 20th century; nor did it prevent the gradual replacement of Hupa with English as the main language acquired by children in Hoopa Valley. Language shift is in fact one of the phenomena that Hill (2001:277) suggests her theory might be well-suited to explaining, as groups adopt a distributed stance with respect to a language that is linked to "new cultural construals of environmental rights and values." The suggestion here is merely that particular kinds of linguistic behaviors – adopting English-sourced vocabulary in Hupa-framed discursive practice – did not become indexically linked to asserting claims on resources.

This view is not without problems, however. As noted, linguistic conservatism in Hupa, Yurok, and Karuk relates to pre-contact trends in California as a linguistic area, where lexical borrowing was not common across the region (Golla 2011). Yet as Bright (1960) originally pointed out, and data in Brown (1994, 1999) confirm, there is a much higher incidence of borrowing from colonial languages in the southern California region than in the northern California region. If the resistance to lexical borrowing in the north was simply a continuation of pre-contact linguistic conservatism, why was there not a similar continuation in the south? As noted above, Brown (1994, 1999) follows Bright (1960) in attributing this to differences in the incidence of bilingualism the two regions. He moreover argues that the differences in bilingualism are in turn due to different strategies of colonization adopted by English-speaking versus Spanish-speaking powers. These strategies may well be relevant to understanding lower rates of lexical borrowing in the north, but not because they led to a lower incidence of bilingualism per se – certainly this does not seem to be the case for Hupa, where bilingualism in English was achieved rapidly in the reservation setting. Rather, under Hill's (2001) theory the difference between outcomes in the north and south could instead be due to lexical borrowing (rather than wholesale language shift, at least initially) becoming a valid expression of new, distributed stances as assertions of secondary claims on resources in one colonial context but not the other. Developing this point in more detail is left for future work, but it offers a plausible explanation for different outcomes in the two cases.

### 2.1.2. Phonological Impact of Loanwords

In many cases of language contact, lexical borrowing leads to changes in the recipient language's phonological system, and can sometimes have an even more profound impact in its grammar (Thomason and Kaufman 1988:74-76). However, because Hupa borrowed so few English words, the overall impact of loanwords on Hupa phonology is quite minimal. Moreover, for the most part the phonological adaptation of English loanwords in Hupa is unsurpris-

[6] This view may be problematic if under-documented phenomena such as nonce borrowing and code switching, discussed below, are considered.

ing. English voicing and minor place of articulation contrasts especially appear to have been susceptible to substitution with Hupa phonemes. For example, Hupa has a phoneme /s/ but not /z/, and accordingly English /z/ is realized as Hupa /s/ in the loanwords in Tables 5.1 and 5.2 (e.g., in the English plural suffix in 'apples', 'beans', etc.). Similarly, Hupa /ʍ/ replaces English /f/ in 'coffee', and /e:/ (Hupa's only non-high front vowel) replaces English /æ/ in 'apples', 'candle', and 'matches'. Also unsurprising is the insertion of an epenthetic vowel to break up complex codas with final /s/, phonotactically ill-formed in Hupa, in 'apples', 'beans', and 'plums'. (Epenthesis fails to apply in 'turnips', a form that is also aberrant in having initial /k$^h$/ for English /t/ and final /š/ rather than /s/.)

One general effect of English loanwords on Hupa phonology is the expansion of marginal or specialized Hupa phonological categories preserving English manner and major place of articulation. This is seen first of all in the labial stops found in 'apples', 'cat', 'wheelbarrow', 'beans', 'peaches', and 'plums'. Before these loanwords entered the language, Hupa's labial consonants were /m/, /w/, and /ʍ/; /b/ existed, but only in the word *'ijibeh* (also *'idzibeh*), an exclamation expressing fright, possibly a diminutive sound-symbolic variant of *'ijiweh* (Golla 1970:27). The addition of English loanwords to the Hupa lexicon expanded the category /b/ to a small number of additional lexical items. The fact that English labial stops were preserved as such rather than substituted with more frequent Hupa phonemes (say, /m/ or /w/) may have been facilitated by the fact that many Hupa people in pre-contact times were at least passively competent in Yurok, Karuk, Chimariko, and other indigenous languages of the region that do have labial stops.

The expansion of specialized phonemes is also seen in English /k/ (phonetically [k$^h$] word-initially) mapping to Hupa /k$^h$/ (orthographic <K> in Table 5.1) in 'coffee', 'corn', 'cloth', 'candle', and 'quince', and word-initially in 'cookies'. There are two primary velar places of articulation in Hupa: front-velar, described by Golla (1970:27-28) as "dorso-palatal," and back-velar, described as "considerably further back than in any English form" (i.e., approaching uvular articulation); as discussed in chapter 3, the back-velar aspirated stop (P-A *q) merges with the fricative /x/ in Hupa. A third dorsal place of articulation is intermediate between the two, described by Golla as "dorso-palatal/velar, much as in English ken or can" (1970:27). These velars are found in endogenous Hupa words, but only as diminutive sound-symbolic variants of the corresponding front velars: *whikya:y* 'my grandchild' (with IPA /k$^{jh}$/) but *whiKa:y* 'my dear little grandchild' (with IPA /k$^h$/) (Golla 1970:28). Loanwords with English /k/ realized as Hupa /k$^h$/ thus led to the expansion of a phonemic category that had previously been restricted to a particular sound-symbolic function. This is similarly the case for /š/ (in 'shawl' from Curtin 1888-1889 and 'turnips' in Table 5.2), which exists in Hupa only as a variant of /s/: phonemically as a diminutive sound-symbolic substitution and allophonically before palatal affricates. In this loanword, and perhaps in some of the cases treated under nonce borrowings below (e.g., [ši:p] for 'sheep'), contact with English led to a slight expansion of a phonological category that had previously been restricted to particular morphological and phonological contexts.

Another interesting point is the variable treatment of English voiceless stops (including affricates), allophonically aspirated as onsets in word-initial position and in stressed syllables. In some cases English /p/ is realized in Hupa as unaspirated /b/ (i.e., IPA /p/). This is found word-initially in 'cat' and in Verdena Parker's imitation of an archaic pronunciation of 'peaches', and as a word-internal syllable onset before the epenthetic vowel [ɪ] in 'apples'. Mrs. Parker produced an aspirated [p$^h$] in 'plums', in two tokens of 'peaches' in texts, and in some of the

words treated below under the rubric of nonce borrowings (e.g. [pʰi:s] for 'peas'). However, the existence of invariant /b/ in 'cat' and the stereotyped pronunciation of 'peaches' suggest an early stratum of borrowing when bilabial stops were imported into Hupa uniformly without whatever allophonic aspiration might have been present in source language tokens of these words. Similarly, English /tš/ is uniformly realized in Hupa as unaspirated /dž/, even in 'chicken' where it is phonetically aspirated in word-initial position.[7] By contrast, English /k/ is realized as Hupa /kʰ/ where it is allophonically aspirated in English: word-initially in 'corn', 'coffee', 'cloth', 'cookies', and 'quince' and in the final syllable with secondary stress in 'cloth' (< 'calico'). It is realized as unaspirated /g/ (or /gʲ/) in the unstressed syllables in 'chicken' and 'cookies'.[8]

The key point here is that aspiration in the English phonetic signal is not treated uniformly in Hupa loanwords, even though the language has a phonemic aspirated ~ unaspirated laryngeal contrast. These data bear on a current debate in the literature on lexical borrowing concerning the extent to which loanword adaptation is primarily driven by perceptual phonetic cues versus more abstract phonological features (cf. de Jong and Cho 2012 and Chang 2012 for discussion and references). In the former view, loanword adaptation is due to L1 speakers attempting to fit L2 phonetic signals to their native phonological systems; subphonemic properties of the acoustic signal can influence the categorization of sounds and "similarity is determined by the acoustic substance of the lexifier language" (de Jong and Cho 2012:342). In the latter view, developed most extensively by LaCharité and Paradis (2005), loanword adaptation is a process of mapping L2 phonological categories to L1 phonological categories via contrastive features, ignoring subphonemic properties of the acoustic signal.[9]

The Hupa data are particularly germane to LaCharité and Paradis' theory, since one of the main empirical phenomena it accounts for is that loanword adaptation of laryngeal contrasts sometimes abstracts away from their phonetic implementation. Thus, while voiced and voiceless stops – i.e., the categories distinguished by the phonological feature [voice] – have very different VOT implementations in English, French, and Spanish, voiceless maps to voiceless and voiced to voiced in loanword adaptation involving these languages. However, this theory has difficulty accounting for the treatment of English /k/ in Hupa. If this category is underlyingly [-voice] in English regardless of surface allophonic aspiration, then there is an expectation that it should be realized uniformly as its [-voice] counterpart in Hupa as well regardless of its position in a given word or whether the syllable it occurs in is stressed. Thus, if Hupa /kʰ/ is phonologically [-voice], as has been suggested by some researchers for aspiration in other Athabaskan languages (e.g., Howren 1971), English 'cookies' should be adapted as /kʰokʰɪs/

---

[7] Golla (p.c.) points out that some of these early loanwords match similar forms documented by Bill Bright in the neighboring language Karuk (e.g. *pičas* 'peaches' in Bright 1950), whose aboriginal territory was one of the earliest loci of contact during the California gold rush. Karuk lacks contrastive aspiration, so some of the early English loanwords lacking aspiration in Hupa might have been borrowed indirectly via Karuk.

[8] English /k/ in the unstressed second syllable of 'chicken' is realized in Hupa as front-velar /gʲ/ before the high front vowel [ɪ]. In the same context in recorded tokens of 'cookies', it is impressionistically at the intermediate velar position. Little depends on this for present purposes; the important point is that these are realized as unaspirated rather than aspirated in this position.

[9] It should be noted that some of the adaptations in Table 5.2 are not accounted for by either theory, e.g., Hupa /kʰ/ for English /t/ in 'turnips' and /h/ for English /t/ in 'tea'. Phonetic accounts probably fare somewhat better in this regard in offering a direct connection between mis-perception and adaptation.

rather than the attested /k$^{h}$ogɪs/. Even if Hupa aspirated stops are not [-voice] (cf. Rice 1994), it is expected that the mapping of English /k/ should be uniformly realized as either /k$^{h}$/ or /g/ in Hupa simply because Hupa lacks an allophonic aspiration rule.

The fact that English /k/ is sometimes adapted as /k$^{h}$/ in Hupa, other times as /g/, suggests that subphonemic properties of the L2 acoustic signal were relevant to determining the L1 phonological representation of the adapted loanwords. This is consistent with phonetic accounts where adaptation is driven by perceptual cues rather than by the mapping of abstract phonological categories. If aspiration is present in the input signal it is preserved; if it is absent from the input signal, it is not inserted. This also is expected given the fact that voice onset time (VOT) for English /k/ in word-initial position (average 80 ms for words produced in isolation according to Lisker and Abramson 1967) is comparable to Hupa /k$^{h}$/ (average 84 ms for words produced in isolation – Cho and Ladefoged 1999). Unaspirated English /k/ would not be misperceived as Hupa /k$^{h}$/, and a phonetic account correctly predicts that Hupa would realize English /k/ as unaspirated /g/ in unstressed syllables.

A phonetic-perceptual account of loanword adaptation is also well-suited to explaining why English /tš/ is realized as Hupa /dž/ even word-initially in 'chicken'. The aspirated counterpart of unaspirated /dž/ in Hupa is labialized /tš$^{wh}$/. According to the phonetic-perceptual view, the lack of labialization in English [tš$^{h}$] makes it unlikely that it would be perceived as /tš$^{wh}$/ in Hupa. Moreover, while Hupa /dž/ is phonologically unaspirated, according to Gordon (1996) the average VOT for word-initial unaspirated affricates produced in isolation is on the order of 70 ms in Hupa. Gordon provides data only for unaspirated dental /dz/, but his discussion suggests that the same would be found for /dž/ as well.[10] Gordon also reports that the mean VOT for Hupa /tš$^{w}$/ is 167 ms. Lisker and Abramson (1967) do not provide VOT data for English affricates, but it is plausible that the mapping of allophonically aspirated English [tš$^{h}$] to Hupa /dž/ rather than /tš$^{wh}$/ would also have been supported by perceptual cues related to VOT.

However, the phonetic-perceptual view has some difficulty explaining why English /p/ often maps to Hupa unaspirated /b/ even where it is allophonically [p$^{h}$], even if only in the earlier stratum of loanwords like 'cat' and 'peaches'.[11] If allophonic aspiration is preserved in the phonological adaptation of English /k/, why is it not also preserved in the adaptation of English /p/? A key difference between the two cases is that there was no phoneme /p$^{h}$/ already existing in the Hupa phonological system.[12] Although marginal, /b/ did exist, so a process of fitting the phonetic substance of English allophonic [p$^{h}$] to a Hupa phonological category might have favored assigning a segment to an existing category over creating an entirely new one. More importantly, there is circumstantial evidence that aspiration on English [p$^{h}$] might have fallen below a critical threshold for phonemic aspiration in Hupa. According to Cho and Ladefoged (1999), there is a cross-linguistic tendency for VOT to increase the farther back in the mouth the stop closure is made. Gordon (1996), however, shows that Hupa is exceptional in this regard: average VOT for aspirated /t$^{h}$/ in Hupa is 82 ms for words produced in isolation,

[10] This is consistent with impressionistic analysis of VOT for /dž/ found in recordings made with Verdena Parker.

[11] As noted above, this earliest stratum of loanwords may have been borrowed indirectly via Karuk, which lacks contrastive aspiration. If so, these items are not problematic for a perceptual account.

[12] This is handled straightforwardly by the phonological account of LaCharité and Paradis (2005), where phonological mappings are determined by a least-distance principle. Since Hupa had no pre-existing aspirated labial consonant, its /b/ (i.e., IPA /p/) would have been the nearest phonological category to English /p/.

comparable to VOT for aspirated /k$^h$/. Crucially, the average VOT for English [p$^h$] is much shorter – 58 ms for words in isolation, and 33 ms for words produced in sentences (Lisker and Abramson 1967). The VOT duration on English [p$^h$] might simply have been too short for Hupa speakers to identify it as belonging to the class of aspirated phonemes.[13]

### 2.1.3. Nonce Borrowings and Code Switching

As alluded to above, the prevalence of loanwords in Hupa might be under-documented due to the context of elicitation encouraging purist linguistic stances among bilingual speakers of a given language (Hill 1994). This may also be due to the fact that as bilingualism in a given community increases, the range of exogenous-sourced material available to speakers comes to encompass virtually their entire L2 lexicon, what Matras (2009) has called bilingual speakers' "repertoire." Oswalt (1994:111) describes the situation for Kashaya (a Pomoan language of northern California) thus:

> "The possibilities include their entire English vocabulary, as they are so completely bilingual that they can use any English word in a Kashaya context and be understood. However, such uses are perceived by most linguists and native speakers not as Kashaya words but as English words appearing within a Kashaya context, and I suspect it is this perception that leads to many of the claims that there are few loanwords from English in the native American languages."

Oswalt also points out that English words that entered Kashaya when there was a lower incidence of bilingualism across the speech community underwent a high degree of phonological adaptation and are more likely to be identified as loanwords; words that are used by fluent bilinguals undergo little phonological adaptation and may not be considered loanwords per se.[14] These considerations suggest that there might have been more English-sourced material in use among Hupa-English bilinguals when speaking Hupa than is evident from the available documentation.

This is a reflection of the larger problem in the study of language contact of distinguishing between single-word code switches (nonce borrowings) and established loanwords. By the early 20$^{th}$ century, there were probably many words straddling this line in Hupa, with English-sourced lexical items in use alongside Hupa-sourced ones. Some examples of loanwords in use alongside Hupa neologisms (i.e., where the English word underwent phonological adaptation) are given in §2.4 below, but might also include cases like *k'iɬixun-nehwa:n* 'sheep' (lit., 'deer – it resembles'), documented in Curtin (1888-1889) and Golla (1984), alongside Verdena Parker's memory of English [ši:p] being in common use. Whether these were in fact established borrowings or what might be considered serial nonce borrowings is difficult to say. Such considerations are part of the motivation for Matras (2009) dispensing with the idea of separate bilingual codes, proposing instead the notion of a broader linguistic repertoire.

---

[13] A phonetic account would then need to maintain that aspirated [p$^h$] in 'plums' and in nonce borrowings was the result of more fully bilingual speakers being able to identify and reproduce short-duration aspiration. It is noteworthy that two tokens of [p$^h$] in 'peaches' produced in texts by Verdena Parker have VOT values (44 ms and 48 ms) on the order of [p$^h$] in English reported by Lisker and Abramson (1967).

