

UC Berkeley

Dissertations, Department of Linguistics

Title

Studies in the History and Geography of California Languages

Permalink

<https://escholarship.org/uc/item/7zn0q7hr>

Author

Haynie, Hannah

Publication Date

2012

Studies in the History and Geography of California Languages

By

Hannah Jane Haynie

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, Chair

Professor Leanne Hinton

Professor Johanna Nichols

Fall 2012

Abstract

Studies in the History and Geography of California Languages

by

Hannah Jane Haynie

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Andrew Garrett, Chair

This dissertation uses quantitative and geographic analysis techniques to examine the historical and geographical processes that have shaped California’s linguistic diversity. Many questions in California historical linguistics have received diminishing attention in recent years, remaining unanswered despite their continued relevance. The studies included in this dissertation reinvigorate some of these lines of inquiry by introducing new analytical techniques that make effective use of computational advances and existing linguistic data. These studies represent three different scales of historical change – and associated geographic patterns – and demonstrate the broad applicability of new statistical and geographic methodologies in several areas of historical linguistics.

The first of these studies (Chapter 2) focuses on the dialect scale, examining the network of internal diversity within the Eastern Miwok languages of the Central Sierra Nevada foothills. This study uses dialectometric measures of linguistic differentiation and geographic distance models to characterize the dialect geography of this language family and examine how human-environment relations have influenced its development. This study finds three primary linguistic divisions in the Eastern Miwok dialect network, corresponding to Plains Miwok, Southern Sierra Miwok and Northern/Central Sierra Miwok, as well as a number of smaller patterns of regional variation. It also identifies elevation, vegetation, and surface water as influences on the dialect network in the region and establishes the utility of cost distance modeling for studying historical linguistic contact networks in situations where our historical knowledge is limited.

The second study (Chapter 3) evaluates the hypothesis that the small families and isolate languages of California form a few, deep genealogical “stocks”. While attempts to validate two of these – Hokan and Penutian – have not met with widespread approval, the classifications themselves have been adopted widely. This study examines the statistical evidence for such deep, stock-level relationships among California languages by implementing a metric of recurrent sound correspondence and a Monte Carlo-style test for significance. The multilateral comparison involved in the clustering component of this method makes it particularly sensitive to the types of large clusters that might represent “Hokan” and “Penutian” groups. However, this test finds no evidence for such groupings and casts further doubt on the genealogical status of these categories.

The scale of the final study (Chapter 4) is broader both temporally and geographically. Chapter 4 examines the idea that Northern California functions as a linguistic area. Uncertainty regarding the genealogical and contact-related influences on individual languages in the region and links between Northern California and other linguistic areas make it difficult to evaluate existing proposals about the region's areal status based only on the regional similarities such studies offer as evidence. This chapter uses measures of spatial autocorrelation to determine whether the spatial patterns exhibited by individual features and cumulative patterns in the region as a whole are likely to reflect a history of geographic trait diffusion. While there is good evidence for areal feature spread in Northern California, and particularly in the Northwestern California and Clear Lake areas, many of the features that occur in Northern California extend up the Pacific coast and suggest that Northern California may be better characterized as a peripheral part of the better-supported Northwest Coast linguistic area.

For my father, Gary D. Haynie

Contents

List of Figures	vi
List of Tables	x
Acknowledgments	xiii
1 Introduction	1
1.1 Background	1
1.2 Theoretical Questions	2
1.3 Outline of Studies	4
2 Eastern Miwok Dialect Geography	6
2.1 Introduction	6
2.1.1 Eastern Miwok	7
2.1.2 Archival Resources	17
2.1.3 Eastern Miwok Dialectology	19
2.2 Eastern Miwok Dialect Features	22
2.2.1 Phonological Changes	23
2.2.2 Lexical Changes	27
2.2.3 Summary	34
2.3 Quantitative Analysis and Hypothesis Testing	34
2.3.1 Aggregate Measures of Linguistic Dissimilarity	35
2.3.2 Relating Linguistic Differences to Geography	36
2.3.3 Distance Models	38
2.4 Quantitative Analysis Results	42
2.4.1 Linguistic Distance	42
2.4.2 Correlations with Geographic Distances	47
2.4.3 Regional Results	48
2.5 Discussion	51
2.6 Conclusions	55
3 Deep Relationships Among California Languages	57
3.1 Introduction	57
3.2 Hokan and Penutian as Genealogical Classifications	58
3.2.1 Origins of the Hokan and Penutian Stocks	58

3.2.2	Evidence in Support of Hokan and Penutian as Genealogical Stocks . . .	61
3.2.3	Controversy	62
3.3	Kessler's R^2 : A Suitable Method for California	63
3.3.1	Reviving Lexicostatistics	63
3.3.2	Methodology	65
3.3.3	Statistical Significance	69
3.3.4	Comparing Kessler's R^2 across Languages with Different Phonologies	71
3.3.5	Multilateral Clustering	73
3.4	Applying this method to California	75
3.4.1	The Language Sample	75
3.4.2	Word List Data	77
3.4.3	Predictions	78
3.5	Results	78
3.5.1	Clustering	79
3.5.2	Unexpected Findings	81
3.5.3	Evaluating the Results	84
3.6	Conclusions	85
4	The Geography of the Northern California Linguistic Area	88
4.1	Introduction	88
4.1.1	Northern California in Areal Linguistics	89
4.1.2	Language Areas in Linguistic Theory	96
4.2	Methods	97
4.3	Results	102
4.3.1	Phonological Traits	102
4.3.2	Morphological Traits	116
4.4	Discussion	127
4.4.1	Evidence for Geographic Feature Diffusion in Northern California . .	127
4.4.2	Geographic Patterns	128
4.4.3	Overlap of Features	132
4.5	Conclusions	134
5	Conclusion	136
5.1	Summary of Findings	136
5.2	Discussion of Methodological Advances	138
5.3	Theoretical Impacts	139
5.4	Directions for Future Work	140
	Bibliography	142
	Appendix A: Lexical Isogloss Maps of Eastern Miwok	159
	Appendix B: Select Mantel Correlations for Eastern Miwok Language/Geography Relationships	259

Appendix C: Expanded Kessler's R^2 Results	261
Appendix D: California Word List	273

List of Figures

2.1	Miwok family tree	7
2.2	Location of the Eastern Miwok language family in California	8
2.3	Eastern Miwok region including towns and rivers	9
2.4	Merriam's divisions of the 'Moquelumnan' territory (after Merriam 1907: plate XXV)	11
2.5	Kroeber's Miwok geography (after Kroeber 1925: plate 37)	13
2.6	Barrett's Miwok geography (after Barrett 1908a: 369)	14
2.7	Diffusion of $*\eta \rightarrow n$ change in Eastern Miwok	24
2.8	Reflexes of Proto-Eastern Miwok $*\check{s}$	25
2.9	Percent occurrence of u as a reflex of Proto-Eastern Miwok $*y$ in 24 forms, represented by percent gray	26
2.10	Lexical isogloss density in Eastern Miwok	27
2.11	Clustering of Eastern Miwok varieties based on phonological (Levenshtein) distances	43
2.12	Clustering of Eastern Miwok varieties based on lexical distances	44
2.13	3-Dimensional scaling of Eastern Miwok phonological (Levenshtein) distances	45
2.14	3-Dimensional scaling of Eastern Miwok lexical distances	46
3.1	California language map showing stock-level groupings	58
3.2	Dendrogram representation of identified California language relationships	80
3.3	Average linking clustering of California languages	84
4.1	Campbell's Northern California area	92
4.2	Northern California areal linguistic study area	99
4.3	Effect of study area size on spatial autocorrelation	101
4.4	Distribution of plain, glottalized, and aspirated stop series contrast	103
4.5	Distribution of glottalized stops	104
4.6	Distribution of five-vowel systems	105
4.7	Distribution of "back" apical stops	106
4.8	Distribution of "back velar" consonants	107
4.9	Distribution of "back velar" stops	108
4.10	Distribution of "back velar" fricatives	109
4.11	Distribution of fricated or affricated lateral consonants	110
4.12	Distribution of lateral fricatives	111
4.13	Distribution of plain lateral affricates	112