[14] LaCharité and Paradis (2005) also present evidence that the greater the incidence of bilingualism there is across a community, the less phonological adaptation loanwords undergo.

It has also been noted already that narrative genres in published collections of Hupa texts like Goddard (1904) and Sapir and Golla (2001) are not ones that would be likely to invoke English-sourced vocabulary, since many are set in a pre-contact or mythological past or describe practices strongly affiliated with traditional Hupa culture. However, there are a few glimpses of what Hupa as spoken by competent bilinguals might have been like in texts dealing with contemporary issues published in Golla (1984) and produced by contemporary speaker Verdena Parker. The following examples are from speakers Ned and Louisa Jackson (NJ and LJ respectively) in Golla (1984):[15]

(1) de:-q'ung-it na:'undiyay 'aywe:-ding-mił school-ding-mił 'a'de:ne' "nawhtsung!"
after she had come home some-place-from school-place-from she said "I've found it!"
When she came home from somewhere, home from school, I heard her shouting, "I've found it!"
(LJ, Golla 1984:5)

(2) ya:dehsdilts'e:-ne'en, hay-de: hospital-ding, nohxontaw' sa'a:n-ne'en hayah
we were living-PST, this hospital-place, our house lies-PST there
... where we used to live ... at the (old) hospital – our house used to be there. (NJ, Golla 1984:11)

(3) woodshed-me'-ch'ing' noh'ing!
woodshed-inside you all look!
Go look in the woodshed! (NJ, Golla 1984:13)

(4) haya:ł k'iłixun se:sehłwe:n xutł'e'-dung' for breakfast!
then deer I had killed in the morning for breakfast
I had killed a deer in the morning for breakfast! (NJ, Golla 1984:13)

The last of these examples is clearly a case of code switching since it involves a multi-word constituent. The others involve single words with the Hupa locative enclitic *-ding* and/or postpositions *-mił* 'with/from it', *-me'-ch'ing'* 'into it' attached, and it is unclear whether they are loanwords or nonce borrowings (and hence perhaps code switches as well). In some of these cases there were well-established Hupa neologisms with the same meaning that might have been used instead, e.g., the use of English *school* in example (1) is in lieu of *'a:k'iwilaw-me'-ch'inehł'e:n* (lit. 'writing – in it – someone looks at') from the Hupa Language Dictionary (Golla 1996b), a variant of which is found as early as Curtin's (1888-1889) vocabulary.

Similar examples can be found in texts narrated by contemporary Hupa speaker Verdena Parker. For example, in one story about an illness that she experienced in her youth, she first refers to the hospital where she received treatment as *hospital ch'idinch'a:t me'-łe:na:wh-ding*, where the English word *hospital* is followed by a Hupa phrase translated literally as 'sick people – in it – they come together – place'; no pause separates the two. She subsequently refers to the hospital as *me:ya'dinch'a:t* (lit., 'people are sick at it'). Mrs. Parker also sometimes uses English-sourced vocabulary together with *'o:lye:* 'it is called'. This is found especially with placenames outside of traditional Hupa territory, as in (5), but also sometimes with common nouns, as in example (6):

[15] The orthography has been modified slightly from the source to conform to the community orthography commonly in use today.

(5) yo:w yiduq Washington-'o:lye:-ding ya:ydil-e:
there uphill/north Washington – it is called – place we went
'We went way up there to Washington.' (VP-051210-06, 3:30)

(6) hay radio 'o:lye: na:tehsintse:-de' ...
the radio it is called you turn it on – if
'If you turn the radio on ...' (VP-051210-07, 3:35)

In such cases, *'o:lye:* flags the preceding word as being non-Hupa in some sense; compare this use of English *radio* with *'a:da:-yixine:wh* in Table 5.3 above. Collectively, examples (1)-(6) show that the use of English-sourced vocabulary, whether as established loanwords, nonce borrowings, or code switches, was probably more common among Hupa-English bilinguals in everyday contexts discussing contemporary issues than is represented in early published collections of Hupa texts.

## 2.2. Semantic Adaptation and New Usages

The second strategy employed by Hupa speakers to label the plethora of new items introduced by Euro-American society was to use an existing Hupa word. Table 5.4 provides examples with glosses for both the old and new items:

| form | gloss (traditional) | gloss (modern) |
|---|---|---|
| q'osta:n | '(women's) basket-hat' | 'hat' |
| yehch'itul | 'moccasins' (lit. 'what someone steps into') | 'shoes' |
| kya' | '(traditional) dress' | 'dress' |
| ts'iłting' | 'bow' | 'rifle' |
| dinday | 'obsidian arrowhead' | 'bullet' |
| nahdiyaw | 'dentalia shell money' | 'money' |
| de:diwiłiq' | 'baked acorn mush' (lit. 'doughy object that has been put into the fire') | 'bread' |
| mił-k'idiłwis | 'fire-drill' (lit. 'with it – one drills for fire') | 'auger'; 'matches' |
| xontah | '(traditional) house' | 'house' |
| no:na:witse | 'sliding door of traditional house' lit. 'what is shoved back' | 'door' |
| minde'iłchwe | 'wild native tobacco (Nicotiana bigelowii)' | 'tobacco' |
| mide'xole:n | 'buck deer or elk' (lit. 'its horns – plenty') | 'cow' |

Table 5.4: Hupa words adapted to items introduced in the 19th century

In each of these cases, the introduced item was considered similar enough to something in the pre-contact Hupa world that an existing word could be used to label it. Often the similarity was functional, as in the use of *mił-k'idiłwis* to mean 'matches', but sometimes other salient dimensions were relevant too, as in the use of the same word to mean 'auger' (the drilling motion independent of the function of making fire being salient). In many cases the equivalence was no doubt facilitated by the semantic transparency of morphologically complex Hupa nominals, but in others no such decomposition was available (e.g., 'hat', 'bow ~ rifle', 'arrow-head ~ bullet'). These equivalences were also often made despite obvious differences between the old and new items in form (e.g., bows vs. rifles) and materials used in construction (metal

vs. stone, wood, shell). The differences were sometimes subtler. For example, traditional caps made of basketry were typically worn only by women (Goddard 1903:18-20), but *q'osta:n* came to designate European-style hats worn by men as well. Similarly, the traditional Hupa *xontah* was constructed largely underground, but the word nonetheless was used for frame houses built aboveground after contact. The introduction of new items, then, led to changes in typical usage if not the actual denotation of these words. In some cases a referential broadening took place, followed by semantic shift: *mide'xole:n* was used for female cows even though originally the word had designated male deer and elk only, a meaning listed as 'archaic' in the Hupa Language Dictionary. Golla (p.c.) notes that some toponyms denoting places outside the Hupas' aboriginal territory underwent semantic shift in coming to denote post-contact settlements that were established nearby, e.g., *dahwilahɬ-ding* (lit. 'the place where it floats on top') for the town 'Eureka', which in pre-contact times referred to a Wiyot village on Indian Island in Humboldt Bay (Sapir and Golla 2001:895).

Where semantic shift didn't occur, both the old and new items could be identified by the same word. To distinguish them – highlighting the bifurcation of the world into objects typically associated with traditional cultural practices and those introduced by Euro-American society – words designating introduced items were sometimes combined with *yima:n'dil* 'white person', itself a neologism meaning literally 'from across – they go around'. The earliest example of this usage is found in Curtin (1888-1889), where 'cloth' is rendered as *yima:n'dil t'e'* 'white person's blanket' (*t'e'* was a traditional blanket made of buckskin). Similarly, traditional Hupa gambling sticks were originally known simply as *kin* 'stick'; playing cards came to be called *yima:n'dil mikine'* 'white person's sticks', with the possessed form of the root *kin*. More commonly, however, the original Hupa item would be prepounded with *xo'ji*. This word is defined in the Hupa Language Dictionary (Golla 1996b) as 'true, real (thing); well, really, thoroughly; in a moderate, correct way'. Typical uses in texts are shown in examples (7) - (10):

(7) hai-ya-miL-ûñ **xōtc** min-nōl-weL-miL
And **quite** it was midnight
'Well after midnight ... ' (Goddard 1904:293.1-2)

(8) hayahujit **xo'ji** nikyah-xo ɬe:na'uliwh
then **really** in a big way they build a fire
'Then they build a really big fire.' (Sapir and Golla 2001:89, 13.57)

(9) hai **xōtc** min-nē-djit na-naL-tsis
the **very** middle it hung
'Right in the middle was hanging (the blanket).' (Goddard 1904:207.8-9)

(10) q'u(t) **xo'ji** mije'e:din sile'n
now **a real** baby it has become
'Now it has truly become a baby.' (Sapir and Golla 2001:162, 13.40)

Uses such as (7) and (8), where *xo'ji* is an intensifier modifying a verbal expression, are by far the most common. Uses such as (9) and (10), where *xo'ji* arguably modifies a noun (as suggested by the gloss lines), are fairly rare. Nonetheless, the word was recruited to discriminate traditional items designated by nouns, as shown in the following examples from the Hupa Language Dictionary:

| form | gloss |
|---|---|
| xo'ji-q'osta:n | 'traditional hat, basketry cap' |
| xo'ji-yehch'itul | 'moccasins' |
| xo'ji-kya' | 'ceremonial abalone-shell dress' |
| xo'ji-ts'iɬting' | 'traditional (sinew-backed) bow' |
| xo'ji-minde'iɬchwe | 'wild native tobacco (Nicotiana bigelowii)' |
| xo'ji-dinday | 'obsidian arrowhead' |
| xo'ji-king | 'traditional gambling sticks' |
| xo'ji-xontah | 'Indian house, living house' |
| xo'ji-nahdiyaw | 'dentalia, Indian money' |

Table 5.5: Uses of *xo'ji* 'true, real' to designate traditional Hupa items

This appears to be a completely novel use of *xo'ji*: there are no documented cases of *xo'ji* used to distinguish, say, something prototypically Yurok or Karuk from something prototypically Hupa.

The new use of *xo'ji* occurs only once in Goddard's collection of Hupa texts, in a story told by the wife of Dan Miskût about the arrival of the first white people in Hoopa Valley: <xōtc tciL-Liñ> (*xo'ji ɬing'* in the contemporary community orthography), glossed as 'real dogs' and rendered as 'native dogs' in the free translation (Goddard 1904:200.6). Nonetheless, Goddard comments in a footnote that "the word xōtc is used to indicate that the Indian thing is meant before names which have been transferred to things introduced by white men." That Goddard was able to generalize in this way suggests that the usage was already well-established by the end of the 19th century; the examples 'hat', 'moccasins', 'bow', 'dress', and 'tobacco' in Table 5.5 are found as early as Curtin (1888-1889). It is also found in a Chilula story in Goddard (1914b) told by Tom and Dan Hill about a war with the Lassik Indians. There is no indication in the story itself that this war took place in post-contact times, but for one clue: some of the fighting involves the use of guns rather than bows and arrows. The interplay between *xo'ji* and *yima:n'dil* is illustrated in (11):

(11)

| | | | | |
|---|---|---|---|---|
| **xōtc siL tiñ** | miL | kit tce xa in ye | sa a | kit tce xa in ye |
| **Bows** | with | they were fighting. | Long time | they fought. |

| | | | | |
|---|---|---|---|---|
| hai ya hit djit | **yī man dil** | **mit siL tin ne** | miL | kyū wim mût |
| Then | **white man** | **his gun** | with | they shot. |

'[They] began shooting at them with bows and arrows. After they had been fighting for some time they began to shoot with white man's guns.' (Goddard 1914b:294.8-10)

This new use of *xo'ji* was not fully lexicalized, since unmodified nouns could still be interpreted to refer to the traditional Hupa item. Thus, while 'bow' is translated as *xo'ji-ts'iɬting'* in the wordlist in Curtin (1888-1889:88), in a myth text transcribed in his fieldnotes *ts'iɬting'* occurs without *xo'ji* meaning 'bow'. Similarly, although *xo'ji-xontah* 'real, Indian house' and *xo'ji-k'iwidits* 'real, Indian rope' both occur in Sapir and Golla (2001), elsewhere the same objects are referred to without *xo'ji*, even within the same text.[16] Still, some of the *xo'ji* formations might have achieved a quasi-lexicalized status for some speakers, as indicated by their

[16] These examples occur in text 23, lines 1 and 30, and text 26, line 83.

inclusion in the Hupa Language Dictionary and possessed nominals elicited from Verdena Parker: *whixo'ji-q'osta'n* 'my Indian hat' or *whixo'ji-xontaw'* 'my Indian house', with the possessive prefix *whi-* attached to *xo'ji* instead of the head noun.

## 2.3. Neologisms

Probably the most common strategy Hupa speakers used to label introduced items was the creation of new words using derivational resources, as productive in Hupa as elsewhere in the Athabaskan language family. This was not a new phenomenon in the language. The number of monomorphemic nouns in Hupa is fairly small: Golla (1964) lists 243 noun roots total, which includes a few bound elements occurring only in old compounds that are arguably unanalyzable synchronically (Goddard 1905:16). The vast majority of Hupa nouns are morphologically complex, a situation described thus by Goddard (1905:17): "There is a large and increasing number of nouns, formed by suffixes or composition, the descriptive force of which is ever present in the mind of the speaker."[17] Some of Goddard's examples (transliterated to the contemporary community orthography) include the following:

| form | gloss | literal translation |
|---|---|---|
| misah-niłchwin | 'buzzard' | 'its mouth – stinks' |
| xine:wh | 'language' | 'it speaks' |
| ch'iłtul | 'a dance' | 'he stamps with his foot' |

Table 5.6: Hupa deverbal nouns from Goddard (1905)

Goddard attributed the pervasiveness of this type of noun to two factors: taboos proscribing the use of the names of deceased people, which in some cases were common nouns, and the introduction of new items due to contact with Euro-American society. Elsewhere, Goddard (1911:152) also mentions "the pleasure which the Hupa find in poetical descriptive names," an aesthetic consideration that might have favored the coining of new vocabulary.[18]

Examples of lexicalized nominal expressions denoting both traditional and introduced items

[17] Some of these formations are clearly quite old, as shown by comparison with other California Athabaskan languages. For example, Hupa *dilxich* 'fawn' corresponds to Kato *del kûcts*, literally 'spotted' (Golla 1996b:33, Goddard 1912:30), with the regular correspondence Hupa /x/ ~ Kato /k/. However, as noted in Chapter 2, Hupa seems to have innovated lexically more than other California Athabaskan languages, even in the basic vocabulary from the Swadesh-100 list. Goddard (1912:19) points out several cases where "[t]he Hupa have descriptive names in the place of these Kato nouns, the apparently original ones. In several instances the change appears to be recent," giving the example of Hupa *ta nan* (ta'na:n) 'water', lit. 'what one drinks' versus *tō* 'water', still found in Hupa in compounds like *to:-nehwa:n* 'obsidian', lit. 'water-it resembles'.

[18] Taboo name avoidance may not have had the lasting impact that Goddard implies. One of the key examples Goddard (1901) offers to illustrate this point is the use of *mił-ky'o:xe:t* (lit., 'with it – one buys') in lieu of *nahdiyaw* for 'money', but *nahdiyaw* is found throughout Goddard's own Hupa Texts (1904). Another is *łehq'onch-yiditile* (lit., 'salt – it relishes') instead of *xa'* for 'goose'. Yet while Goddard maintains (1909:73, fn. 32) that Hupa *xa'* 'goose' "disappeared a generation ago," it is found both in Curtin (1888-1889:116) and in the Hupa and Whilkut vocabularies collected by C. Hart Merriam between 1910 and 1930.

are given in Table 5.7:[19]

| form | gloss | literal translation |
|---|---|---|
| me'-chwich-ch'iɬchwe | 'burden basket' | 'in it – firewood – one gathers' |
| me'-no:na'mil | 'clothes basket' | 'in it – one throws back down' |
| k'e:da'ay-miɬ-ch'iɬwe | 'fish club' | 'fish head – with it – one beats it' |
| miɬ-'a'k'iɬ'e:n | 'pen, pencil' | 'with it – one writes' |
| k'ima:w-ch'iɬchwe | 'doctor (Indian or modern)' | 'medicine – one makes' |
| tse:ɬch'e'-ch'itsit | 'blacksmith' | 'iron – one pounds' |

Table 5.7: Lexicalized relativized nominals in Hupa

Such relativized nominals are fully inflected Hupa verbs (typically imperfective) whose arguments can be retained explicitly. The noun 'burden basket', for example, is composed of a verb 'gather' inflected with the 3rd person animate/proximate subject prefix *ch'i-* plus its object *chwich* 'firewood'. Instrumental relativizations using *miɬ* 'with it', as in 'fish club' and 'pen, pencil' are an especially common way to label tools, both traditional and introduced.