4.14	Distribution of glottalized lateral affricates	113
4.15	Distribution of contrast between <i>r</i> and <i>l</i>	114
4.16	Distribution of velar nasal consonants	115
4.17	Distribution of sound symbolic diminutivization processes	116
4.18	Distribution of quinary numeral systems	117
4.19	Distribution of nominal reduplication associated with plural meanings	118
4.20	Distribution of verbal reduplication with iterative or distributive meanings	119
4.21	Distribution of case marking systems	120
4.22	Distribution of <i>n</i> -initial first person pronoun	121
4.23	Distribution of <i>m</i> -initial second person pronoun	122
4.24	Distribution of prefixed possessive markers	123
4.25	Distribution of prefixed subject markers	124
4.26	Distribution of pre-verbal tense and aspect markers	125
4.27	Distribution of numeral classifier systems	126
4.28	Total number of significant areal features found in Northern California and surrounding areas	133
5.1	Eastern Miwok isogloss map: Acorn	160
5.2	Eastern Miwok isogloss map: Afraid	161
5.3	Eastern Miwok isogloss map: Arm	162
5.4	Eastern Miwok isogloss map: Arrow	163
5.5	Eastern Miwok isogloss map: Ashes	164
5.6	Eastern Miwok isogloss map: Bad	165
5.7	Eastern Miwok isogloss map: Basket (generic)	166
5.8	Eastern Miwok isogloss map: Black	167
5.9	Eastern Miwok isogloss map: Black Oak	168
5.10	Eastern Miwok isogloss map: Bluejay	169
5.11	Eastern Miwok isogloss map: Bone	170
5.12	Eastern Miwok isogloss map: Bow	171
5.13	Eastern Miwok isogloss map: Boy	172
5.14	Eastern Miwok isogloss map: Buzzard	173
5.15	Eastern Miwok isogloss map: Chief	174
5.16	Eastern Miwok isogloss map: Coyote	175
5.17	Eastern Miwok isogloss map: Dance	176
5.18	Eastern Miwok isogloss map: Deer	177
5.19	Eastern Miwok isogloss map: Doctor	178
5.20	Eastern Miwok isogloss map: Door	179
5.21	Eastern Miwok isogloss map: Ear	180
5.22	Eastern Miwok isogloss map: Earth	181
5.23	Eastern Miwok isogloss map: Eat	182
5.24	Eastern Miwok isogloss map: Eye	183
5.25	Eastern Miwok isogloss map: Fingernails	184
5.26	Eastern Miwok isogloss map: Fingers	185
5.27	Eastern Miwok isogloss map: Fire	186

5.28	Eastern Miwok isogloss map: Fish (generic)	187
5.29	Eastern Miwok isogloss map: Fly	188
5.30	Eastern Miwok isogloss map: Foot	189
5.31	Eastern Miwok isogloss map: Friend	190
5.32	Eastern Miwok isogloss map: Girl	191
5.33	Eastern Miwok isogloss map: Good	192
5.34	Eastern Miwok isogloss map: Grasshopper	193
5.35	Eastern Miwok isogloss map: Green	194
5.36	Eastern Miwok isogloss map: Hair	195
5.37	Eastern Miwok isogloss map: Hand	196
5.38	Eastern Miwok isogloss map: Head	197
5.39	Eastern Miwok isogloss map: House	198
5.40	Eastern Miwok isogloss map: Hummingbird	199
5.41	Eastern Miwok isogloss map: Infant	200
5.42	Eastern Miwok isogloss map: Jackrabbit	201
5.43	Eastern Miwok isogloss map: Jump	202
5.44	Eastern Miwok isogloss map: Kill	203
5.45	Eastern Miwok isogloss map: Knife	204
5.46	Eastern Miwok isogloss map: Large	205
5.47	Eastern Miwok isogloss map: Leg	206
5.48	Eastern Miwok isogloss map: Lizard (generic)	207
5.49	Eastern Miwok isogloss map: Man	208
5.50	Eastern Miwok isogloss map: Meat	209
5.51	Eastern Miwok isogloss map: Medicine	210
5.52	Eastern Miwok isogloss map: Mosquito	211
5.53	Eastern Miwok isogloss map: Mountain	212
5.54	Eastern Miwok isogloss map: Mouth	213
5.55	Eastern Miwok isogloss map: Mush Paddle	214
5.56	Eastern Miwok isogloss map: Mushroom (generic)	215
5.57	Eastern Miwok isogloss map: Neck	216
5.58	Eastern Miwok isogloss map: Nine	217
5.59	Eastern Miwok isogloss map: Nose	218
5.60	Eastern Miwok isogloss map: Old Man	219
5.61	Eastern Miwok isogloss map: Old Woman	220
5.62	Eastern Miwok isogloss map: One	221
5.63	Eastern Miwok isogloss map: Owl (generic)	222
5.64	Eastern Miwok isogloss map: Pestle	223
5.65	Eastern Miwok isogloss map: Pipe	224
5.66	Eastern Miwok isogloss map: Poison	225
5.67	Eastern Miwok isogloss map: Quail (generic)	226
5.68	Eastern Miwok isogloss map: Rain	227
5.69	Eastern Miwok isogloss map: Rattlesnake	228
5.70	Eastern Miwok isogloss map: Red	229
5.71	Eastern Miwok isogloss map: Rib	230

5.72	Eastern Miwok isogloss map: Run	231
5.73	Eastern Miwok isogloss map: Salmon	232
5.74	Eastern Miwok isogloss map: Sand	233
5.75	Eastern Miwok isogloss map: See	234
5.76	Eastern Miwok isogloss map: Seven	235
5.77	Eastern Miwok isogloss map: Shoot	236
5.78	Eastern Miwok isogloss map: Sing	237
5.79	Eastern Miwok isogloss map: Older Sister	238
5.80	Eastern Miwok isogloss map: Small	239
5.81	Eastern Miwok isogloss map: Smoke	240
5.82	Eastern Miwok isogloss map: South	241
5.83	Eastern Miwok isogloss map: Star	242
5.84	Eastern Miwok isogloss map: String	243
5.85	Eastern Miwok isogloss map: Sun	244
5.86	Eastern Miwok isogloss map: Ten	245
5.87	Eastern Miwok isogloss map: Thunder	246
5.88	Eastern Miwok isogloss map: Tobacco	247
5.89	Eastern Miwok isogloss map: Tree (generic)	248
5.90	Eastern Miwok isogloss map: Up	249
5.91	Eastern Miwok isogloss map: Valley	250
5.92	Eastern Miwok isogloss map: West	251
5.93	Eastern Miwok isogloss map: White	252
5.94	Eastern Miwok isogloss map: White Man	253
5.95	Eastern Miwok isogloss map: White Oak	254
5.96	Eastern Miwok isogloss map: Wind	255
5.97	Eastern Miwok isogloss map: Wood	256
5.98	Eastern Miwok isogloss map: Woodpecker	257
5.99	Eastern Miwok isogloss map: Yellowjacket	258

List of Tables

2.1	Eastern Miwok documentation by location	19
2.2	Sound changes proposed in Callaghan and Broadbent's reconstruction of Proto-Miwok (1960:302)	23
2.3	Lexical forms represented by Eastern Miwok isogloss bundle A	29
2.4	Lexical forms whose boundaries fall near Eastern Miwok isogloss bundle A	30
2.5	Lexical forms found only in Ione	30
2.6	Lexical forms represented by Eastern Miwok isogloss bundle B	31
2.7	Lexical forms whose boundaries fall near Eastern Miwok isogloss bundle B	31
2.8	Lexical forms represented by Eastern Miwok isogloss bundle C	31
2.9	Lexical forms whose boundaries fall near isogloss bundle C	32
2.10	Lexical forms represented by Eastern Miwok isogloss bundle D	33
2.11	Lexical forms represented by Eastern Miwok isogloss bundle E	33
2.12	Lexical forms represented by Eastern Miwok isogloss bundle F	33
2.13	Lexical forms represented by Eastern Miwok isogloss bundle G	33
2.14	Human walking coefficients from Giovani and Goldman (1971)	40
2.15	Surface water crossing coefficients	41
2.16	R^2 values for significant correlations between linguistic and geographic distances – all Eastern Miwok locations	47
2.17	R^2 values for significant correlations between linguistic and geographic distances – Central Sierra, Northern Sierra and Plains Miwok	49
2.18	R^2 values for significant correlations between linguistic and geographic distances – Southern Sierra Miwok	49
2.19	R^2 values for significant correlations between linguistic and geographic distances – Central Sierra Miwok	49
2.20	R^2 values for significant correlations between linguistic and geographic distances – Northern Sierra and Plains Miwok	49
3.1	Pairwise counts of “stem resemblances” among California languages reported by Dixon and Kroeber (1919:51)	60
3.2	Contingency tables representing different distributions of 100 items	67
3.3	Contingency tables with 1 subtracted from each non-zero cell to eliminate non-recurring correspondences	68
3.4	Contingency tables with recurrence values squared	68
3.5	Random rearrangement of simulated word list data	70
3.6	Even distributions of 100 items in contingency tables of different sizes	71

3.7	Contingency tables for randomly reorganized simulated data	72
3.8	Sample table of normalized recurrence metrics	73
3.9	Contingency table for simulated languages 1 and 2	74
3.10	Contingency table for a simulated comparison between a cluster (languages 1 and 2) and language 3	74
3.11	Round by round building of California phylogenetic groupings	79
4.1	Spatial autocorrelation of plain/glottalized/aspirated stop series contrast . .	103
4.2	Spatial autocorrelation of glottalized stops	104
4.3	Spatial autocorrelation of five-vowel system	105
4.4	Spatial autocorrelation of “back” apical stops	106
4.5	Spatial autocorrelation of “back velar” consonants	107
4.6	Spatial autocorrelation of “back velar” stops	108
4.7	Spatial autocorrelation of “back velar” fricatives	109
4.8	Spatial autocorrelation of lateral fricatives and affricates	110
4.9	Spatial autocorrelation of lateral fricatives	111
4.10	Spatial autocorrelation of plain lateral affricates	112
4.11	Spatial autocorrelation of glottalized lateral affricates	113
4.12	Spatial autocorrelation of the <i>r/l</i> contrast	114
4.13	Spatial autocorrelation of velar nasal consonants	115
4.14	Spatial autocorrelation of sound symbolic diminutivization processes	116
4.15	Spatial autocorrelation of quinary numeral systems	117
4.16	Spatial autocorrelation of nominal reduplication with plural semantics	118
4.17	Spatial autocorrelation of verbal reduplication with iterative or distributive semantics	119
4.18	Spatial autocorrelation of case marking systems	120
4.19	Spatial autocorrelation of <i>n</i> -initial first person pronouns	121
4.20	Spatial autocorrelation of <i>m</i> -initial second person pronouns	122
4.21	Spatial autocorrelation of prefixed possessive markers	123
4.22	Spatial autocorrelation of prefixed subject markers	124
4.23	Spatial autocorrelation of pre-verbal tense and aspect markers	125
4.24	Spatial autocorrelation of numeral classifiers	126
4.25	Spatial autocorrelation of significant areal feature counts	133
5.1	<i>r</i> values for significant Mantel correlations between linguistic and geographic distances – all Eastern Miwok locations	259
5.2	<i>r</i> values for significant Mantel correlations between linguistic distance and rejected models of cost distance	259
5.3	<i>r</i> values for significant Mantel correlations between phonological distance (based on data with segment length distinctions) and models of geographic distance	260
5.4	Multilateral clustering: Round 1	262
5.5	Multilateral clustering: Round 2	263
5.6	Multilateral clustering: Round 3	264

5.7	Multilateral clustering: Round 4	265
5.8	Multilateral clustering: Round 5	266
5.9	Multilateral clustering: Round 6	267
5.10	Multilateral clustering: Round 7	268
5.11	Multilateral clustering: Round 8	269
5.12	Multilateral clustering: Round 9	270
5.13	Multilateral clustering: Round 10	271
5.14	Multilateral clustering: Round 11	272

Acknowledgments

The members of my committee have made many contributions to this dissertation and to my overall development as a scholar, and I owe them each many thanks. Before taking Andrew Garrett's historical linguistics course at Berkeley I was fairly certain that I had no interest in diachronic linguistics. Andrew's teaching opened my eyes to many compelling questions and ideas, and his advising has instilled in me a sense of excitement for research as well as rigorous standards of scholarship. He has made it possible to pursue some somewhat unconventional ideas, and then has made these studies better through his feedback. Leanne Hinton introduced me to California languages through the Breath of Life workshop for language revitalization and a field methods class on Southeastern Pomo. She has been a tremendous source of knowledge about California languages and cultures and a kind but uncompromising advisor. Johanna Nichols has brought new ideas, important articles, and interesting data to my attention in every conversation I have had with her. Her encouragement has been as valuable to me as her expert advising. I thank each of these mentors for the effort, time, and care they have dedicated to my work and scholarly development.

While her training is in geophysics, my sister Jill has also offered insightful comments in conversations about historical linguistics. Without her input on statistical and geographic analysis, the many methodological decisions that went into this project would have been much more daunting and some of the inevitable hiccups in implementation would have turned into real headaches.

I also owe a great deal to members of the Berkeley linguistics community. The staff who keep the department running smoothly have also helped me in many ways, from course registration to conference planning, and even physically filing this dissertation. My hat is off to Belen Flores, Paula Floro, Natalie Babler, Jane Paris, and Ronald Sprouse. Berkeley's linguistics faculty has also been tremendously supportive during my time in the department, and I thank them for the knowledge and unique perspectives they have shared with me. I would like to give special thanks to Line Mikkelsen for her mentorship and her continual encouragement. I am also grateful for the graduate student colleagues who have served as sounding boards for ideas and provided friendship within the halls of Dwinelle and beyond. Michael Houser has been a particularly loyal and kind friend and has provided feedback on much of the material in this dissertation, and for that I thank him.

The Bancroft Library research involved in this project was generously supported by a *Bancroft Library Study Award*. I have also received support as a Graduate Student Researcher in the Survey of California and Other Indian Languages. These two institutions hold incredible collections of California linguistic materials, and their dedication to preserving these records and creating communities of scholarship is commendable. I owe an even greater debt

of gratitude to the speakers and scholars whose work is archived in these repositories.

Finally, I am thankful to my family, who are an extraordinarily caring and supportive group of people. Without their helpfulness and humor I wouldn't have graduated from high school, let alone a Ph.D. program. To my family, to the many friends whose good nature has brightened my days over the last several years, and to the creatures whose companionship has helped me keep an even keel, thank you.