For the most part relativized nominals labeling introduced items are similar to those labeling traditional items. An exception is things that operated under their own power, which led to the coining of a several nouns that included *'a:da:*, literally 'by itself': *'a:da:-nayk'iɬxut'* 'sewing machine' (lit., 'by itself – it sews'), *'a:da:-nahɬ'its*, originally 'train' and later 'car' (lit., 'by itself – it runs around'), *'a:da:-k'iɬ'e:n* 'typewriter, computer' (lit., 'by itself – it writes'). Such things were previously unknown in Hupa and can be considered equivalent to the use of *auto-* or *automatic* in English. However, although this use of the reflexive with the postposition *-a:* in such nominalizations is innovative, structurally these examples are equivalent to other subject-relativized nominals with a postpositional object.

In §2.2, examples were given of existing Hupa words that came to designate introduced items. Sometimes, however, the similarity between the traditional and introduced items was deemed only partial. This led to the expansion of a strategy for creating neologisms by postpounding *-nehwa:n* 'it resembles' to an existing word. This is found in pre-contact Hupa words designating colors especially, e.g. *tse:l-nehwa:n* 'red' (lit., 'blood – it resembles'), *xon'-nehwa:n* 'orange' (lit., 'fire – it resembles'), and in nouns like *to:-nehwa:n* 'black obsidian' (lit., 'water – it resembles') and *k'iɬts'o:s-nehwa:n* 'ceremonial headgear' (lit., 'Indian paintbrush – it resembles'). This strategy was used to label a number of introduced items, especially foods:[20]

[19] One of the most common means of deriving nouns in Hupa is via the so-called "relative enclitic." This is historically a nominalizing enclitic vowel *i found throughout the Athabaskan family (Sapir 1923). The vowel has been lost in Hupa, but its phonological effects remain in alternations affecting vowel length, spirantization, voicing, and the timing of glottalization in closed stems and as an offglide in open stems (Sapir and Golla 2001:823-824).

[20] These and other Hupa neologisms are analyzed as metonymic in the semantic analysis of Ahlers (1999:176-178).

| form | gloss | literal translation |
|---|---|---|
| qo:-nehwa:n | ‘rice’ | ‘worms – it resembles’ |
| k’imit’-nehwa:n | ‘watermelon’ | ‘belly – it resembles’ |
| k’iqude’-nehwa:n | ‘carrot’ | ‘willow root – it resembles’ |
| k’iɬixun-nehwa:n | ‘sheep’ | ‘deer – it resembles’ |
| k’e:dze’-nehwa:n | ‘corn’ | ‘penis – it resembles’ |
| tsintehɬ-nehwa:n | ‘canteen’ | ‘turtle – it resembles’ |

Table 5.8: Neologisms with *-nehwa:n* ‘it resembles’

Notably rare in the dossier of Hupa neologisms denoting introduced items is the strategy sometimes known as loan translations or calques, where the abstract structure of a source language word is borrowed but is fleshed out with recipient language morphemes (Haugen 1950). A handful of possible calques are identified in footnotes to the texts in Sapir and Golla (2001). The most likely are *k’ich’ing’* ‘overturned’ used to mean ‘upset (emotionally)’ (text 12, line 3), and *dahk’iɬtse:tɬ’* ‘he broke it (a bone) in two’, where the prefix *dah-*, normally an adverbial prefix meaning ‘above, on top’, is added to the verb *k’iɬtse:tɬ’* ‘break something’, possibly a calque of English ‘break up’. Contemporary speaker Verdena Parker reports that the Hupa noun *-xe’* ‘foot’ (the body part) was recruited to talk about the unit of measurement, as in *ɬa’ xoxe’* ‘one foot’.[21] Further analysis of other text collections (Goddard 1904, Golla 1984, and contemporary recordings with Verdena Parker) may reveal other such cases, but they appear to be rare overall. Just as was found with regular English loanwords, English seems to have had a minimal direct impact on the Hupa lexicon in this regard.

## 2.4. Diffusion of Lexical Innovations

The three strategies for labeling introduced items outlined in §2.1-2.3 are not mutually exclusive. It is common to find Hupa neologisms alongside loanwords, for example, even ones that appear to be fairly well-established and are not mere nonce borrowings. Examples include the following:

| gloss | form | strategy |
|---|---|---|
| ‘matches’ | me:jis | loanword |
| | miɬ-k’idiɬwis (< ‘fire-drill’) | adaptation |
| | xong’-min (lit., ‘fire – for’) | neologism |
| ‘coffee’ | Ka:whe | loanword |
| | ta’na:n-ɬiwhin (lit., ‘water –dark’) | neologism |
| | xosa:wilin-ɬiwhin (lit., ‘it flows into mouth – dark’) | neologism |
| ‘corn’ | Ka:n | loanword |
| | k’e:dze’-nehwa:n (lit., ‘penis – it resembles’) | neologism |
| ‘chicken’ | jigin | loanword |
| | mining’-q’eh dahna:da’ay (lit., its face – along – it lies on) | neologism |

Table 5.9: Multiple strategies for labeling introduced items

[21] The same calque is also found in the neighboring language Karuk (Bright 1952).

One of these examples, *Ka:n* alongside *k'e:dze'-nehwa:n* 'corn', is found as early as Curtin (1888-1889:95). It is also extremely common to find new items labeled with different neologisms, as in the case of 'coffee' in Table 5.9, or the following examples:

| gloss | form | literal translation |
|---|---|---|
| 'bathtub' | me' na'me: | 'in it – one swims/bathes' |
| | me' tehna:na'k'iɬdiw | 'in it – one washes up' |
| 'eyeglasses' | xona:' wilchwe:n | 'his eye – it is made to be' |
| | xona:' na:wilchwe:n | 'his eye – it is made again to be' |
| | xona:'-xwo-sa'a:n | 'at his eye – a round object lies' |
| | xona:' tse:ɬch'e' | 'his eye - iron' |
| 'cup, glass' | me' ta'dina:n | 'in it – one drinks' |
| | me' ta:wina:n | 'in it – something is drunk' |
| 'school' | me' 'a:k'iwilaw ya(')nehɬ'e:n | 'in it – writing – they look at' |
| | 'a:k'iwilaw me' ch'nehɬ'e:n | 'writing – in it – one looks at' |

Table 5.10: Multiple neologisms for labeling introduced items

In some cases, entirely different neologisms are used, although there may be partial similarities (such as both deverbal formations meaning 'bathtub' formed with the locative *me'* 'in it'). In other cases, however, the same abstract lexical item ('theme' in the Athabaskanist idiom) is used with different inflectional or closely related derivational variants: the verb 'look at' occurs with or without the plural prefix *ya:-* in 'school', 'drink' is found with passive vs. active voice in 'cup', and *wilchwe:n* 'made to be' appears with or without the iterative/reversative prefix *na:-* in two of the neologisms denoting 'eyeglasses'.

Overall, there a great deal of variation documented in nouns labeling introduced items in Hupa; in this regard it is similar to the *ni-* ~ *'i-* variation discussed in chapter 3. Contemporary speaker Verdena Parker describes the situation thus in relation to the many words denoting 'eyeglasses': "some families make a name of something for their own, and they call it that in families, so this is probably one of them ... [because] that's a new thing, so that's why I think... in families they called it different things" (VP-230308-10). She further suggests that this could be a reflection of micro-regional variation within Hoopa Valley, specifically differences between families living in the northern vs. southern end of the valley (cf. §5.1 in chapter 3).[22] These statements are reminiscent of the findings of Bright (1952:57) in his discussion of contact-induced lexical innovations in Karuk. Bright found that it would be impossible to make a complete list of all such innovations because "... different speakers tend to use different words for new culture objects. That is, there have been repeated borrowings, meaning transfers, and inventions, but many have not diffused widely enough to become standard for the language" (Bright 1952:57). A similar situation seems to have obtained in Hupa.

The precise nature of this variation is uncertain, but it is likely that many Hupa speakers had at least passive command of more than one variant. This is evident insofar as different ones are sometimes offered by contemporary speaker Verdena Parker, even in a single elicitation

[22] As noted in chapter 3, this reflects a traditional division between the northern and southern ends of Hoopa Valley that was important in many ceremonial contexts (Goddard 1903:12).

session; other times she recognizes a form from other documented sources but has a different way of expressing the same thing. And while many aspects of Curtin's (1888-1889) elicitation are unknown – how many speakers he worked with and in what combinations – the fact that some items in his vocabulary have multiple Hupa translations suggests that some speakers might have been offering more than one. While the details of the social-indexical values of different variants of particular items are probably for the most part unrecoverable, it is likely that such values existed, if Mrs. Parker's strong reaction to some forms being from "down the valley" or "Redwood Creek" is any indication.

There are several ways to interpret the failure of these recent lexical innovations to spread more widely through the Hupa-speaking community and for a single variant to be used by all speakers. The first and perhaps simplest explanation is that there simply wasn't enough time for the large number of labels for new items to stabilize between their introduction in Hoopa Valley and the point at which the language became obsolescent. Another is to see it as pathological, a sign of incipient language obsolescence, a view found in studies such as Cook (1989). This might be due to communicative isolation of sub-populations of Hupa speakers as Hupa was used in fewer and fewer public domains. Or it might be due to the development of highly charged social-indexical interpretations of variation reflecting social fissures that were a direct consequence of the colonial experience, a view that has been proposed by Furbee (1995) in her important unpublished paper on language obsolescence in the Siouan language Chiwere (cf. §3 in chapter 1). Furbee argues that variation introduced in the post-contact era is indicative of "family dialects" that developed due to divergent accommodation – social pressure to distinguish one's speech from that of other speakers. Her theory essentially maintains that innovations failed to diffuse widely not because speakers were unaware of how other people said things, but because they were not inclined to follow their lead.

Still another view is to consider lexical and other variation a normal part of any language. From this perspective, variation involving nouns denoting introduced items is not necessarily a sign of language attrition, nor is it a linguistic problem that would inevitably have been solved had Hupa remained a vital language for a few more decades. According to this view, it may be the case that hyperlocal "family dialects" were a longstanding fact of Hupa linguistic culture even in pre-contact times. The semantic transparency of many Hupa nominal formations means that lexical variation would not have presented major impediments to communication. Without functional pressure to eliminate variation on purely referential grounds, there might have been room for considerable variation of this sort. This is not to suggest that Hupa dispensed entirely with lexical changes that spread throughout the language, since as discussed in chapter 2 many innovative deverbal nominals did achieve this status in the development of Hupa from Proto-Athabaskan. Rather, the existence of a high degree of lexical variation might have been diachronically stable.

Essentially this point has been made with regard to other indigenous speech communities in North America. The notion that speech communities might tolerate more variation than is commonly acknowledged is suggested by Scollon (1979), who noted a high degree of variation in the articulation of Dene Sųłiné consonants spoken at Fort Chipewyan. Even Cook (1980), who disagrees with some of Scollon's specific points of interpretation, maintains that some aspects of Dene Sųłiné variation "have existed perhaps from the beginning of Athabaskan history." More recently, in a discussion of what he calls "personal dialects" in one Munsee community, Goddard (2010) argues that the expectation of a high degree of uniformity in a speech community is an artifact of the Bloomfieldian notion of "community norm" as the

basis for linguistic description. Goddard found in his fieldwork that "[a]gain and again, fully competent speakers who had been born on the same six-square-mile tract of land turned out to have different ways of saying everyday things" (2010:24), differences that cut across many linguistic domains.

That Hupa might have tolerated a fair degree of lexical variation even in pre-contact times is suggested by the existence of multiple words documented not just for introduced items, but for traditional Hupa items as well. Some examples for technologies related to hunting and fishing gleaned from the Hupa Language Dictionary are given in Table 5.11:

| gloss | form | |
|---|---|---|
| 'arrow' | na:tse:s | |
| | dahch'iwile:l | (lit., 'one holds them up') |
| 'arrowhead(s)' | dinday | |
| | dahk'islay | (lit., 'objects that lie on top') |
| 'A-frame lifting net' | k'ixa:q' | |
| | miɬ-dahsiday | (lit., 'with it – one sits on top') |
| 'small dip net' | k'ixa:q'ich | (*k'ixa:q'* + diminutive) |
| | miɬ-no:'oɬwul | (lit., 'with it – one throws down') |
| 'fishing platform' | ta:lqe:t | (lit., 'what is shoved into the water') |
| | dahk'iwe:wita:n | (lit., 'stick-like object that it put on top') |

Table 5.11: Multiple lexical items labeling traditional Hupa items

Here too it is possible that the documented variation is a relatively recent phenomenon – that some variants might have entered Hupa when Chilula and Whilkut people from Redwood Creek were resettled in Hoopa Valley, for example. Or perhaps the use of different morphologically complex forms even for traditional Hupa items was due to lexical innovations in the post-contact period for the reasons given by Furbee (1995). It is difficult to argue definitively against such views given the nature of the data at hand, but one consideration is that such variation is found even in early sources of Hupa documentation. For example, for 'arrowhead' Curtin (1888-1889) registered both *dinday* (as *xo'ji-dinday*) and *na:tse:s k'islay* (with *k'islay* the derivational base of *dahk'islay* in Table 5.11). And Goddard (1905:191) transcribed one speaker using <kin-nal-mats> and <kin-na-kyū-wil-mats> glossed as 'cribs (of hazel)' in adjacent sentences of the same text.[23] The earlier such variation is found, the less likely that it is due to factors introduced after contact.

These considerations suggest that lexical variation in Hupa encountered in vocabulary denoting items introduced in the post-contact period is neither pathological nor reflective of deeper social divisions than existed in pre-contact society. Hupa was in all likelihood perfectly able to cope with lexical and other variation for centuries, at many levels of granularity (individual, and involving various sub-portions of the overall Hupa-speaking population). From this perspective, lexical variation introduced in the wake of the colonial encounter is simply a

[23] Goddard (1905:22) gives the literal translation of both forms as 'around it is coiled'. Goddard's <kin-nal-mats> appears as *kin-na:lma:ts'* in the Hupa Language Dictionary glossed as 'storage cask' with literal translation 'wood-encircled'.

continuation of earlier Hupa linguistic culture. The more general point – perhaps the deeper philosophical division separating demographic determinism from indexical approaches invoking Trudgill's (2008a) "identity factors" – is that eliminating variation is not an inevitable outcome of human social interaction given enough time. Nor is variation in minority speech communities necessarily a sign of incipient obsolescence – even if in this particular case variation is observed in the documentary record of a language that is today no longer widely spoken.

## 3. Non-Lexical Effects of English

The overall impression created by the data presented in §2 is that Hupa-English bilingualism did not have a major impact on the Hupa lexicon. For the most part, Hupa people coped with the plethora of new material culture introduced in the mid-19th century by adapting existing words and coining new ones using Hupa's productive morphological resources. The number of words borrowed from English into Hupa is quite small, although there may have been more extensive use of English terms as nonce borrowings among bilinguals than is found in the documentary record. Accordingly, loanwords did not penetrate very deeply into the Hupa phonological system, the main effect being to expand the reach of some phonemes that had previously been somewhat marginal and/or specialized. Moreover, what loanwords exist are exclusively nouns: there are no examples of English verbs, function words or bound morphemes as established borrowings. Even the nouns are all non-basic vocabulary and related to introduced items – there are no cases where traditional Hupa technologies came to be replaced by English words, for example.[24]

Given the minor impact of English on the Hupa lexicon even in the face of intense functional pressure to label introduced items, Thomason and Kaufman's (1988:74-76) borrowing scale leads to an expectation that non-lexical effects of English on Hupa will also be relatively minor.[25] At the "casual" end of the borrowing scale, "[s]yntactic features borrowed at this stage will probably be restricted to new functions (or functional restrictions) and new orderings that cause little or no typological disruption" (Thomason and Kaufman 1988:74). It is in fact difficult to find clear cases of syntactic differences between Hupa as spoken in the early 20th century (in the texts in Goddard 1904 and Sapir and Golla 2001) versus later. This may be because none exist, but the study of Hupa syntax is still very much in its infancy and detailed analysis of texts from different periods is required before these kinds of subtler distributional effects can be explored systematically.