Chapter 1

Introduction

1.1 Background

California is a region of extraordinary linguistic diversity, with at least a dozen language families within its borders, numerous linguistic isolates, and relatively high levels of diversity within and across its families. Its history – how this diversity came to exist in this region – was not recorded before the arrival of Euro-American settlers in the region in the late 18th century. Traces of this history were further obscured by the resettlement of indigenous people and the language loss that occurred between that time and the early 20th century, when these languages became the focus of scholarly documentation efforts. Many questions about California’s linguistic history were asked as new data became available in the 20th century, and while some found clear answers, others were met with controversy or only prompted more questions. Scholarly interest in many aspects of California historical linguistics has diminished in recent decades because of perceived barriers to their study, including the continuing trend of language loss in the region and the limits this has imposed on the amount of new linguistic information that is discoverable. However, recent advances in computational technology and statistical methodologies make it possible to use the existing data more effectively and gain new perspectives on these unanswered historical questions.

This dissertation breathes new life into several of these old questions in California historical linguistics, each representing a different temporal scale of linguistic development and a different tradition within historical linguistics. While some new data from California languages have become available through recent publications and archival contributions, it is primarily the exponential advance of computer technology over the last thirty years that makes it possible to revisit these topics, which previously appeared to be intractable.

The shallowest scale of study represented in this project is the dialect scale, and this project’s dialect study (Chapter 2) combines new techniques from dialectometry with Geographic Information Systems analysis to investigate a dialect network that developed through regional contact during a relatively recent period of history. At an intermediate temporal scale, but one that represents thousands of years of history, the second study of this dissertation (Chapter 3) examines genealogical relationships between California’s languages. This component of the dissertation employs a relatively new measure of the recurrence of sound correspondences and a Monte Carlo-style significance test that would have

been computationally prohibitive during the early- to mid-20th century when classification of California's languages was a new and particularly active area of research. The third scale at which this dissertation focuses is the broadest. This final study (Chapter 4) investigates geolinguistic patterns resulting from events that occurred during the period that is traceable through phylogenetic studies and likely even earlier. The areal study that is aimed at this vast historical scale is also geographically broad, looking not only at the patterns within Northern California, but also the surrounding regions. Geographic analysis, made possible by computationally intensive programs, provides a new perspective on areal trait diffusion in this region.

Thus, these studies not only highlight different temporal scales of linguistic development, but they also touch upon several areas of diachronic linguistics – namely dialectology, phylogenetics, and areal linguistics – and encompass different geographic extents, from the Sierra foothill region, to the state of California, to the West Coast. For each of these diverse studies, methods were carefully chosen or developed to extract as much information as possible from the available data while taking into consideration the methods and standards that would traditionally be applied to each question. The diversity of topics and spatiotemporal scales represented in this work demonstrates the broad spectrum of historical linguistic questions to which new computational and geographic methods can be fruitfully applied.

One other thread that unites at least some of these studies is an emphasis on geographic perspectives and techniques. This theme represents a belief that spatial patterns in human organization and culture generally encode evidence of history, and that these historical signals can be advantageously exploited in situations where more direct historical evidence is limited. This view is gaining popularity within linguistics, but its implementation in linguistic research has thus far been limited. A geospatially sophisticated approach to diachronic questions helps in identifying genealogical and geographic signals in linguistic data and understanding how each of these mechanisms have contributed to linguistic development over time. As a result, the types of analysis introduced in this dissertation are likely to produce interesting results even in areas that are not subject to the same data limitations as California.

1.2 Theoretical Questions

The theoretical implications of this collection of studies extend beyond just the introduction of a geographic viewpoint and a set of analytical tools, however. In very general terms, this dissertation is focused on identifying evidence of the historical processes that have shaped languages and language families. The studies in Chapters 2–4 are broadly organized around the questions of how genealogical inheritance and geographic diffusion interact at various stages of historical development and how the manifestations of these processes differ as a function of time depth.

At the dialect scale we see diversification in progress, with regional patterns in variation representing the emerging divisions between lects that may over time develop into languages, which over even greater spans of time would continue to evolve and could branch into generations of daughter languages. The incipient divisions between linguistic varieties that are

found at this scale of study may not, in many cases, be perceived by speakers as different ways of speaking, and the subtler variation between nearby communities is even likelier to go unnoticed by speakers. Yet the patterns of contact between these closely related varieties have a profound impact on the development of family-internal diversity that eventually leads to branching in the family tree. How specific situations of contact within dialect networks relate to detailed patterns of dialect diversity is therefore an important question about the relationship between contact and genealogical descent. The literature relating contact conditions to specific patterns in language development has discussed straightforwardly polylectal communities in both broad terms (e.g. Thomason and Kaufman 1988) and with more detailed consideration of social network characteristics (e.g. Ross 2003), but the type of internal contact described and analyzed in Chapter 2 is discussed within the field of dialectology more often than it is discussed in relation to more general historical linguistic questions (e.g. as in Milroy and Milroy 1985; Labov 2007). The case study in Chapter 2 describes the dialect diversity of Eastern Miwok using methods from dialectology, but perhaps more importantly it relates patterns in that diversity to specific conditions of contact within the dialect network that are associated with these outcomes.

Identifying relationships between conditions of contact and their outcomes within linguistic groups, such as Eastern Miwok, helps to answer several larger questions. The first is how the dynamics of contact in essentially monolectal communities condition processes of language diversification and leveling, and whether the generalizations described in broader studies of contact and language change apply in these situations. Whereas Ross's (2003) framework was intended to assist in diagnosing prehistoric contact between languages, insights from studies like Chapter 2 are useful for developing a way to understand historical contact between similar language varieties based on observable patterns.

The dialect-scale study also leads to insights about how concrete environmental factors impact processes of linguistic change. Sociolinguistic context is commonly considered to be an important contributor to linguistic development, and even historical, political, and cultural elements may be considered part of the context of language change. But the identification of links between elements of the physical environment and patterns of linguistic development suggests that to truly understand linguistic development in an ecological context, we must also pay attention to potential influences of physical phenomena on human behavior. Just as Nichols has found geography to be a strong predictor of global typological patterns (1992) and has continued to demonstrate the impact of geography on the history and spread of linguistic traits (e.g. 1993; 1999), this dissertation looks to geography for answers about why patterns in dialect and language diversity have evolved the way they have. The role of physical factors in conditioning linguistic developments is demonstrated here in a dialect network in a rugged environment, but just as social and cultural conditions are relevant to multiple scales of linguistic history, we should also expect the impact of the physical environment to be manifested in the outcomes of linguistic development over different spans of time.

At deeper temporal scales, the role of geography in linguistic development is most often discussed in terms of areal diffusion and the relative contributions of diffusion and descent to the resemblances between languages. While these processes of language change can in many cases be traced through historical reconstructions and careful analysis of diffusion events,

deeper prehistoric changes and relationships are more difficult to detect. The limitations on our ability to decipher evidence of processes of prehistoric language change have created as much contention as consensus about the relative roles of diffusion and inheritance in deep linguistic relationships (e.g. the debate in Dunn et al. 2005, 2007; Dunn 2009; Donohue and Musgrave 2007; Donohue et al. 2008). This dissertation proposes that formalizing some terms and tools that make it possible to identify signals of distant genealogical relationships and geographic diffusions is a significant step toward understanding the relative importance of these processes in the development of specific linguistic outcomes. Though it may never be possible to examine prehistoric events in as detailed a manner as would be necessary to reconstruct millennia of organic language evolution, increasing the rigor with which we evaluate our hypotheses allows us to continually elevate the level of the discussion about processes of prehistoric language change.