One area where non-lexical effects of English on Hupa might be expected is word order. Thomason and Kaufman maintain that "word order seems to be the easiest sort of syntactic feature to borrow or acquire via language shift" (1988:55). Most Athabaskan languages have a fairly rigid SOV word order. Hupa, however, has more flexibility, with postverbal NPs (includ-

[24] Although the form of the English plural suffix was borrowed in several nouns (*'e:bilos* for 'apples', *bijis* for 'peaches' and so on), the plural semantics of the suffix was not necessarily borrowed. Golla (1970:27) documents *'e:bil* 'apple', but Verdena Parker produces singular forms such as *ła' 'e:bilos* 'one apple' and *ła' bijis* 'one peach'. See Bright (1952) for discussion of similar data in Karuk.

[25] As noted in §1, the overall long-term profile of contact between Hupa and English is one of language shift rather than borrowing in Thomason and Kaufman's narrow sense (1988:37), with Hupa speakers shifting to English. However, effects of English on Hupa in the interim can be characterized as borrowing (speakers with Hupa as L1 incorporating features of English as an L2), and hence the borrowing scale is relevant.

ing postpositional objects) commonly found in texts.[26] This is illustrated in example (12) from consecutive lines of a text in Sapir and Golla (2001, 22.22-23, transliterated to the contemporary community orthography). In (12a), the NP *k'it'ung'* 'leaves' is the object of the postposition 'after' and occurs preverbally; in (12b), it is a direct object and appears postverbally:

| | | | | | | |
|---|---|---|---|---|---|---|
| (12a) | hayahujit | k'itung' | mixa: | te'iwha:wh | tse:mitah-q'eh | yiduq |
| | then | leaves | after | I go off | mouth of Hostler Creek – along | up country |
| | 'Then I go off in search of (maple) leaves up along Hostler Creek.' | | | | | |

| | | | | |
|---|---|---|---|---|
| (12b) | hayah | 'e'iwhchwe' | hay | k'itung' |
| | there | I always make, pick | the | leaves |
| | 'I always collect the leaves there.' | | | |

Conathan (2004:76-79) argues that postverbal nominals in Hupa tend to be referents that have previously been introduced to the discourse, i.e., are old information. Noting that this order is rare in the Athabaskan language family (Koyukon being an exception – Thompson 2000), Conathan suggests that the relatively flexible word order in Hupa is likely due to contact with neighboring non-Athabaskan languages.

Newbold (2010) has studied the phenomenon of Hupa postverbal NPs in detail for a wide range of argument types using a subset of the texts in Sapir and Golla (2001). Her study on the whole confirms Conathan's generalization that postverbal NPs tend to be discourse-old referents but also reveals some systematic exceptions, such as presentational locative inversions and copular clauses. Crucially for present purposes, the data tabulated in Newbold (2010) show that 45 out of 156 (29%) overt subject NPs and 51 out of 181 (28%) overt direct object NPs in her sample appeared in postverbal position. That is, regardless of grammatical function and status as discourse-new versus discourse-old, just over a quarter of all overt NPs occurred postverbally.[27]

Although the simple division of preverbal vs. postverbal is a very coarse-grained metric (ignoring the more nuanced information-structural effects discussed by Newbold), it nonetheless provides a baseline for considering the effects of English bilingualism on word order in Hupa. English is canonically discussed as having a fairly rigid SVO word order in matrix declarative clauses. While postverbal subjects and preverbal objects are perhaps more common in English than is often supposed (cf. Birner and Ward 1998), especially in spoken discourse, it is unlikely that the aggregate incidence of postverbal main clause subjects approaches the 28% found in Hupa, or that preverbal direct and oblique objects approach 70%. This leads to the following prediction, as a starting point at least: if Hupa word order was influenced by Hupa speakers' English bilingualism, this might be reflected in less frequent use of postverbal subjects and more frequent use of postverbal objects.

Analysis of a selection of texts in Goddard (1904) reveals a somewhat lower rate of postverbal subject NPs than Newbold (2010) found in her study of the texts in Sapir and Golla (2001), and a higher rate of postverbal object NPs:

---

[26] Hupa is arguably still basically SOV typologically: although not unattested, postverbal nominals are rare in embedded clauses.

[27] Newbold (2010) found that the same was true for postpositional/oblique NPs as well, with 29 out of 106 (27%) oblique NPs occurring postverbally.

| | subject | object |
|---|---|---|
| preverbal | 88 (79%) | 73 (65%) |
| postverbal | 23 (21%) | 40 (35%) |
| total | 111 | 113 |

Table 5.12: Distribution of preverbal and postverbal nominals in Goddard (1904)

The extent to which the speakers Goddard worked with were bilingual in English is not known, although he discusses working through interpreters (1904:93), suggesting that some speakers at least were not fully able to communicate with Goddard with English as an elicitation metalanguage. Although the postverbal subjects are somewhat less frequent and the postverbal direct objects somewhat more frequent in Goddard's collection than they were for Sapir a generation later, the difference between the two corpora is not statistically significant ($p = .174$ for subjects, $p = .2407$ for objects).[28]

There are three sources of Hupa text material collected subsequent to Sapir's visit to Hoopa Valley in 1927: Woodward (1953) with Stella Jarnaghan, Golla (1984) with Ned Jackson, Louisa Jackson, and Minnie Reeves in the early 1960s, and a growing number of transcribed texts recorded with contemporary speaker Verdena Parker. Summaries of their pre- and postverbal subjects and objects are provided in Tables 5.13-5.15:[29]

| | subject | object |
|---|---|---|
| preverbal | 56 (84%) | 42 (75%) |
| postverbal | 11 (16%) | 14 (25%) |
| total | 67 | 56 |

Table 5.13: Distribution of preverbal and postverbal nominals in Woodward (1953)

| | subject | object |
|---|---|---|
| preverbal | 92 (93%) | 48 (63%) |
| postverbal | 7 (7%) | 28 (37%) |
| total | 99 | 76 |

Table 5.14: Distribution of preverbal and postverbal nominals in Golla (1984)

| | subject | object |
|---|---|---|
| preverbal | 84 (86%) | 53 (84%) |
| postverbal | 14 (14%) | 10 (16%) |
| total | 98 | 63 |

Table 5.15: Distribution of preverbal and postverbal nominals in texts by Verdena Parker

The data related to subjects can be summarized as follows. In all three collections, overt subject NPs appear postverbally less frequently than they do in Goddard (1904) and Sapir and Golla (2001). This is especially the case for the speakers Golla worked with in the early 1960s, who

[28] Statistical calculations in this section were made with a chi-square test using the prop.test() function in R (R Core Team 2013).

[29] Newbold (2010) includes an analysis of a subset of the texts recorded with Mrs. Parker that are included here. Her findings are consistent with these trends.

produced only seven tokens of postverbal subject NPs, and five of those were from just one of the three speakers (Minnie Reeves) that Golla worked with. Combining the figures from these three corpora and comparing them with the total for the combined Goddard and Sapir corpora, the difference between the two partitions is highly significant ($p < .0002$).[30]

The situation for object NPs is more complex. Stella Jarnaghan, the main speaker who worked with Woodward (1953), places overt objects postverbally at roughly the same rate as found in the Sapir and Golla (2001) corpus, and the speakers in Golla (1984) produced post-verbal objects at a rate comparable to that found in Goddard (1904). Contemporary speaker Verdena Parker places objects postverbally at roughly the same rate as she places subjects postverbally, substantially less frequently than what was found in the Goddard and Sapir corpora. Performing the same partition of the data for objects (Sapir and Goddard versus others), there is no significant difference between the two ($p = .3583$). This in itself can be taken as evidence that Hupa speakers did not alter their word order with respect to objects based on an English model. Moreover, if the data are partitioned to compare Mrs. Parker's use of postverbal objects with the other corpora combined, the difference becomes significant ($p < .02$). These results imply that Mrs. Parker uses the English-like postverbal objects even less than Hupa speakers in the past did.

The question is whether these differences in the distribution of overt subject and object NPs in the mid-20th century and later are an effect of Hupa speakers' bilingualism in English. The fact that postverbal subjects are less frequent is consistent with the hypothesis that the relatively rigid SVO word order of English might have influenced bilingual speakers to use VS order less frequently when speaking Hupa. However, the incidence of postverbal object NPs is no greater in later texts than it was in Goddard (1904). In all collections, the majority of overt objects are preverbal, and Verdena Parker, born later than all of the other documented speakers and without doubt a fully fluent speaker of both Hupa and English, uses postverbal objects less frequently than anyone. If the decline in the use of Hupa postverbal subjects is due to influence from English, it is surprising that there was no corresponding increase in the use of postverbal objects. In the absence of a theory of structural borrowing that predicts word order effects asymmetrically targeting subjects and objects, the decreased use of postverbal subjects in recent collections of Hupa texts does not seem to be straightforwardly due to contact with English.

What, then, explains the lower rate of postverbal subjects among the later Hupa speakers? The data discussed here relate to cases where reference is achieved with an overt nominal expression, whether a full NP or a pronoun. What is not shown is the incidence of reference achieved without a syntactically independent nominal expression at all. Hupa, like all Athabaskan languages, has robust nominal cross-referencing encoded in the verbal system, and it is extremely common for independent nominal arguments to be omitted entirely in texts. Such nominals are typically old information, i.e., precisely the kinds of things that would tend to appear in postverbal rather than preverbal position according to the analysis of Newbold (2010). It is possible that Hupa speakers in the mid-20th century and later tended to make use of the null pronominal option rather than the postverbal position prototypically used for discourse-old (or inferable) referents. This would tend to deflate the number of postverbal NPs and might asymmetrically affect subjects instead of objects insofar as subjects tend to be old information. Reduction of the use of postverbal subjects could therefore be due to incipient language attrition whereby the range of pragmatic and stylistic devices available in a language undergoes

[30] The difference is still significant even if the texts from Golla (1984) are excluded from consideration ($p < .02$).

contraction (cf. discussion of similar issues in Campbell and Muntzel 1989). If postverbal NPs are a marked option in Hupa, then decreases in their occurrence would suggest a diachronic trajectory preferring less-marked options. Determining whether this alternative explanation is empirically supported is a topic left for future research, but it is one in which English as a structural model is unlikely to play a role, since English does not typically sanction the complete omission of nominal arguments. Overall, even in an area of grammatical structure that is most prone to casual contact effects, there is no unambiguous evidence that Hupa speakers were directly influenced by English bilingualism.

## 4. Discussion

This chapter has considered ways that contact with English-speaking populations influenced the Hupa language in the period before it ceased to be acquired by children. The overall conclusion from the data presented above is that there was very little borrowing, lexical or structural, from English into Hupa, even among speakers recorded in the mid-20th century or later. This is similar to the case of Montana Flathead, discussed by Thomason and Kaufman (1988) as a typical case of gradual language obsolescence in indigenous languages of North America. Flathead "has apparently borrowed nothing from English except a few words," and the shift from Flathead to English "will have been completed within a few generations from the onset of intense cultural pressure from English speakers" (1988:101-102). This is roughly the same timeline of language shift that Hupa has experienced. The absence of pervasive structural borrowing and of lexical borrowing affecting basic vocabulary in Hupa is thus unsurprising, since these typically only develop after centuries of intense contact (Thomason and Kaufman 1988:41).

More unexpected, however, is the fact that there was not greater use of English-sourced material to label the vast quantity of introduced items that in principle might have been loci for lexical borrowing in the post-contact era, especially among fluent Hupa-English bilinguals in the 20th century. Given a choice between adopting an English word together with a new item versus generating a new word with Hupa-sourced material, Hupa speakers overwhelmingly opted for the latter strategy. As noted by Goddard (1905) and discussed in §2.3 above, this was a continuation of Hupa speakers' pre-existing tendency to coin new words using productive structural resources in their language, perhaps in part because they simply enjoyed doing so (Goddard 1911). This suggests an additional component of analysis to complement the approach to lexical acculturation in Brown (1999): quantifying the extent to which languages' pre-contact lexicons are composed of morphologically complex (and semantically transparent) words. This would be an enormous undertaking on the scale of Brown's hemispheric study, but could be fruitfully applied in more local, regional settings like northwestern California. It might also be useful to measure pre- versus post-contact rates of lexical diversity within communities, which would address the issue of the extent to which such variation is a product of the colonial encounter or a continuation of pre-contact linguistic culture.

With regard to structural borrowing, the word order data presented in §3 showed that some tendencies in the placement of overt NP arguments in Hupa seem to have changed between the turn of the 20th century and the mid-20th century, with postverbal subjects coming to be used much less frequently. However, this was not straightforwardly modeled on an English SVO pattern insofar as VO order did not become more common in the same period. Further study is needed to determine whether or not the observed changes can be correlated with some other factor such as overall decreased use of overt discourse-old NP arguments – factors that might

even suggest Hupa diverging with English tendencies in some respects. More generally, it would be of interest to discover areas of morphosyntactic variation and change in the documentation of Hupa; some ideas in this regard are included in chapter 6.

# Chapter 6: Conclusion

## 1. Main Findings

The focus of this dissertation has been the diachronic development of the Pacific Coast Athabaskan languages, from their earliest separation from the rest of the Athabaskan language family over a millennium ago to the effects of language and dialect contact in the past century and a half. In addressing language change at such diverse time scales and with sometimes sparsely documented languages the study has of necessity been eclectic, with computational, comparative, and philological methods applied to data drawn from a variety of published and unpublished sources and original fieldwork with a contemporary speaker of Hupa. Portions of the analysis have been speculative simply because the linguistic and social documentary record is in many cases too meager for firmer conclusions to be drawn. Nonetheless, the common thread throughout has been the linguistic phylogeny as an object of study that can be fruitfully examined at different granularities. The main results of this line of inquiry are presented in this section.

The main finding of chapter 2 is that Bayesian phylogenetic analysis of the Athabaskan language family detects a Pacific Coast Athabaskan subgroup subsuming the Oregon and California Athabaskan languages. This result is robust across conditions: with multi-state and binary codings, with or without polymorphic characters, with lexical and non-lexical characters, and with different model assumptions about the constancy of rates of evolution across lineages. Additional scrutiny of this result is warranted, especially in light of the conceptual issues surrounding the use of non-lexical characters in the rates-across-sites model used in this study. Uncertainty also remains concerning the locus of differentiation of this subgroup given Golla's (2011) observations about the lack of shared lexical innovations in PCA languages in semantic domains that would indicate a shared history of migration into the region. However, the results obtained here lend support to the view of Hoijer (1960) that a Pacific Coast Athabaskan subgroup is part of the true evolutionary history of the Athabaskan language family.

One of the central theoretical questions explored in chapters 3 and 4 was the extent to which social-indexical values associated with linguistic variables have empirical consequences for understanding the transgenerational outcomes of variation introduced when mixed dialect communities are formed. Theories that afford a prominent role to indexical considerations were compared with approaches that seek to largely remove them from the equation and replace them instead with demographic factors such as relative population size and density (Trudgill 1983, 2004, 2008a,b). Evidence presented in chapter 3 was equivocal in this regard. While indexical theories certainly could be invoked to explain the outcomes of dialect contact in Hoopa Valley, they do not seem to be required: simple majority-rule principles also correctly predict the leveling of one parameter of variation ([k] vs. [x]) and the non-leveling of another (*ni-* vs. *'i-*). It was suggested, however, that demographic approaches might have difficulty explaining the source of [k] ~ [x] variation in Whilkut, arguably an extension of a pre-contact status quo. This is because the group that exhibited the variation is likely to have had a population at least as large and as dense, if not moreso, than the other Athabaskan groups in their immediate vicinity (Baumhoff 1958).

The case study of dialect contact at Siletz in chapter 4 argued more strongly that a theory of demographic determinism does not predict the retention of minority variants in the reservation setting (Galice, Mikwanutni) or that minority variants might in fact spread in such cases (fronting of /s/). Overall, the evidence considered in this chapter suggests that something more

than simple demographic arithmetic is sometimes needed in order to understand the diachronic trajectory of dialect contact, even in the limited tabula rasa scenarios that Trudgill (2004, 2008a) considers. Trudgill's aims in developing a non-indexical theory of contact are commendable: he seeks a parsimonious and predictive theory that does not depend on the vagaries of social meanings associated with linguistic variation in all cultures throughout human history. There is certainly an appeal to such minimalist approaches, and no one would seriously dispute that demographic factors are an important consideration in many, perhaps most, dialect contact situations. However, it remains to determine whether they can in fact explain the results of dialect contact (and of the spread of variants within a community, and language change more generally) across a broad range of cases. The evidence considered in this dissertation suggests that in some cases at least something needs to be said about the social and historical contingencies of dialect contact in order to understand the observed outcomes.