Finally, while historical linguistic research has been conducted on languages and linguistic groups around the globe, questions remain regarding whether and how the dynamics of language change differ in societies or linguistic areas of different types. This issue is sometimes invoked in questioning the applicability of methods developed for Indo-European languages to very different sociolinguistic or historical situations, and has been pursued as a topic of research in its own right. This dissertation focuses on the mechanisms of language change in a linguistically and geographically diverse region that is markedly different from the political, cultural, and geographic context of Indo-European language change. Some specific concerns about the application of established methods and perspectives in a very different sociolinguistic setting are addressed, including the interpretation of dialect patterns in a region without prominent centers of population or prestige and the comparison of languages with very different sound systems. In addition to presenting insights on the processes of change in California languages, this dissertation also investigates the universality of the techniques and theories used to examine processes of language change.

1.3 Outline of Studies

Chapter 2 presents a dialect study of Eastern Miwok, focused primarily on the Sierra Miwok subgroup due to the paucity of information about Plains Miwok and Saclan. This study explores the network of dialect diversity that exists in the Miwok-speaking areas of the western Sierra Nevada foothills and mountains and the Sacramento Valley plains. Both traditional dialect mapping techniques and dialectometric analysis are used to characterize the patterns of dialect differentiation found in this region. Models of geographic distance and connectivity are then compared to measures of linguistic differences to test hypotheses about how spatial patterns in the region's dialect diversity have arisen. This chapter not only aims to describe the Eastern Miwok dialect network, but it also explores ideas about how patterns in dialect geography have developed.

Chapter 3 takes aim at claims that California's small language families and isolates form a few large stocks that represent very historically deep genealogical relationships, and specifically examines the Hokan and Penutian hypotheses. This study uses a metric of sound correspondence recurrence, a Monte Carlo style test for significance, and a multilateral

comparison-inspired approach to clustering to detect significant relationships between California languages. This analysis includes a broad sample of California languages as well as a Uto-Aztecan sample, a control used to evaluate the ability of the method to identify large, multi-branching clusters. This chapter introduces several methodological adaptations that tailor this phylogenetic method to the languages of California.

Chapter 4 investigates the proposal that Northern California is a linguistic area. This region was described as a linguistic area by Haas (1976), and though little work has been done to evaluate the evidence for the region's areal status since that time, the region's linguistic area designation has been widely assumed to be fact. However, many of the features that are associated with Northern California are also found in other nearby regions, while others occur only in small sub-regions of Northern California. This study uses geographic analysis techniques to identify spatial patterns that are likely to represent areal feature spreads, and evaluates the overall feature distribution patterns that occur in and around Northern California to assess whether the scale and location of geographic patterns suggests that Northern California is a region of areal diffusion.

Finally, Chapter 5 summarizes the results of these studies and presents several conclusions.

Chapter 2

Eastern Miwok Dialect Geography

2.1 Introduction

This chapter presents a study of Eastern Miwok dialect geography that uses traditional isogloss mapping, quantitative methods for analyzing linguistic data in aggregate, and geographic modeling to understand patterns of dialect diversity. By focusing on the dialect scale, this investigation uses the spatial component of language change to help understand the social contact networks across which dialect leveling occurs. This study is unique in its focus on historical dialect diversity in a region for which little sociohistorical information is available and for which the available language data consists of archival field notes gathered by several early 20th century researchers. To accomplish this, the contact network geography is modeled using data describing elements of the physical environment that are hypothesized to affect contact.

This chapter aims to answer several questions. First, what is the spatial distribution of dialects within Eastern Miwok and how do the spatial patterns in dialect diversity align with the boundaries conventionally drawn within this language group? Secondly, is the linguistic dissimilarity of local speech varieties significantly correlated with geographic distances and cost-modeled geographic distances, and how much of the dialect variation in this region can be explained based on geographic information? Finally, can the approach used in this study, which correlates dialectometry metrics with costs associated with landscape features and barriers, be employed fruitfully to better understand the ways in which social contact shaped the early 20th century Eastern Miwok dialect network?

The remainder of this section introduces the Eastern Miwok language group, the data used in this study, and the approach to dialectology employed here. Section 2.2 maps the occurrence of dialect features across Eastern Miwok locations and discusses the patterns that occur in the lexical and phonological isoglosses. Section 2.3 explains the quantitative methods and geographic analysis techniques used to investigate relationships between linguistic patterns and the underlying landscape. Section 2.4 presents the results of this analysis, and Section 2.5 discusses the implications of these results for Eastern Miwok dialect geography and the utility of these methods more generally.

2.1.1 Eastern Miwok

Eastern Miwok and its sister Western Miwok (the branch that contains the Lake Miwok language south of Clear Lake and the Coast Miwok languages that stretch from Sonoma to Bodega Bay) make up the Miwok family. Miwok is commonly classified as part of the Penutian “stock” (Golla 2011; Goddard 1996), and more thorough analysis has demonstrated that Miwok groups genealogically with a subset of other Penutian languages¹. Catherine Callaghan has demonstrated a genealogical link between Miwok and the Ohlone languages of the San Francisco and Monterey Bay areas, forming the Utian subgroup, and she has presented further evidence to support a Yok-Utian group comprised of Utian and the Yokuts languages (1997; 2001). The Miwok languages are relatively well documented, especially the Eastern Miwok languages that survived after the establishment of Spanish missions in the late 18th and early 19th centuries. Traditionally, Eastern Miwok is divided into five languages: Plains Miwok, Saclan or Bay Miwok, and three Sierra Miwok languages. While the boundaries between these languages and their classification as languages rather than dialects are generally accepted, it is unclear what evidence these designations are based on. However, existing documentation is sufficient for evaluating patterns of variation within this group more thoroughly, including boundaries and degrees of difference.

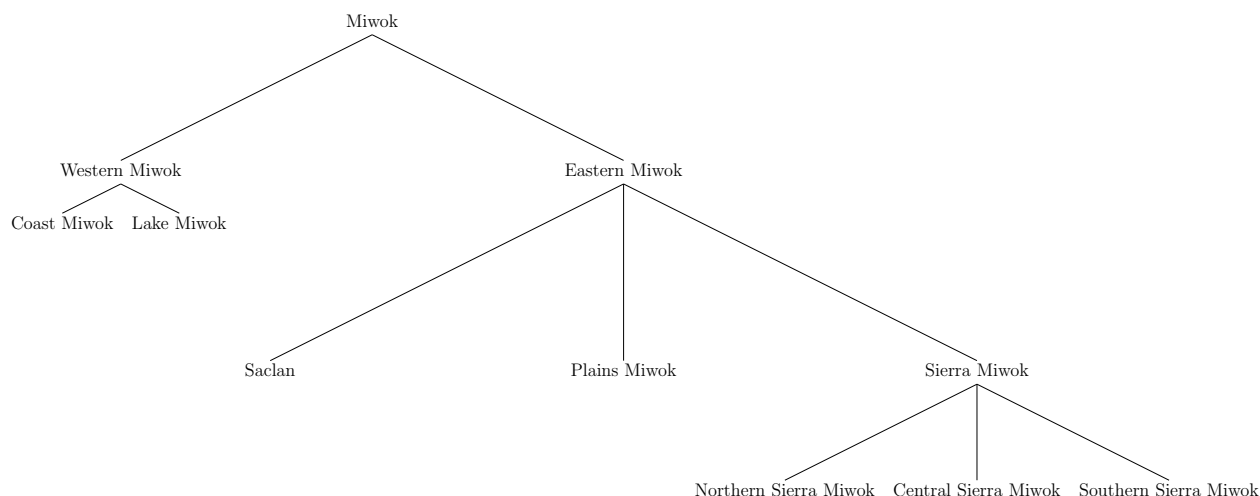


Figure 2.1: Miwok family tree

The Miwok family was originally grouped by Powell as part of the “Mutsun” family, which included Ohlone languages. The Miwok members of this proposed family were thought to have spread from the Ohlone area on California’s central coast directly east through Yokuts territory to the western slopes of the Sierra Nevada. A later revision to Powell’s language map of California identified “Moquelumnan”, or Miwok, as its own family — a great advance from the Mutsun grouping — but the internal boundaries within the family remained unclear (Powers 1877). The early ethnographers — notably Barrett, Kroeber and Merriam

¹The term “Penutian” is used here a way to orient the reader to this language’s position in California, not necessarily as a genealogical classification. For a discussion of whether the evidence supports Penutian as a genealogical grouping, see Chapter 3.