The main finding of chapter 5 is that Hupa's contact with English had only a minor impact on how people spoke Hupa, despite what appears to have been a high incidence of Hupa-English bilingualism by the turn of the 20th century. This is reflected in the overall lack of lexical borrowing even for new items introduced at the time of contact and in the absence of direct English interference in the domain of main-clause constituent order. This was interpreted as a continuation of conservative linguistic stances that existed in pre-contact times, both within the Athabaskan language family (Sapir 1921) and in California considered as a linguistic area (Golla 2011).

Interesting aspects of the theme of linguistic conservatism are found elsewhere in this dissertation as well, especially when analyzed in terms of Hill's (2001) theory of localist versus distributed stances towards the adoption of new linguistic material. Consider first variation involving [k] and [x] in the Whilkut dialect of Hupa. Chapter 3 presented evidence that this variation extended back into pre-contact times, and moreover that it may have been due to relatively recent diffusion of the [k] variant from a neighboring Eel River language. This raises the following question, however: if Athabaskan groups in the region were so linguistically conservative, borrowing but little from their non-Athabaskan neighbors over many centuries of contact (Golla 2011:77), why were the Whilkut open to adopting the [k] variant from other Athabaskans?

A partial answer to this question can be developed by pointing out that, despite an overall tendency towards linguistic conservatism among Athabaskan-speaking groups with respect to non-Athabaskans, there is evidence that they have historically been receptive to borrowing from one another. This is one way to understand Krauss' (1973) view of the Athabaskan family as an extended dialect network lacking barriers to the diffusion of innovations, and has been demonstrated by Kari (1977), for example, for contact involving the Alaskan Athabaskan languages Ahtna and Dena'ina.[1] This may be due to structural factors, if borrowing from one's linguistic congeners is easier than borrowing from unrelated languages because typological distance is not a consideration. However, it can also be understood in sociocultural terms. Linguistic conservatism is not an absolute, across-the-board phenomenon, but may be more or

[1] It is perhaps important to point out that Dena'ina is one of the Alaskan languages that appears to have been susceptible to at least some kinds of areal contact effects involving non-Athabaskan languages, in particular changes in its vowel system under the influence of Yup'ik Eskimo (Krauss 1973:948), suggesting that it might be exceptional within the family in its susceptibility to language contact effects. Dena'ina is also the Athabaskan language that was the most open to lexical borrowing from European colonial languages (Brown 1999:85).

less salient with respect to innovations introduced by different reference groups. Cast in terms of Hill's (2001) theory of stance-taking, the Whilkut might have adopted a distributed stance with respect to neighboring Athabaskan groups like the Nongatl and lower Trinity Hupa (which would also have facilitated leveling in the reservation setting), but maintained localist stances with respect to non-Athabaskan languages they were in contact with.[2] Affordance for this relativized, dyadic view of linguistic stances is implied in Hill's discussion when she notes that "[t]he speech of any single person and the patterns of variation in a community will always be the product of a combination of these two strategies" (2001:261).

Of course, English had a catastrophic influence on Hupa overall in coming to replace it as the primary language spoken in Hoopa Valley. This leads to a second consideration related to the fact that there was so little direct influence from English even as language shift was underway in Hoopa Valley.[3] In Hill's theory of localist and distributed stances, language shift occurs when a community adopts a distributed stance towards an exogenous language. Given the fact that Hupa speakers did not, in the transitional period of Hupa-English bilingualism, engage in borrowing from English suggests that a further refinement of Hill's theory is needed. The relativization of distributed stances must extend not only to the various reference groups that speakers come into contact with, but also to particular linguistic instantiations of them: in the present case, a distributed stance favoring the adoption of English as a whole, but disfavoring the adoption of English elements into Hupa. The two issues are not, in all likelihood, unrelated, since a distributed stance realized as code mixing versus the wholesale adoption of a new but distinct code is likely determined by how one strategy versus another would be perceived by various sets of interlocutors and what material advantages would be conferred as a result.

## 2. Future Directions

Each chapter of this dissertation has included suggestions about areas for future investigation, some of which are summarized and expanded here.

### 2.1. Computational Studies

As discussed at length in chapter 2, one of the main points of disconnect between current computational work in historical linguistics and traditional comparative approaches is in the treatment of non-lexical data. The results in chapter 2 are based in part on an attempt to incorporate non-lexical characters into the analysis, but using computational models that may be conceptually inappropriate. Moreover, non-lexical (especially morphological) characters were collected opportunistically rather than based on a principled theory of why they might be useful for phylogenetic inference. Thus, moving forward it will be necessary both to employ more appropriate computational models (perhaps using Maximum Compatibility as in Nakhleh, Ringe, and Warnow 2005 and related studies) and to assemble a theoretically grounded set of non-lexical characters.

One of the most fruitful areas for future exploration in the latter regard is the use of

[2] This is also suggested by the asymmetric use of Whilkut in a multilingual household observed by Reichard (1922) in her field work at Blue Lake, noted in §4.1 of chapter 3.

[3] This again comes with the caveat that some kinds of contact effects like nonce borrowing and other species of code switching may be under-documented in Hupa. Unfortunately, it is unlikely that this gap in the documentary record will ever be filled.

templatic ordering as a basis for establishing subgroups – taking Rice's (2000) "local variability" in Athabaskan verb morphology as a means to infer historical relationships. Ideally this would be done in terms of the substantive morphological content of the morphemes involved, but could also be pursued in terms of abstract ordering considerations (e.g., "plural precedes incorporate" regardless of the particular plural prefix involved, as in character #7 in Appendix B to chapter 2). One advantage of this for computational work is that the positions in a verb template might have a direct analogy with biological evolution, where the sites for evolutionary development involve strings of bases occupying positions within an inherited DNA template. This might be a fruitful testing ground for computational modeling of morphological change that could be applied in other language families whose morphology is robust enough to support such analysis.

It can also be noted that the lexical component of the computational study presented in chapter 2 could be enhanced in a variety of ways as well, especially by using a more robust wordlist (Swadesh-200 instead of Swadesh-100, or the Leipzig-Jakarta list discussed in Tadmor, Haspelmath and Taylor 2012), and by explicitly including meanings that are likely to be well-attested across the Athabaskan family and potentially diagnostic for subgrouping.

## 2.2. Morphosyntactic Variation and Change

The particular cases of variation and change discussed in chapters 3, 4, and 5 were for the most part phonological and lexical, mainly because these are the areas where the documentation of PCA languages is robust enough to allow for comparison from the earliest stages of language and dialect contact in the mid-19th century to the present day. Chapter 5 did, however, consider changes with respect to main clause constituent order in Hupa, probably the only PCA language with adequate text material to permit this kind of diachronic analysis. This suggests another fruitful area of investigation of Hupa: discovering morphological and syntactic-pragmatic parameters of variation in the language that underwent changes in the reservation setting, due either to contact between Hupa dialects or to contact with English (either with English as a direct model for change or indirectly as a catalyst for language attrition of various sorts).

For example, there is evidence that there was a degree of variation in the application of a phonological rule deleting /w/ after /n/, typically in the context of /n/-final allomorphs of aspect prefixes in immediate pre-stem position.[4] Some speakers in the 20th century seem to have generalized non-application of the rule, perhaps a species of paradigm leveling that created greater transparency in the morphological realization of /w/-initial roots. Variation involving morpheme ordering at the left periphery of the Hupa verb template has also been documented, e.g., the iterative/reversative prefix *na:-* with respect to the bipartite reflexive element *'a:-* …

[4] This was originally noted by Goddard (1905) in entries for particular verb roots, for example: "-wan, -ñan, -wûn, -ñuñ; to sleep. This root usually appears in the form of -ñan or -ñûñ, w following -n of the preceding syllable of the definite tenses being assimilated to it" (p. 223). Exceptions to this generalization are also found however: in Goddard's entry for the root *-wiwh/-we:n* 'carry on the back', alongside <ya na kiñ en> 'he packed up' (with /w/ deleted) there is <ya kin wen ne> 'he had carried it off' and <ya-kiñ-wūw> 'carry it' (with /w/ retained) (1905:226).

*di-*.[5] Analysis of Hupa's complex verbal morphology, with its discontinuous dependencies across templatic positions and stem alternations, is likely to reveal more cases of this sort, which could shed light on how morphological variation was resolved in the context of contact in the reservation setting.

Casual analysis of Hupa texts has not yet uncovered parameters of syntactic variation across speakers or corpora collected in different decades that are comparable to the word order effects discussed in chapter 5. However, more rigorous analysis and tabulation of data may reveal subtler differences, and these could in turn allow for exploration of the extent to which changes occurred that are attributable to bilingual interference modeled on English. For example, there are many meaning types that can be encoded either analytically or morphologically in Hupa, and one might expect to find skewing towards patterns that are more typologically consistent with English. Hupa has both analytic and morphological causatives, and speakers with more proficiency in English in recent decades might have tended to use more analytic constructions (although it should also be noted that the types of causation encoded in the different strategies are systematically different – cf. Escamilla 2012). Similarly, Hupa often encodes path information morphologically in the verb, but can also encode it (sometimes redundantly) in satellite postpositional phrases – English influence might have led to less frequent use of the verbal encoding strategy in later decades. Impressionistically there do not seem to be major differences in these areas in earlier versus later corpora, but more precise analysis is needed.

Taken together, these topics point to a nascent diachronic study of Hupa morphology, syntax, pragmatics, and discourse stylistics that interlocks with the complex nexus of cultural and linguistic contact found in Hoopa Valley from the mid-19th century to the present. In each case, a negative result quantifying the absence of changes would be of as much interest as a positive one would be, since retention of the status quo when language and dialect contact occur is as much in need of explanation as change is (an inversion of the actuation problem in historical linguistics discussed by Weinreich, Labov, and Herzog 1968).

## 2.3. Heterogeneous Profiles of Language Shift

One of the interesting results of chapter 2 (§5.2) involves differences in the level of support for the California and Oregon Athabaskan subgroups depending on the nature of the characters being considered. When lexical characters alone were analyzed, California Athabaskan was only weakly supported as a subgroup whereas Oregon Athabaskan received strong support (as in Hoijer 1962). When non-lexical characters were considered instead, support for California Athabaskan was roughly the same as found with lexical characters (only marginally higher), but support for Oregon Athabaskan was relatively weak. This raises the question of whether there are principled reasons that linguistic diversity in Oregon Athabaskan was concentrated in phonology and morphology, whereas linguistic diversity in California Athabaskan appears to have encompassed the lexicon as well – not via lexical borrowing, as might be expected in the "maintenance" profile of language contact in the typology of Thomason and Kaufman (1988), but instead via extensive coining of new vocabulary, in Hupa especially.

This difference is all the more puzzling since both the California and Oregon groups are likely to have reached their pre-contact territories via similar historical processes. Jacobs (1937)

[5] For example, 'mirror', lit. 'in it – one looks back at oneself', is rendered as *me'-na:'a'dinehɬ'e:n* (iterative precedes reflexive) and *me'-'a:na'dinehɬ'e:n* (iterative splits the bipartite reflexive) (Verdena Parker, LA 256.33-033_B, 10:15).

proposed a scenario whereby the Athabaskan migration into the region was achieved via gradual, village-by-village movement of populations over the course of many generations from relatively sparsely populated upstream areas into more densely populated downstream areas.[6] Evans and McConvell (1998) have argued that such upstream-to-downstream migrations result in substrate effects of the kind associated with language shift (see also Hill (to appear) for an application in this idea in California). In the scenario proposed by Jacobs, where the spread of Athabaskan languages was due to intermarriage and language shift, the non-lexical diversity in Oregon Athabaskan is expected in Thomason and Kaufman's (1988) typology; the lexical diversity in California Athabaskan is not.

Resolving this issue is beyond the scope of the present study, but the outlines of a solution can be found if it is possible to consider the high degree of lexical innovation in California Athabaskan as a species of substrate effect, hence more closely connected with language shift than with maintenance. Conathan (2004:81-90) considered some of the lexical innovations in Hupa to be cases of functional convergence – languages becoming more similar to one another in abstract semantic and pragmatic domains without corresponding convergence in form or structure. Conathan noted that there are a number of morphologically complex nominals in Hupa and the neighboring non-Athabaskan languages Yurok and Karuk that are loan translations of one another. For example, 'otter' in Hupa is *łoːqʼi-yiditile*, lit. 'what is attracted to salmon'; in Yurok, *nepeʼwish-neg* 'one who regularly eats fish'; and in Karuk, *ʔamváːmvaːn*, lit. 'salmon-eater'. These can be plausibly treated as being due to bilingual speakers transferring the semantic content of complex nominals from one language into another, perhaps a consequence of language shift.[7] This would not in and of itself explain the different outcomes in California and Oregon, but at least would make them no more mysterious than the unpredictability of language shift in other situations: Thomason and Kaufman themselves emphasize "the impossibility of making specific predictions about the outcome of interference, even in comparable shift situations" (1988:121).

## 2.4. Hupa among Contemporary Speakers

One more area of potential interest would be to study the effects of English on Hupa as spoken by a broader range of remaining first-language speakers. Contemporary speaker Verdena Parker, who has been working with the present author and other researchers at UC Berkeley

[6] Jacobs points out that the paradox of this situation is that relatively wealthy and densely populated downstream communities shifted linguistically in favor of the languages spoken by relatively impoverished newcomers moving from upstream. One possible mechanism for such downstream migration is a combination of patrilocal settlement in marriages where bride-price was paid in full, and matrilocal settlement where only a portion of bride-price was paid ('half- marriage') (Goddard 1903:56; Drucker 1937; Gould 1966). This scenario would have been highly favorable to an upstream-to-downstream movement of population, both because downstream communities could afford more brides and because upstream communities might have to engage in half-marriage arrangements more often. There were apparently limitations on this and other factors favoring downstream spread of languages: Jacobs notes that the Athabaskans only spread all the way to the coast along certain watercourses and failed to supplant languages in the most densely populated areas.

[7] It is important to point out, however, that there are also many morphologically complex lexical innovations in Hupa that are non-convergent in nature. For example, Hupa vocabulary for the Swadesh-100 glosses 'red', 'head', 'tongue', and 'heart' are all morphologically complex innovations that make Hupa less similar to Yurok and Karuk insofar as the corresponding items in those languages are morphologically simplex.

since 2005, may be exceptional insofar as she continued to speak Hupa daily with members of her family until the late 1990s. Major differences in how she speaks Hupa as compared with previous decades of documentation have not yet emerged (although there are more subtle differences, such as tendencies in the placement of overt NPs noted in chapter 5). Close examination of the speech of other speakers documented in recent years might reveal a greater range of variation in the last generation to acquire first-language fluency in the language. One such study has already been conducted by Gordon and Luna (2005), a surprising result of which is the discovery of more complex patterns of stress assignment among recent speakers than are found in the texts transcribed by Edward Sapir (Sapir and Golla 2001). Similar transgenerational variation might emerge in a number of domains, e.g., phonological transfer effects of the sort discussed by Babel (2009b), which are commonly found when wholesale language shift is underway in a speech community. A corpus of recordings and transcriptions prepared with speakers in the 1990s have recently been deposited by Sean O'Neill in the Survey of California and Other Indian Languages, and analysis of their speech may be illuminating in this regard.

Also of interest would be a systematic study of Hupa as spoken by second-language learners. This could include structural differences in their speech versus Hupa as spoken by fluent first-language speakers as well as sociolinguistic analysis of the domains of use for Hupa in contemporary society and the role of language revitalization and reclamation efforts that have been underway in Hoopa Valley for over three decades. This would be relevant to the theoretical study of Second Language Acquisition since the typological distance between English and Hupa is much greater than the distance between English and other Indo-European languages that are commonly (if not exclusively) studied in the field. Moreover, it would have the very practical application of highlighting areas where pedagogical intervention in Hupa and other Athabaskan languages might be most needed or likely to be successful.

## 3. Final Remarks

Overall, this dissertation can be considered a preliminary step in a long-term project of discovering parameters of variation and change that are relevant to understanding the diachronic development of Pacific Coast Athabaskan based on exhaustive analysis of the entire documentary record of these languages. The research is therefore closely related to current discussion in the field of linguistics surrounding the creation of linguistic documentation and ensuring its long-term accessibility to a broad audience of potential users. One current within this overall movement emphasizes the task of assembling robust, well-annotated corpora of primary recordings that future scholars can access to explore a wide variety of research questions (Himmelmann 2006). Such corpora, however, are only useful insofar as they are in fact rich enough to make such exploration possible. If the present study can produce even limited results of general interest to the field of linguistics, even while relying on relatively problematic early archival material that was haphazardly collected by researchers with no long-term community of users in mind, then it is *a fortiori* a justification for a more measured approach to the collection, annotation, and preservation of language data. The window for doing so, at least among speakers with no break in transgenerational transmission, has nearly closed for Pacific Coast Athabaskan. However, ongoing collaborative documentation efforts with remaining speakers offer hope that an adequate record of the language will be created that can assist new generations of speakers in their efforts to develop fluency.

## References


IJAL = International Journal of American Linguistics
UCPAAE = University of California Publications in American Archaeology and Ethnology

Abbott, G.H. 1858. *Letter to George Gibbs, from Siletz Agency, O.T., August 31, 1858, with Map and Census of Coquille and Adjoining Tribes*. National Anthropological Archives: NAA MS 125.

Ahlers, Jocelyn. 1999. *Proposal for the Use of Cognitive Linguistics in Hupa Language Revitalization*. Ph.D. dissertation, University of California, Berkeley.

Atkinson, Quentin and Russell D. Gray. 2006. How Old is the Indo-European Language Family? Illumination or More Moths to the Flame? In Forster and Renfrew (2006), 91-110.

Azpell, Thomas F. 1870. *Comparative Vocabulary of Hoopa (Noh-tin-oah) and Klamath (Sa-ag-its) August 14, 1870*. National Anthropological Archives: NAA MS 83.

Babel, Molly. 2009a. *Phonetic and Social Selectivity in Speech Accommodation*. Ph.D. dissertation, University of California, Berkeley.

Babel, Molly. 2009b. The Phonetics and Phonology of Obsolescence. In *Variation in Indigenous Minority Languages*, edited by D. Preston and M. Stanford, 23-45. Amsterdam: Benjamins.

Babel, Molly. 2012. Evidence for Phonetic and Social Selectivity in Spontaneous Phonetic Imitation. *Journal of Phonetics* 40: 177-189.

Babel, Molly, Andrew Garrett, Michael J. Houser, and Maziar Toosarvandani. 2013. Descent and Diffusion in Language Diversification: A Study of Western Numic Dialectology. *IJAL* 79: 445–489.

Barnett, Homer G. 1934. *Indian Tribes of the Oregon Coast*, books 1 & 2. National Anthropological Archives: Homer Garner Barnett Papers 1934-1973. Series 3, container 2.

Barnett, Homer G. 1937. Culture Element Distributions: VII. Oregon Coast. *Anthropological Records* 1(3): 155-204.

Barnhardt, W.H. 1859. *Comparative Vocabulary of the Languages Spoken by the Umpqua, Lower Rogue River [Takelma], and Calapooia Tribes of Indians May, 1859*. National Anthropological Archives: NAA MS 218.

Baumhoff, Martin A. 1958. California Athabascan Groups. *University of California Anthropological Records* 16(5): 157-237.

Bell, Alan. 1984. Language Style as Audience Design. *Language in Society* 13: 145-204.

Bergsland, Knut and Hans Vogt. 1962. On the Validity of Glottochronology. *Current Anthropology* 3(2): 115-153.

Birner, Betty J. and Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English.* Philadelphia: Benjamins.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Bommelyn, Loren, ed. 1989. *Xus we-yo: Tolowa (Tututni) Language Dictionary.* Crescent City, CA: Tolowa Language Committee.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 2012. Mapping the Origins and Expansion of the Indo-European Language Family. *Science* 337: 957-960.

Bourhis, Richard Y. and Howard Giles. 1977. The Language of Intergroup Distinctiveness. In *Language, Ethnicity, and Intergroup Relations*, edited by Howard Giles, 119-136. London: Academic.

Bowern, Claire and Quentin Atkinson. 2012. Computational Phylogenetics and the Internal Structure of Pama-Nyungan. *Language* 88(4): 817-845.

Bright, Jane O. 1964. The Phonology of Smith River Athapaskan (Tolowa). *IJAL* 30(2): 101-107.

Bright, William. 1950-1952. [Field notes on languages of northwestern California.] Survey of California and Other Indian Languages, University of California, Berkeley, BrightW.003.001.

Bright, William. 1952. Linguistic Innovations in Karok. *IJAL* 18(2): 53-62.

Bright, William. 1960. Animals of Acculturation in the California Indian Languages. *University of California Publications in Linguistics* 4(4): 215-246. Berkeley: University of California Press.

Brown, Cecil H. 1994. Lexical Acculturation in Native American Languages. *Current Anthropology* 35(2): 95-108.

Brown, Cecil H. 1999. *Lexical Acculturation in Native American Languages*. New York: Oxford University Press.

Bryant, David. 2006. Radiation and Network Breaking in Polynesian Linguistics. In Forster and Renfrew (2006), 111-118.

Bureau of Indian Affairs. 1967. Indian Census Rolls, 1885-1940. National Archives and Records Administration microfilm M595.

Bureau of Indian Affairs. 1979. *Siletz Reservation Plan*. [Online resource: http://soda.sou.edu]

Callaghan, Catherine A. and Geoffrey Gamble. 1996. Borrowing. In *Handbook of North American Indians*, vol. 17, *Languages*, edited by Ives Goddard, 111-116. Washington, DC: Smithsonian Institution.

Campbell, Lyle, and Martha C. Muntzel. 1989. The Structural Consequences of Language Death. In *Investigating Obsolescence*, edited by N. Dorian, 181–196. Cambridge University Press.

Chang, Charles B. 2012. Phonetics vs. Phonology in Loanword Adaptation: Revisiting the Role of the Bilingual. In *Proceedings of the Thirty-Fourth Annual Meeting of the Berkeley Linguistics Society*, edited by S. Berson, A. Bratkievich, D. Bruhn, A. Campbell, R. Escamilla, A. Giovine, L. Newbold, M. Perez, M. Piqueras-Brunet, and R. Rhomieux, 61-72. Berkeley, CA: Berkeley Linguistics Society.

Cho, Taehong and Peter Ladefoged. 1999. Variation and Universals in VOT: Evidence from 18 Languages. *Journal of Phonetics* 27: 207-229.

Collins, James. 1985. Pronouns, Markedness, and Stem Change in Tolowa. *IJAL* 51: 358-372.

Collins, James. 1989. Nasalization, Lengthening, and Phonological Rhyme in Tolowa. *IJAL* 55(3): 326-340.

Computational Phylogenetics in Historical Linguistics [CPHL] Project. 2007a. Phonological Characters. [Online resource available on the Computational Phylogenetics in Historical Linguistics project website: http://www.cs.rice.edu/~nakhleh/CPHL/code-p-07.pdf]

Computational Phylogenetics in Historical Linguistics [CPHL] Project. 2007b. Morphological Characters. [Online resource available on the Computational Phylogenetics in Historical Linguistics project website: http://www.cs.rice.edu/~nakhleh/CPHL/code-m-07.pdf]

Conathan, Lisa. 2004. The Linguistic Ecology of Northwestern California: Contact, Functional Convergence, and Dialectology. Ph.D. dissertation, University of California, Berkeley.
Conathan, Lisa. 2006. Recovering Sociolinguistic Context from Early Sources: The Case of Northwestern California. *Anthropological Linguistics* 48(3): 209-232.
Cook, Eung-Do. 1980. On "Variability of Chipewyan Consonants" by Scollon. *IJAL* 46(4): 313-315.
Cook, Eung-Do. 1984. *A Sarcee Grammar*. Vancouver: University of British Columbia Press.
Cook, Eung-Do. 1989. Is Phonology Going Haywire in Dying Languages? Phonological Variations in Chipewyan and Sarcee. *Language in Society* 18(2): 235-255.
Cook, Eung-Do. 1996. Third-Person Plural Subject Prefix in Northern Athapaskan. *IJAL* 62(1): 86-110.
Cook, Eung-Do. 2004. *A Grammar of Dëne Sųłiné (Chipewyan).* Winnipeg: Algonquian and Iroquoian Linguistics.
Cook, Sherburne F. 1956. The Aboriginal Population of the North Coast of California. *University of California Anthropological Records* 16(3): 81-130.
Crook, George. 1852-1861. *Klamath River Vocabularies; Aliquah [Yurok], Arra-Arra [Karok], Hopah [Hupa] ca. 1852-1861*. National Anthropological Archives: NAA MS 209.
Curtin, Jeremiah. 1888-1889. *Hupa Vocabulary December 1888-January 1889*. National Anthropological Archives: NAA MS 2063.
de Jong, Kenneth and Mi-Hui Cho. 2012. Loanword Phonology and Perceptual Mapping: Comparing Two Corpora of Korean Contact with English. *Language* 88(2): 341-368.
Delvaux, Véronique and Alain Soquet. 2007. The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. *Phonetica* 64: 145-173.
Dorsey, J. Owen. 1884a. *Dakube tede Vocabulary from the Athapascans Formerly Living on Applegate Creek, Oregon October 1, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.1) [372].
Dorsey, J. Owen. 1884b. *Galice Creek (Talt uct un tude) Vocabulary and Grammatical Notes, Formerly Spoken by the Athapascans in Josephine County, Oregon (30 miles north of Kerby) September 18-October 9, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.2) [373].
Dorsey, J. Owen. 1884c. *Chasta Costa (Cistakawusta) Vocabulary from the Athapascans Formerly Living on Rogue River, Oregon September-October 1884*. National Anthropological Archives: NAA MS 4800 (4.1.3) [374].
Dorsey, J. Owen. 1884d. *Upper Coquille (Miciqwutme tunne) and Grammatical Notes August, October 1884*. National Anthropological Archives: NAA MS 4800 (4.1.4) [375].
Dorsey, J. Owen. 1884e. *Chetco (Tceti tunne) Vocabulary and Grammatical Notes, Formerly Spoken by the Athapascans Living on the Chetco River, Oregon September 20, 25; October 4-21, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.5) [376].
Dorsey, J. Owen. 1884f. *Mikwunu tunne Vocabulary and Grammatical Notes, Spoken by the Tututni Athapascans Formerly Living on Lower Rogue River, Oregon October 15, 1884.* National Anthropological Archives: NAA MS 4800 (4.1.6) [377].
Dorsey, J. Owen. 1884g. *Tutu tunne or Tutu and Joshua Vocabulary, as Spoken in Several Athapascan Villages, with Grammatical and Miscellaneous Notes August 20-October 28, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.7) [378].
Dorsey, J. Owen. 1884h. *Tceme tunni or Joshua Vocabulary and Grammatical Notes August 1884*. National Anthropological Archives: NAA MS 4800 (4.1.7) [379].

Dorsey, J. Owen. 1884i. *Tutu Vocabulary and Grammatical Notes August 20-?, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.7) [380].
Dorsey, J. Owen. 1884j. *Tutu Vocabulary: Original Notes and Vocabulary on Slips 1884*. National Anthropological Archives: NAA MS 4800 (4.1.7) [381].
Dorsey, J. Owen. 1884k. *Yukitce Vocabulary and Grammatical Notes, Formerly Spoken by the Athapascans Living on Euchre Creek, Oregon September 2, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.8) [386].
Dorsey, J. Owen. 1884l. *Kwatami Vocabulary and Grammatical Notes, Spoken by the Athapascans Formerly living on Sixes Creek, Oregon August 29, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.9) [387].
Dorsey, J. Owen. 1884m. *Naltûnnétûnně Vocabulary and Grammatical Notes, as Spoken by the Athapascans Formerly Living on Naltunne River, Oregon October 1884*. National Anthropological Archives: NAA MS 4800 (4.1.10) [388].
Dorsey, J. Owen. 1884n. *Qaunwate Vocabulary from the Athapascans Formerly Living on Smith River, California September 20, 1884*. National Anthropological Archives: NAA MS 4800 (4.1.11) [389].
Dorsey, J. Owen. 1889. Indians of Siletz Reservation, Oregon. *American Anthropologist* 2(1): 55-62.
Dorsey, J. Owen. 1890. The Gentile System of the Siletz Tribes. *Journal of American Folklore* 3(10): 227-237.
Dorsey, J. Owen. 1891. *Omaha and Ponka Letters.* Washington: Government Printing Office.
Drucker, Philip. 1934. [Field notes from the Siletz Reservation.] National Anthropological Archives: Philip Drucker Papers 1933-ca. 1954, NAA MS 4516.
Drucker, Philip. 1937. The Tolowa and Their Southwest Oregon Kin. *UCPAAE* 36(4): 221-300. Berkeley: University of California Press.
Du Bois, Cora. 1934. [Tututni (Rogue River Athapaskan) field notes.] Typescript. Bancroft Library CU-23.1, no. 6.
Dunn, Michael. 2013. Computational Phylogenetic Approaches to Language Change. Presentation at the 87th Annual Meeting of the Linguistic Society of America, Boston, MA.
Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesnik, and Angela Terrill. 2008. Structural Phylogeny in Historical Linguistics: Methodological Explorations Applied in Island Melanesia. *Language* 84(4): 710-759.
Dyen, Isidore and David F. Aberle. 1974. *Lexical Reconstruction: The Case of the Proto-Athapaskan Kinship System*. New York: Cambridge University Press.
Eckert, Penelope. 2008. Variation and the Indexical Field. *Journal of Sociolinguistics* 12(4): 453-476.
Escamilla, Ramon. 2012. An Updated Typology of Causative Constructions: Form-Function Mappings in Hupa (California Athabaskan), Chungli Ao (Tibeto-Burman), and Beyond. Ph.D. dissertation, UC Berkeley.
Evans, Nicholas and Patrick McConvell. 1998. The Enigma of Pama-Nyungan Expansion in Australia. In *Archaeology and Language II*, edited by Roger Blench and Matthew Spriggs, 174-192. London: Routledge.
Evans, Steven N., Donald Ringe, and Tandy Warnow. 2006. Inference of Divergence Times as a Statistical Inverse Problem. In Forster and Renfrew (2006), 119-130.
Everette, Willis E. 1882. *Vocabulary of "Tu-tu-tene" and Nine Confederated Tribes, Siletz River, Western Oregon, Dec. 12, 1882*. National Anthropological Archives: NAA MS 78.

Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. Centers and Peripheries: Network Roles in Language Change. *Lingua* 120: 2061-2079.

Felsenstein, Joseph. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.

Fishman, Joshua. 1967. Bilingualism with and without Diglossia; Diglossia with and without Bilingualism. *Journal of Social Issues* 23(2): 29-38.

Furbee, N. Louanna. 1995. Accommodation in Language Death. Paper presented at the Symposium on Language Loss and Public Policy, University of New Mexico Linguistic Institute.

Gatschet, Albert S. 1877. *Umpqua (Athapascan) Vocabulary, Recorded at Grande Ronde Indian Reservation, Polk Co., Oregon, December 30, 1877*. National Anthropological Archives: NAA MS 76.

Gibbs, George. 1851-1852. *Hopa Vocabulary 1851-1852*. National Anthropological Archives: NAA MS 128.

Gibbs, George. 1853a. Journal of the Expedition of Colonel Reddick M'Kee, United States Indian Agent, through North-Western California. Performed in the Summer and Fall of 1851. *Historical and Statistical Information Respecting the History, Condition and Prospects of the Indian Tribes of the United States*, edited by Henry R. Schoolcraft. Vol. 3, 99-177. Philadelphia: Lippencott.

Gibbs, George. 1853b. Vocabularies of Indian Languages of Northwest California. In *Historical and Statistical Information Respecting the History, Condition and Prospects of the Indian Tribes of the United States*, edited by Henry R. Schoolcraft. Vol. 3, 428-445. Philadelphia: Lippencott.

Givón, Talmy. 2000. Internal Recontstruction: As Method, as Theory. In *Reconstructing Grammar: Comparative Linguistics and Grammaticalization Theory*, edited by Spike Gildea, 107-160. Philadelphia: John Benjamins.

Givon, Talmy and Loren Bommelyn. 2000. The Evolution of De-Transitive Voice in Tolowa Athabaskan. *Studies in Language* 24(1): 41-76.