Figure 2.2: Location of the Eastern Miwok language family in California

— traveled through Miwok territory just after the turn of the twentieth century in search of both ethnographic and linguistic information. Their interest in the categorization of Miwok languages was tied to their interest in understanding the associated cultures more generally. Hence, the way in which their work touched on linguistic geography was through their simultaneous gathering of ethnogeographic information and linguistic data. This allowed Kroeber (1906), Merriam (1907) and Barrett (1908a) to propose the first level of branching of the Eastern Miwok group into four languages² and to propose some approximate language boundaries. For Barrett and Merriam this process necessarily involved mapping, and though Kroeber did not produce an Eastern Miwok map until his 1925 *Handbook of the Indians of California*, his work also emphasized an understanding of the geographic territories associated with these languages and cultures. With an established sense of the geographic boundaries between these languages and declining speaker populations limiting the ethnogeographic work that could be done, later scholars (e.g. Freeland, Broadbent, Callaghan)

²These four languages are Plains Miwok, Northern Sierra Miwok, Central Sierra Miwok, and Southern Sierra Miwok; Saclan was extinct by the time of the first ethnographic/linguistic surveys.

took note of dialect differences but did not focus on the spatial extents of these varieties.

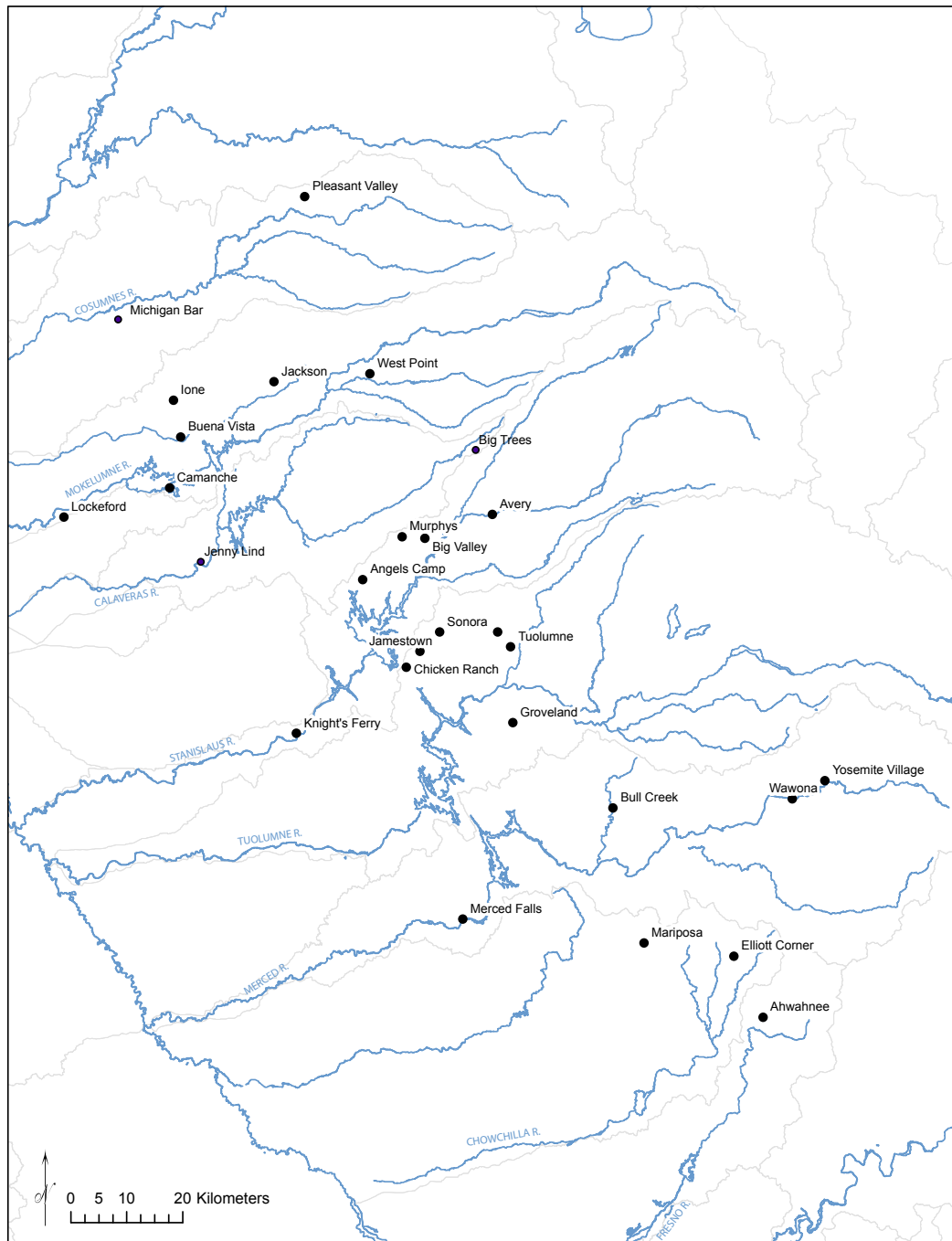


Figure 2.3: Eastern Miwok region including towns and rivers

The earliest descriptions of Sierra and Plains Miwok settlement date back to the late 18th and early 19th century exploration of these regions, recorded primarily by Spanish missionaries and explorers. Cook (1955) draws upon the logs of these early expeditions and other 19th century travelers in the San Joaquin valley and western foothills in his anal-

ysis of the population patterns of the region. Aiming for a broad picture of the region's aboriginal demographics rather than a community-by-community analysis of settlement patterns, Cook's population estimates are organized by geographic domains — primarily by river valleys, which seem to have been important for constraining the settlement patterns of what Levy calls the Eastern Miwok "tribelets" (Levy 1978:400). Cook estimates the ca. 1850 population of the Cosumnes, Mokelumne, and Calaveras River basins (roughly the traditionally mapped extent of Northern Sierra Miwok) to have been about 1,000 people, but notes that the population may have been closer to 4,000 people before the gold rush. Cook confidently reports an estimate of about 2,000 people in the Stanislaus and Tuolumne basins (traditional Central Sierra Miwok territory), with the figure again reflecting some decline in the wake of the gold rush. The Merced River, Chowchilla River, and Mariposa Creek basins (Southern Sierra Miwok) are estimated by Cook to have had about 2,500 total inhabitants. Although Spanish missions and gold mining led to greater population losses in the Cosumnes through Tuolumne basins (Northern-Central Miwok) than the Merced through Chowchilla basins (Southern Sierra Miwok), the pre-gold-rush population density throughout the Miwok-inhabited western Sierra was probably fairly even. Plains Miwok settlement, on the other hand, was much denser and perhaps the area of California's greatest population density before European contact (Levy 1978:402). Cook estimates an aboriginal population in the Miwok-inhabited northern portion of the San Joaquin valley of around 11,000 people (1955). However, this population declined rapidly after Euro-American contact, leaving only a few hundred people in the traditional geographic range of Plains Miwok by the early 20th century.