Goddard, Ives. 1975. Algonquian, Wiyot, and Yurok: Proving a Distant Genetic Relationship. In *Linguistics and Anthropology: In Honor of C.F. Voegelin*, edited by M.D. Kinkade, K.L. Hale, and O. Werner, 249-262. Lisse: Peter de Ridder.

Goddard, Ives. 2010. Linguistic Variation in a Small Speech Community: The Personal Dialects of Moraviantown Delaware. *Anthropological Linguistics* 52(1): 1-48.

Goddard, Pliny Earle. n.d. Hupa Text (South Fork of the Trinity). American Philosophical Society MS 30 (Na20a.1).

Goddard, Pliny Earle. 1901. Conscious Word-making by the Hupa. *American Anthropologist*, n.s., 3(1): 208-209.

Goddard, Pliny Earle. 1902-1903. Tolowa Field Notes. American Philosophical Society MS 30 (Na20f.1).

Goddard, Pliny Earle. 1903. Life and Culture of the Hupa. *UCPAAE* 1(1): 1-87.

Goddard, Pliny Earle. 1904. Hupa Texts. *UCPAAE* 1(2): 89-368.

Goddard, Pliny Earle. 1905. Morphology of the Hupa Language. *UCPAAE* 3.

Goddard, Pliny Earle. 1907. Phonology of the Hupa Language, Pt. 1: The Individual Sounds. *UCPAAE* 5(1): 1-19.

Goddard, Pliny Earle. 1908. A Confession as to Errors in Hupa Linguistics. *American Anthropologist*, n.s., 10(1): 170-171.

Goddard, Pliny Earle. 1909. Kato Texts. *UCPAAE* 5(3): 65-238.

Goddard, Pliny Earle. 1911. Athapascan (Hupa). In *Handbook of American Indian Languages, Part 1*, edited by Franz Boas, 85-159. Washington, D.C.: Bureau of American Ethnography B-40(1).
Goddard, Pliny Earle. 1912. Elements of the Kato Language. *UCPAAE* 11(1): 1-176.
Goddard, Pliny Earle. 1914a. Notes on the Chilula Indians of Northwestern California. *UCPAAE* 10(6): 265-288.
Goddard, Pliny Earle. 1914b. Chilula Texts. *UCPAAE* 10(7): 289-379.
Goddard, Pliny Earle. 1916. Beaver Texts. *Anthropological Papers of the American Museum of Natural History* 10(5): 295-397.
Goddard, Pliny Earle. 1917. Beaver Dialect. *Anthropological Papers of the American Museum of Natural History* 10(6): 399-517.
Goddard, Pliny Earle. 1923. Wailaki Texts. *IJAL* 2(3/4): 77-135.
Goddard, Pliny Earle. 1928. Pitch Accent in Hupa. *UCPAAE* 23: 333-8.
Golla, Victor. 1963. [Hupa field notes.] Included on microfilm reel Lamb.m001, Survey of California and Other Indian Languages, University of California Berkeley. Associated audio recordings archived at the Berkeley Language Center, LA 119.
Golla, Victor. 1964. Etymological Study of Hupa Noun Stems. *IJAL* 30(2): 108-117.
Golla, Victor. 1970. *Hupa Grammar*. Ph.D. dissertation, University of California, Berkeley.
Golla, Victor. 1976. Tututni (Oregon Athapaskan). *IJAL* 42(3): 217-227.
Golla, Victor. 1977. A Note on Hupa Verb Stems. *IJAL* 43(4): 355-8.
Golla, Victor. 1984. *Hupa Stories, Anecdotes, and Conversations*. Hoopa, CA: Hoopa Valley Tribe.
Golla, Victor. 1985. *A Short Practical Grammar of Hupa*. Hoopa, CA: Hoopa Valley Tribe.
Golla, Victor. 1996a. Sketch of Hupa, an Athapaskan Language. In *Handbook of North American Indians*, vol. 17, *Languages*, edited by Ives Goddard, 364-389. Washington: Smithsonian Institution.
Golla, Victor. 1996b. *Hupa Language Dictionary*. 2nd. ed. Hoopa, CA: Hoopa Valley Tribal Council. [Available online: http://hdl.handle.net/2148/48]
Golla, Victor. 2000. Language History and Communicative Strategies in Aboriginal California and Oregon. In *Languages of the North Pacific Rim*, vol. 5, edited by Osahito Miyaoka, 43-64. Suita, Japan: Faculty of Informatics, Osaka Gakuin University.
Golla, Victor. 2008. *Lower Rogue River (Tututni) Lexicon*. Manuscript, Humboldt State University.
Golla, Victor. 2011. *California Indian Languages*. Berkeley: University of California Press.
Gordon, Matthew. 1996. The Phonetic Structures of Hupa. *UCLA Working Papers in Phonetics* 93: 164-187.
Gordon, Matthew and Edmundo Luna. 2005. An Intergenerational Investigation of Hupa Stress. In *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, edited by M. Ettlinger, N. Fleischer, and M. Park-Doob, 105-117. Berkeley, CA: Berkeley Linguistics Society.
Gould, Richard A. 1966. The Wealth Quest Among the Indians of Northwest California. *Proceedings of the American Philosophical Society* 110(1): 67-89.
Gray, Dennis J. 1987. The Takelma and their Athapascan Neighbors: A New Ethnographic Synthesis for the Upper Rogue River Area of Southwestern Oregon. *University of Oregon Anthropological Papers* 37. Eugene: University of Oregon Department of Anthropology and Museum of Natural and Cultural History.

Gray, Russell D. and Quentin Atkinson. 2003. Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origins. *Nature* 426: 435-439.

Greenhill, Simon J., Thomas E. Currie, and Russell D. Gray. 2009. Does Horizontal Transmission Invalidate Cultural Phylogenies? *Proceedings of the Royal Society B: Biological Sciences* 276: 2299-2306.

Hall, Roberta L. 1992. Language and Cultural Affiliations of Natives Residing Near the Mouth of the Coquille River before 1851. *Journal of Anthropological Research* 48(2): 165-184.

Hale, Horatio. 1846. Ethnography and Philology. *United States Exploring Expedition During the Years 1838, 1839, 1840, 1841, 1842 under the Command of Charles Wilkes, U.S.N.*, vol. 6. Philadelphia: Lea and Blanchard.

Hargus, Sharon. 2010. Athabaskan Phonetics and Phonology. *Language and Linguistics Compass* 4/10: 1019-1040.

Hargus, Sharon and Siri Tuttle. 1997. Augmentation as Affixation in Athabaskan Languages. *Phonology* 14: 177-220.

Harrington, John Peabody. 1943. Pacific Coast Athapascan Discovered to be Chilcotin. *Journal of the Washington Academy of Sciences* 33: 203-213.

Harrington, John Peabody. 1981. [Oregon Athabaskan field notes.] *The Papers of John Peabody Harrington in the Smithsonian Institution, 1907-1957*. Microfilm edition, part 1: Native American History, Language, and Culture of Alaska/Northwest Coast. "Kwalhioqua-Tlatskanai", reel 19, and Oregon Athabaskan, reels 25-28. National Anthropological Archives.

Haugen, Einar. 1950. The Analysis of Linguistic Borrowing. *Language* 26(2): 210-231.

Hill, Jane H. 1994. [Commentary on Brown (1994)] *Current Anthropology* 35(2): 109-110.

Hill, Jane H. 2001. Languages on the Land: Toward an Anthropological Dialectology. In *Archaeology, Language and History: Essays on Culture and Identity*, edited by John E. Terrell, 257-282. Westport: Bergin and Garvey.

Hill, Jane H. To appear. Uto-Aztecan Hunter-Gatherers: Language Diversity in the Takic Spread and the Numic Spread Compared. In *Hunter-Gatherers and Linguistic History: A Global Perspective*, edited by Tom Güldemann, Patrick McConvell, and Richard Rhodes. Cambridge University Press.

Himmelmann, Nikolaus P. 2006. Language Documentation: What Is It, and What Is It Good for? In *Essentials of Language Documentation*, edited by J. Gippert, N.P. Himmelmann, and U. Mosel, 1- 30. New York: Mouton de Gruyter.

Hinton, Leanne. 1979. Irataba's Gift: A Closer Look at the ṣ > s > θ Sound Shift in Mojave and Northern Pai. *Journal of California and Great Basin Anthropology – Papers in Linguistics* 1: 3-37.

Hinton, Leanne. 1981. Upland Yuman Sibilant Shifts: The Beginning of the Story. *Journal of California and Great Basin Anthropology – Papers in Linguistics* 3: 65-76.

Hoijer, Harry. 1938. The Southern Athapaskan Languages. *American Anthropologist*, n.s., 40(1): 75-87.

Hoijer, Harry. 1945. The Apachean Verb, Part I: Verb Structure and Pronominal Prefixes. *IJAL* 11: 193-203.

Hoijer, Harry. 1956. The Chronology of the Athapaskan Languages. *IJAL* 22(4): 219-232.

Hoijer, Harry. 1960. Athapaskan Languages of the Pacific Coast. In *Culture in History: Essays in Honor of Paul Radin*, edited by S. Diamond, 960-77. New York: Columbia University Press.

Hoijer, Harry. 1962. Linguistic Subgroupings by Glottochronology and the Comparative Method: The Athapaskan Languages. *Lingua* 11: 192-198.
Hoijer, Harry. 1963. The Athapaskan Languages. In Hoijer and Others (1963), 1-29.
Hoijer, Harry. 1966. Galice Athapaskan: A Grammatical Sketch. *IJAL* 32(4): 320-27.
Hoijer, Harry. 1971. Athapaskan Morphology. In *Studies in American Indian Languages*, edited by Jesse Sawyer, 113-147. (University of California Publications in Linguistics 65.) Berkeley: University of California Press.
Hoijer, Harry. 1973. Galice Noun and Verb Stems. *Linguistics* 104: 49-73.
Hoijer, Harry and Janet Joël. 1963. Sarsi Nouns. In Hoijer and Others (1963), 62-75.
Hoijer, Harry and Others, eds. 1963. *Studies in the Athapaskan Languages*. (University of California Publications in Linguistics 29.) Berkeley: University of California Press.
Holden, Clare and Russell D. Gray. 2006. Rapid Radiation, Borrowing and Dialect Continua in the Bantu Languages. In Forster and Renfrew (2006), 19-32.
Howren, Robert. 1971. A Formalization of the Athabaskan 'D-Effect'. *IJAL* 37(2): 96-113.
Huelsenbeck, John, Fredrick Ronquist, and Maxim Teslenko. 2011. Command Reference for MrBayes ver. 3.2. [Online resource: http://mrbayes.sourceforge.net/commref_mb3.2.pdf]
Huson, D.H. and D. Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23(2): 254-267.
Hymes, Dell. 1957. A Note on Athapaskan Glottochronology. *IJAL* 23(4): 291-297.
Jacobs, Elizabeth. 1968. A Chetco Athabaskan Myth Text from Southwestern Oregon. *IJAL* 34(3): 192-193.
Jacobs, Elizabeth. 1977. A Chetco Athabaskan Text and Translation. *IJAL* 43(4): 269-273.
Jacobs, Melville. 1937. Historic Perspectives in Indian Languages of Oregon and Washington. *Pacific Northwest Quarterly* 28(1): 55-74.
Jacobs, Melville. 1968. An Historical Event Text from a Galice Athabaskan in Southwestern Oregon. *IJAL* 34(3): 183-191.
Jacobs, Melville and Elizabeth Jacobs. 1930s. [Oregon Athabaskan field notes.] University of Washington Libraries Special Collections: Melville Jacobs Papers 1918-1974.
Jelinek, Eloise, Sally Midgette, Keren Rice, and Leslie Saxon, eds. 1996. *Athabaskan Language Studies: Essays in Honor of Robert W. Young*. Albuquerque: University of New Mexico Press.
Jetté, Jules and Eliza Jones. 2000. *Koyukon Athabaskan Dictionary*. Fairbanks, AK: Alaska Native Language Center.
Jones, Eliza. [1978]. *Junior Dictionary for Central Koyukon Athabaskan.* National Bilingual Materials Development Center. [Accessible online at the Alaska Native Language Archive: http://www.uaf.edu/anla/collections/dictionaries]
Kalyan, Siva and Alexandre François. To appear. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry In *Let's Talk About Trees: Tackling Problems in Representing Phylogenetic Relationships Among Languages*, edited by R. Kikusawa and L. Reid. Osaka: National Museum of Ethnology.
Kari, James. 1977. Linguistic Diffusion between Tanaina and Ahtna. *IJAL* 43(4): 274-288.
Kari, James. 1989. Affix Positions and Zones in the Athapaskan Verb Complex: Ahtna and Navajo. *IJAL* 55(4): 424-454.
Kari, James. 1990. *Ahtna Athabaskan Dictionary*. Fairbanks, AK: Alaska Native Language Center.

Kari, James. 1993. Diversity in Morpheme Order in Several Athabaskan Languages. In *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society: Special Session on Syntactic Issues in Native American Languages*, edited by David Peterson, 50-56. Berkeley, CA: Berkeley Linguistics Society.

Kari, James. 1994. *Lower Tanana Athabaskan Dictionary*. [Accessible online at the Alaska Native Language Archive: http://www.uaf.edu/anla/collections/dictionaries]

Kari, James. 2007. *Dena'ina Topical Dictionary*. Fairbanks, AK: Alaska Native Language Center.

Kautz, August Valentine. [1855]. *English-"Tou-tout-en" Vocabulary of About 70 Words*. National Anthropological Archives: NAA MS 198.

Kent, William Eugene. 1973. *The Siletz Indian Reservation 1855-1900*. Newport, Oregon: Lincoln County Historical Society.

Kerswill, Paul. 1994. *Dialects Converging: Rural Speech in Urban Norway*. Oxford: Clarendon Press.

Kerswill, Paul. 2002. Koineization and Accommodation. In *The Handbook of Language Variation and Change*, edited by J.K. Chambers, P. Trudgill, and N. Schilling-Estes, 669-702. Oxford: Blackwell.

Kerswill, Paul. 2006. Review of Trudgill (2004). *Language* 83(3): 657-661.

Kerswill, Paul, and Peter Trudgill. 2005. The Birth of New Dialects. In *Dialect Change: Convergence and Divergence in European Languages*, edited by P. Auer, F. Hinskins, and P. Kerswill, 196-220. Cambridge University Press.

Kerswill, Paul, and Ann Williams. 2000. Creating a New Town Koine: Children and Language Change in Milton Keynes. *Language in Society* 29: 65-115.

Kerswill, Paul, and Ann Williams. 2005. New Towns and Koineization: Linguistic and Social Correlates. *Linguistics* 43(5): 1023-1048.

Kingston, John. 2005. The Phonetics of Athabaskan Tonogenesis. In *Athabaskan Prosody*, edited by S. Hargus and K. Rice, 137-184. Philadelphia: John Benjamins.

Krauss, Michael. 1963. [Notes on Athabascan Glottochronology.] Alaska Native Language Archive manuscript CA961K1963c.

Krauss, Michael. 1965. Proto-Athabaskan-Eyak and the Problem of Na-Dene II: Morphology. *IJAL* 31(1): 18-28.

Krauss, Michael. 1973. Na-Dene. In *Linguistics in North America*, edited by Thomas A. Sebeok, 903-978. (Current Trends in Linguistics 10). The Hague: Mouton.

Krauss, Michael. 2005. Athabaskan Tone. In *Athabaskan Prosody*, edited by S. Hargus and K. Rice, 55-136. Philadelphia: John Benjamins.

Krauss, Michael, and Victor Golla. 1981. Northern Athapaskan Languages. In *Handbook of North American Indians*, vol. 6, *Subarctic*, edited by J. Helm, 67-85. Washington, D.C.: Smithsonian Institution.

Krauss, Michael, and Jeff Leer. 1981. *Athabaskan, Eyak, and Tlingit Sonorants*. (Alaska Native Language Center Research Papers 5.) Fairbanks, Alaska: ANLC.

Kroeber, Alfred L. 1920. California Culture Provinces. *UCPAAE* 17(2): 151-169.

Kroeber, Alfred L. 1925. *Handbook of the Indians of California*. Republished 1976. New York: Dover Publications.

Kroeber, Alfred L. 1967. Goddard's California Athabascan Texts. *IJAL* 33(4): 269-275.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1994. *Principles of Linguistic Change*, vol. 1, *Internal Factors*. Cambridge, MA: Blackwell.
Labov, William. 2001. *Principles of Linguistic Change*, vol. 2, *Social Factors*. Cambridge, MA: Blackwell.
Labov, William. 2007. Transmission and Diffusion. *Language* 83(2): 344-387.
Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language* 89(1): 30-65.
Lacharité, Darlene and Carole Paradis. 2005. Category Preservation and Proximity versus Phonetic Approximation in Loanword Adaptation. *Linguistic Inquiry* 36(2): 223-258.
Landar, Herbert. 1977. Three Rogue River Athapaskan Vocabularies. *IJAL* 43(4): 289-301.
Leavelle, Tracy Neal. 1998. "We Will Make It Our Own Place": Agriculture and Adaptation at the Grand Ronde. *American Indian Quarterly* 22(4): 433-456.
Leer, Jeff. 1979. *Proto-Athabaskan Verb Stem Variation*. Fairbanks, AK: Alaska Native Language Center.
Leer, Jeff. 1987. Navajo and Comparative Athabaskan. In *The Navajo Language. A Grammar and Colloquial Dictionary*, edited by Robert W. Young and William Morgan, 264-266. Albuquerque: University of New Mexico Press.
Leer, Jeff. 1996. The Historical Evolution of the Stem Syllable in Gwich'in (Kutchin/Loucheux) Athabaskan. In Jelinek et al. (1996), 193-234.
Leer, Jeff. 2000. The Negative/Irrealis Category in Athabaskan-Eyak-Tlingit. In *The Athabaskan Languages: Perspectives on a Native American Language Family*, edited by Theodore B. Fernald and Paul R. Platero, 101-138. Oxford University Press.
Leer, Jeff. 2005. How Stress Shapes the Stem-suffix Complex in Athabaskan. In *Athabaskan Prosody*, edited by S. Hargus and K. Rice, 277-318. Philadelphia: Benjamins.
Leer, Jeff. 2011. [Comparative Athabaskan lexicon.] Manuscript, University of Alaska, Fairbanks.
Lees, Robert B. 1953. The Basis for Glottochronology. *Language* 29: 113-127.
Le Page, Robert Brock and Andrée Tabouret-Keller. 1985. *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. Cambridge University Press.
Li, Fang-Kuei. n.d. [Wailaki noun and verb stem list.] Photocopy of original manuscript, in the possession of Victor Golla, Humboldt State University.
Li, Fang-Kuei. 1930a. *Mattole: An Athabaskan Language*. Chicago: University of Chicago Press.
Li, Fang-Kuei. 1930b. A Study of Sarcee Verb-Stems. *IJAL* 6(1): 3-27.
Li, Fang-Kuei. 1933. A List of Chipewyan Stems. *IJAL* 7(3/4): 122-151.
Li, Fang-Kuei. 1946. Chipewyan. In *Linguistic Structures of Native America*, edited by Harry Hoijer et al., 398-423. New York: The Viking Fund.
Lisker, Leigh and Arthur S. Abramson. 1967. Some Effects of Context on Voice Onset Time in English Stops. *Language and Speech* 10: 1-28.
Marten, Lutz. 2006. Bantu Classification, Bantu Trees, and Phylogenetic Methods. In Forster and Renfrew (2006), 43-56
Matras, Yaron. 2009. *Language Contact*. Cambridge University Press.
Merriam, C. Hart. 1850-1974. C. Hart Merriam Papers, BANC FILM 1022, Bancroft Library, University of California, Berkeley.

Merriam, C. Hart. 1979. *Indian Names for Plants and Animals among Californian and Other Western North American Tribes*. Socorro, N.M.: Ballena Press.
Milhau, John J. 1856. *Vocabulary of Umpqua Valley (Proper) 1856*. National Anthropological Archives: NAA MS 193-a.
Miller, Jay, and William R. Seaburg. 1990. Athapaskans of Southwest Oregon. In *Handbook of North American Indians*, vol. 7, *Northwest Coast*, edited by W. Suttles, 580-588. Washington, D.C.: Smithsonian Institution.
Minto Nenana Athabaskan Noun Dictionary. 1974. Fairbanks, AK: Alaska Native Language Center. [Accessible online at the Alaska Native Language Archive: http://www.uaf.edu/anla/collections/dictionaries]
Morice, A.G. 1906. The Great Déné Race. Chapter III. The Southern Dénés. *Anthropos* Bd. 1, H. 3, 483-508.
Morice, A.G. 1932. *The Carrier Language*. Mödling bei Wien, Austria: Verlag der Internationalen Zeitschrift "Anthropos."
Nakhleh, Luay, Tandy Warnow, Don Ringe, and Steven N. Evans. 2005. A Comparison of Phylogenetic Reconstruction Methods on an Indo-European Dataset. *Transactions of the Philological Society* 103(2): 171-192.
Nakhleh, Luay, Don Ringe, and Tandy Warnow. 2008. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81(2): 382-420.
Nelson, Byron. 1978. *Our Home Forever: A Hupa Tribal History*. Hoopa, CA: Hoopa Valley Tribe.
Newbold, Lindsey. 2010. Word Order in Hupa. In *Working Papers in Athabaskan Languages 2009*, edited by Siri Tuttle and Justin Spence, 89-106. (Alaska Native Language Center Working Papers 8). Fairbanks, AK: Alaska Native Language Center.
Nichols, Johanna and Tandy Warnow. 2008. Tutorial on Computational Linguistic Phylogeny. *Language and Linguistics Compass* 2/5: 760-820.
Office of Indian Affairs. 1855. *Annual report of the Commissioner of Indian Affairs, Transmitted With the Message of the President at the Opening Session of the Thirty-Third Congress, 1854*. Washington, D.C.: A.O.P. Nicholson.
Office of Indian Affairs. 1858. *Report of the Commissioner of Indian Affairs Accompanying the Annual Report of the Secretary of the Interior for the Year 1857*. Washington, D.C.: William A. Harris.
Office of Indian Affairs. 1861. *Report of the Commissioner of Indian Affairs Accompanying the Annual Report of the Secretary of the Interior for the Year 1861*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1865. *Report of the Commissioner of Indian Affairs for the Year 1865*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1866. *Report of the Commissioner of Indian Affairs for the Year 1866*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1870. *Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior for the Year 1870*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1878. *Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior for the Year 1878*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1882. *Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior for the Year 1882*. Washington, D.C.: Government Printing Office.

Office of Indian Affairs. 1886. *Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior for the Year 1886*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1887. *Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior for the Year 1887*. Washington, D.C.: Government Printing Office.
Office of Indian Affairs. 1889. *Fifty-eighth Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior*. Washington, DC: Government Printing Office.
Office of Indian Affairs. 1891. *Annual Report of the Commissioner of Indian Affairs to the Secretary of the Interior 1891*. Washington, D.C.: Government Printing Office.
Oswalt, Robert L. 1994. [Commentary on Brown (1994).] *Current Anthropology* 35(2): 111-112.
Pagel, Mark and Andrew Meade. 2006. Estimating Rates of Lexical Replacement on Phylogenetic Trees of Languages. In Forster and Renfrew (2006), 173-182.
Peter, Katherine. 1979. *Gwich'in Junior Dictionary*. Anchorage, AK: National Bilingual Materials Development Center. [Accessible online at the Alaska Native Language Archive: http://www.uaf.edu/anla/collections/dictionaries]
Pierce, Joe E., and James M. Ryherd. 1964. The Status of Athapaskan Research in Oregon. *IJAL* 30(2): 137-143.
Poser, William J. 2005. Noun Classification in Carrier. *Anthropological Linguistics* 47(2): 143-168.
Powell, John Wesley. 1880. *Introduction to the Study of Indian Languages with Words, Phrases, and Sentences to Be Collected*. Washington, D.C.: Government Printing Office.
Powers, Stephen. [1870s]. *Comparative Vocabulary of Hupa and Wailaki, without English Equivalents ca. 1870s*. National Anthropological Archives: NAA MS 220.
Powers, Stephen. 1877. *Tribes of California*. Republished 1976, edited by Robert F. Heizer. Berkeley: University of California Press.
R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. [http://www.r-project.org]
Rambaut, Andrew. 2006-2012. FigTree: Tree Figure Drawing Program, version 1.4.0. Institute of Evolutionary Biology, University of Edinburgh.
Randoja, Tiina Kathryn. 1990. *The Phonology and Morphology of Halfway River Beaver*. Ph.D. dissertation, University of Ottawa.
Reichard, Gladys. 1922. *Gladys Reichard Field Notebooks on Wiyot Indians, 1922*. BANC FILM 2732, Bancroft Library, University of California, Berkeley.
Rice, Keren. 1989. *A Grammar of Slave*. New York: Mouton de Gruyter.
Rice, Keren. 1994. Laryngeal Features in Athapaskan Languages. *Phonology* 11(1): 107-147.
Rice, Keren. 2000. *Morpheme Order and Semantic Scope: Word Formation in the Athapaskan Verb*. Cambridge University Press.
Rice, Keren. 2012. The Conservatism of Dene (Athabaskan) Template Morphology. Paper presented at the 86th Annual Meeting of the Linguistic Society of America, Portland, OR.
Rice, Keren and Leslie Saxon. 2005. Comparative Athapaskan Syntax: Arguments and Projections. In *The Oxford Handbook of Comparative Syntax*, edited by Guglielmo Cinque and Richard S. Kayne, 698-774. Oxford University Press.
Rice, Sally. 2012. "Our Language is Very Literal": Figurative Language in Dene Sųłiné [Athapaskan]. In *Endangered Metaphors*, edited by Anna Idström and Elisabeth Piirainen, 21-76. Philadelphia: John Benjamins.

Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. Indo-European and Computational Cladistics. *Trans. of the Philological Society* 100(1): 59-129.

Ronquist, Fredrick, John Huelsenbeck, and Maxim Teslenko. 2011. Draft MrBayes version 3.2 Manual: Tutorials and Model Summaries. [Online resource: http://mrbayes.sourceforge.net/mb3.2_manual.pdf]

Ronquist, F., M. Teslenko, P. van der Mark, D. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, and J.P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology*. 61(3): 539-542.

Sapir, Edward. 1914. Notes on Chasta Costa Phonology and Morphology. *University Museum Anthropological Publications* 2(2): 269-340. Philadelphia: University of Pennsylvania Museum Press.

Sapir, Edward. 1915. The Na-Dene Languages, a Preliminary Report. *American Anthropologist*, n.s., 17(3): 535-558.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech.* New York: Harcourt, Brace & Co.

Sapir, Edward. 1923. A Type of Athabaskan Relative. *IJAL* 2(3/4): 136-142.

Sapir, Edward. 1949. Dialect. In *Selected Writings in Language, Culture, and Personality*, edited by David G. Mandelbaum, 83-88. Berkeley: University of California Press.

Sapir, Edward and Victor Golla. 2001. Hupa Texts, with Notes and Lexicon. In *Collected Works of Edward Sapir, Vol. 14: Northwest California Linguistics*, edited by Victor Golla and Sean O'Neill, 19-1011. New York: Mouton de Gruyter.

Schilling-Estes, Natalie and Walt Wolfram. 1999. Alternative Models of Dialect Death: Dissipation vs. Concentration. *Language* 75(3): 486-521.

Schwartz, E.A. 1997. *The Rogue River Indian War and Its Aftermath, 1850-1980.* Norman: University of Oklahoma Press.

Scollon, Ronald. 1979. Variable Data and Linguistic Convergence: Texts and Contexts in Chipewyan. *Language in Society* 8(2): 223-243.

Scouler, John. 1841. Observations on the Indigenous Tribes of the N.W. Coast of America. *Journal of the Royal Geographical Society of London* 11: 215-251.

Seaburg, William R. 1977. A Wailaki (Athapaskan) Text with Comparative Notes. *IJAL* 43(4): 327-332.

Sicoli, Mark A. and Gary Holton. 2013. Methods and Questions in Applying Computational Phylogenetics to Na-Dene. Paper presented at the 2013 Winter Meeting of the Society for the Study of Indigenous Languages of the Americas, Boston, MA.

Story, Gillian. 1989. The Athapaskan First Duoplural Subject Prefix. In *Athapaskan Linguistics: Current Perspectives on a Language Family*, edited by Keren D. Rice and Eung-Do Cook, 487-532. New York: Mouton de Gruyter.

Swadesh, Morris. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *IJAL* 21(2): 121-137.

Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. 2012. Borrowability and the Notion of Basic Vocabulary. In *Quantitative Approaches to Linguistic Diversity: Commemorating the Centenary of the Birth of Morris Swadesh*, edited by Søren Wichmann and Anthony P. Grant, 35-55. Amsterdam: Benjamins.

Tenenbaum, Joan Marsha. 1978. *Morphology and Semantics of the Tanaina Verb.* Ph.D. dissertation, Columbia University.

Taylor, Alexander. 1860. [Oregon Athabaskan wordlist collected by Lorenzo Hubbard.] *California Farmer and Journal of Useful Sciences* 13.16.
Thomason, Sarah Grey. 2001. *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
Thomason, Sarah Grey. 2007. Language Contact and Deliberate Change. *Journal of Language Contact* 1: 41-62.
Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
Thompson, Chad. 1993. The Areal prefix h~~u~~- in Koyukon Athapaskan. *IJAL* 59(3): 315-333.
Thompson, Chad. 1996. The History and Function of the yi-/bi- Alternation in Athabaskan. In Jelinek et al. (1996), 81-100.
Thompson, Chad. 2000. Iconicity and Word Order in Koyukon Athabaskan. In *The Athabaskan Languages: Perspectives on a Native American Language Family*, edited by Theodore B. Fernald and Paul R. Platero, 228-251. Oxford University Press.
Trudgill, Peter. 1983. *On Dialect: Social and Geographical Perspectives*. Oxford: Blackwell.
Trudgill, Peter. 1986. *Dialects in Contact.* (Language in society 10.) New York: Blackwell.
Trudgill, Peter. 2004. *New-dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh UP.
Trudgill, Peter. 2008a. Colonial Dialect Contact in the History of European Languages: On the Irrelevance of Identity to New-dialect Formation. *Language in Society* 37: 241-254.
Trudgill, Peter. 2008b. On the Role of Children, and the Mechanical View: A Rejoinder. *Language in Society* 37: 277-280.
Tuten, Donald N. 2008. Identity Formation and Accommodation: Sequential and Simultaneous Relations. *Language in Society* 37: 259-262.
Tuttle, Siri. 1998. *Metrical and Tonal Structures in Tanana Athabascan*. Ph.D. dissertation, University of Washington.
Tuttle, Siri. n.d. [Notes on the Galice Creek dialect of Oregon Athabaskan from the Melville Jacobs Papers.] Manuscript, University of Alaska, Fairbanks.
Vajda, Edward J. 2010. A Siberian Link with Na-Dene Languages. In *The Dene-Yeniseian Connection*, edited by James Kari and Ben A. Potter, 33-99. (Anthropological Papers of the University of Alaska, n.s., 5, nos. 1-2). Fairbanks, AK: University of Alaska Department of Anthropology and the Alaska Native Language Center.
Wallace, William J. 1978. Hupa, Chilula, and Whilkut. In *Handbook of North American Indians*, vol. 8, *California*, edited by Robert F. Heizer, 164-179. Washington: Smithsonian Institution.
Waterman, Thomas Talbot. n.d. *The Athapascan Indians of Southwestern Oregon and Northwestern California*. National Anthropological Archives: NAA MS 3183.
Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. Empirical Foundations for a Theory of Language Change. In *Directions for Historical Linguistics: A Symposium*, edited by W.P. Lehmann and Yakov Malkiel, 95-195. Austin: University of Texas Press.
Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden, MA: Blackwell.
Wisniewski, Tomasz. n.d. The Boundaries in the Bipartite Model of the Athabaskan Verb. [Online resource, http://media.leidenuniv.nl/legacy/thesis_wisniewski.pdf]
Wolfram, Walt and Natalie Schilling-Estes. 2003. Dialectology and Linguistic Diffusion. In *The Handbook of Historical Linguistics*, edited by Brian D. Joseph and Richard D. Janda, 713-735. Malden, MA: Blackwell.

Woodward, Mary F. 1953. [Hupa texts and vocabulary.] Survey of California and Other Indian Languages, University of California Berkeley, Woodward.001 to Woodward.003.

Yurok Language Project Digital Archive. 2001-2013. [Online resource: http://linguistics.berkeley.edu/~yurok]

Zenk, Henry. 1988. Chinook Jargon in the Speech Economy of the Grand Ronde Reservation, Oregon: An Ethnography-of-Speaking Approach to an Historical Case of Creolization in Process. *International Journal of the Sociology of Language* 71: 107-124.