It was during this early 20th century era of population decline that scholars — mostly from the University of California, Berkeley — began to organize the aforementioned field expeditions to the plains and foothills to study Native cultures and languages. The first published material to result from this wave of documentary work that touched on the settlement patterns of Eastern Miwok speakers was Merriam's 1904 report on the distribution of indigenous people in the southern Sierra foothills and adjacent plains. This work focused more specifically on the Numic and Yokuts populations of the region, but also described the general range of the 'Muwa' or 'Moquelumnan' people as extending 110 miles from the west flank of the middle Sierra from the Cosumnes River to Fresno Creek, covering the Upper Sonoran and lower Transition ecological zones. In 1907 Merriam published a detailed ethnogeography of the 'Me'-wan' people. In this article Merriam identifies two primary groups within Eastern Miwok — the 'Me'-wuk' or Sierra Miwok and the 'Mew'-ko' or Plains Miwok. Interestingly, Merriam's classification system for Me'wuk, or Eastern Miwok, divides Me'wuk into three main groups (Northern, Middle, and Southern) but finds ten distinctions in the Mew'-ko group: Hul-poom'-ne, Mo-koz'-um-ne, Mo-kal'-um-ne, Chil-lum'-ne, Si-a-kum-ne, Tu-ol'-um-ne, O'-che-hak, Wi'-pā, Han-ne'-suk, and Yatch-a-chum'-ne. In addition to finding a greater level of diversity within Plains Miwok (or Mew'-ko) than has been described by other researchers, Merriam's map of the Miwok territories (Figure 2.4) extends the Plains Miwok territory far south of the boundary described in all other accounts. On this map Miwok territory extends through the plains traditionally described as Yokuts territory as far south as the Tuolumne River. Merriam's extension of the Miwok territory south of the Calaveras River to the Tuolumne River was explicitly refuted in Kroeber's 1908 discussion

of the occupation of these regions by Miwok people. Kroeber's additional work on Yokuts and Bennyhoff's Plains Miwok ethnogeography cast further doubt on the San Joaquin Valley territories claimed for Miwok by Merriam (Kroeber 1907c, 1963; Bennyhoff 1977).

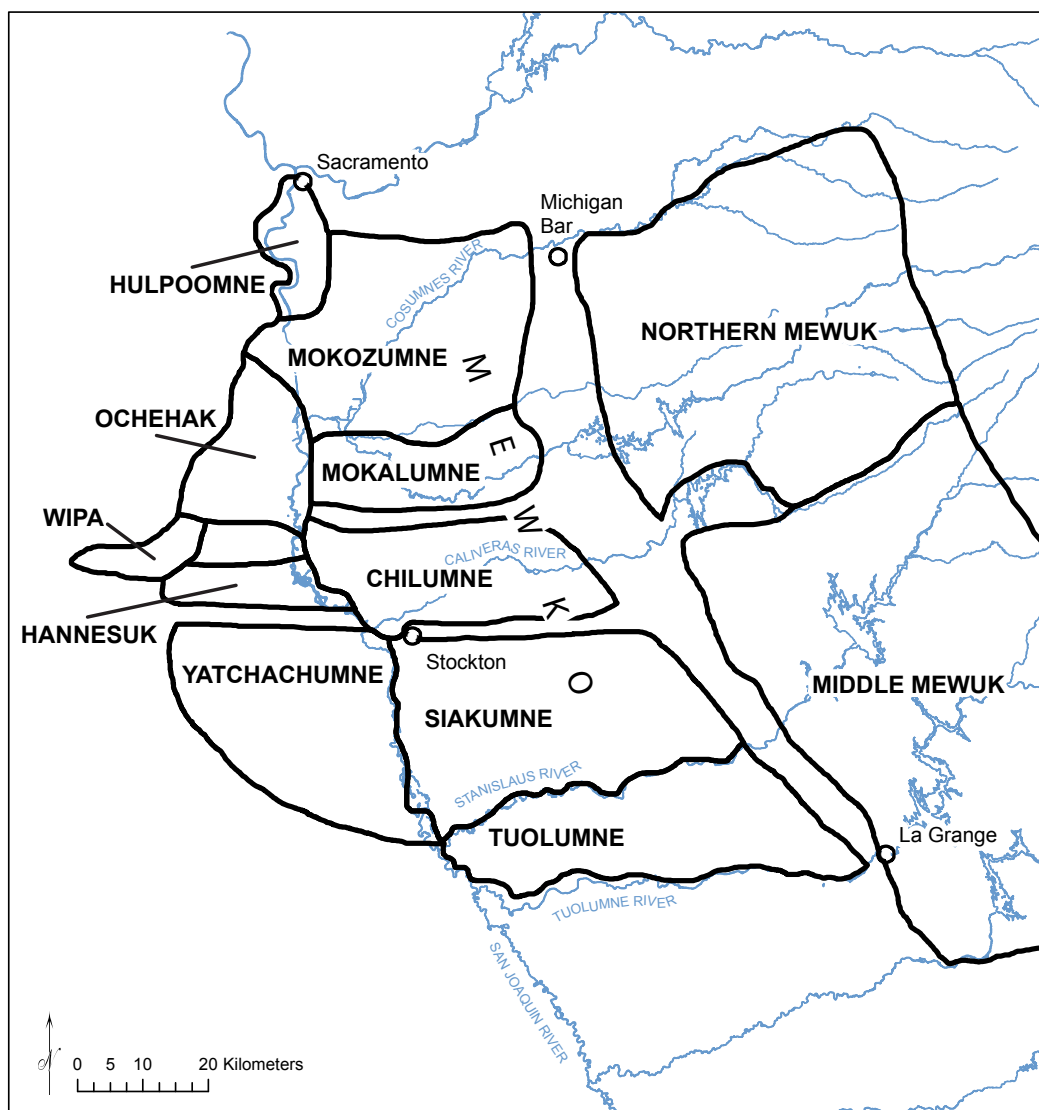


Figure 2.4: Merriam's divisions of the 'Moquelumnan' territory (after Merriam 1907: plate XXV)

Merriam's survey of Miwok language varieties was conducted in conjunction with his work mapping plant and animal species ranges for the U.S. Biological Survey (Merriam 1904), and accordingly his descriptions of the Miwok territories provide some interesting information regarding the ecological features that correspond to the group's geographic range. For the Sierra Miwok groups, Merriam describes their general settlement pattern as covering a 30 mile wide swath along the edge of the foothills, parallel to the trend of the mountains. He notes that their settlements extend 15 or 20 miles up some river valleys, clustered closely along the river channels in these high elevation extensions (Merriam 1907:342). The three primary

divisions of the Sierra Miwok, called ‘Northern’, ‘Middle’, and ‘Southern’ by Merriam, are linked to broadly defined ranges. Northern Mewuk is described as ranging from the Middle Fork of the Cosumnes River to some point slightly south of Calaveras Creek. From an eastern boundary which runs between Grizzly Flat, Big Trees, West Point, and Railroad Flat, the territory extends to the western towns of Michigan Bar, Ione, Buena Vista, Lancha Plain, and Comanche. This downhill boundary, Merriam notes, corresponds with the edge of the range of California foothill pine and blue oak species (Merriam 1907:344). The territory of Central Miwok is described as ranging from Calaveras Creek in the north to the Tuolumne River in the south, extending to the east beyond Hetch-hetchy Valley. The western boundary of this group is described as crossing the Tuolumne River between Jenny Lind and La Grange. From the southern boundary of Central Miwok on the Tuolumne River, the Southern group’s territory is said to range southward to Fresno Creek, following the Merced River through the Yosemite Valley on its east side, and extending to the town of Raymond in the west.

Of the Mew’-ko, or Plains Miwok, tribes listed by Merriam, he reports having obtained ‘original information’ on all but the Yatch’-a-chum’-ne (Merriam 1907:348), and judges that the dialects of the groups he contacted personally all belong to a single language. These groups range through the plains from just south of the American River to the Tuolumne River, a territory that includes the islands between the San Joaquin and Sacramento Rivers. The Yatch’-a-chum’-ne, one of the westernmost groups and one to which Merriam ascribes a large territory west of the San Joaquin River, are located on Merriam’s map on the basis of an authorless article published in 1900, and claims about the linguistic classification of this group are therefore particularly suspect. The eastern boundary for two of the Mew’-ko groups, the Si’-a-kum’-ne and the Tu-ol’-um-ne, is listed as Knight’s Ferry. The general absence of any documentation of Miwok language west of Knights Ferry suggests that this town is on the true western boundary of Eastern Miwok’s geographic range.

Kroeber’s (1906) classification of the Miwok family, like Merriam’s earliest ethnogeographic publication (1904), contains only a brief description of the geographic distribution of this language family. More interesting is his discussion of Miwok geography in the *Handbook of the Indians of California* (1925). In this description he too notes the primary split between Plains and Sierra Miwok. He describes the Northern and Central Sierra Miwok languages as more similar to each other than either is to Southern Sierra Miwok and suggests that some Plains Miwok influence extends into both Northern and Central Sierra Miwok. Though he states that dialects exist within all four of the Eastern Miwok languages, he asserts that all “are rather insignificant and may be disregarded” (Kroeber 1925:442). Kroeber describes the general range of Northern Sierra Miwok as encompassing the drainages of the Mokelumne and Calaveras Rivers, with the Central Sierra Miwok controlling the Stanislaus and Tuolumne Rivers, and the Southern Sierra Miwok settling within the Merced basin and along smaller streams in that vicinity (see Figure 2.5). Kroeber notably discusses the controversy regarding the northern boundaries of the Northern Sierra and Plains Miwok, and the contradictory reports of the Plains Miwok boundaries, the latter group having been reduced almost to extinction by the time ethnogeographic work was begun. Ultimately he concludes that Plains Miwok was spoken only in the north, with southerly parts of the San Joaquin Valley controlled by the Yokuts.

The other notable account of Eastern Miwok geographic coverage is Barrett’s (1908a)

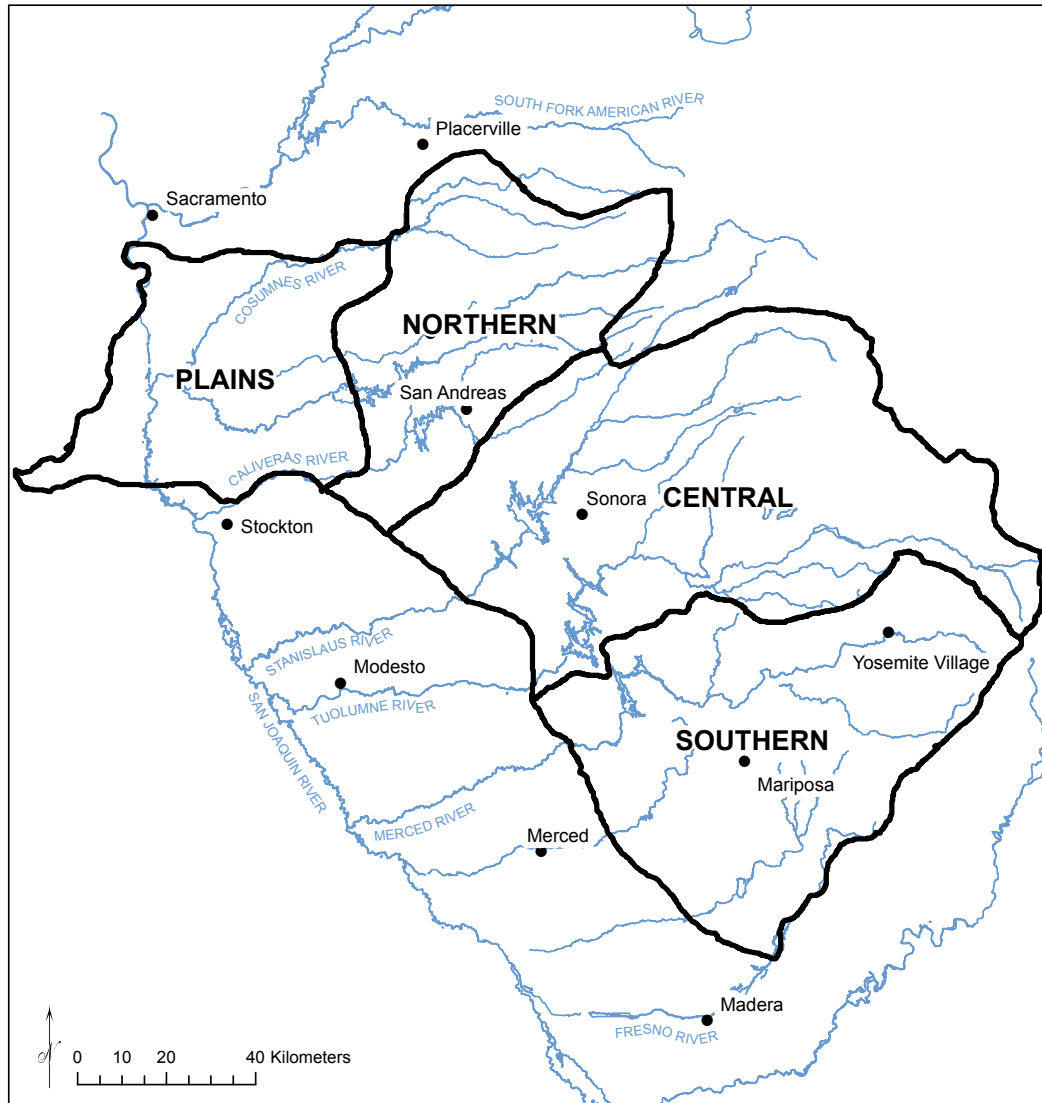


Figure 2.5: Kroeber's Miwok geography (after Kroeber 1925: plate 37)

survey of the Miwok dialects. Barrett describes the general range of Eastern Miwok as extending north to south from the Cosumnes River to the Fresno River and east to west from the crest of the Sierra Nevada to the edge of the San Joaquin Valley plain, extending in Plains Miwok territory out through the Calaveras drainage into the plain (Barrett 1908a:335). He identifies Sacramento, Amador, Calaveras, San Joaquin, Stanislaus, Tuolumne, Mariposa, Merced, and Madera counties as the political units in which this territory lies, and the Cosumnes, Mokelumne, Calaveras, Stanislaus, Tuolumne, Merced, and Fresno Rivers as the drainages in which the Eastern Miwok people settled (Barrett 1908a:335). Like Merriam, Barrett draws some conclusions about the physical environment in which this group chose to settle, describing the high Sierra and some of the valleys of moderate elevation as uninhabit-

able in the winter. He notes, though, that these regions were used by the Indian populations as sources of foods, including pine and oak trees and several animal species. Notably, Barrett observes that the transition between the Sierra foothills and the San Joaquin Valley plains is most abrupt in the southern extent of this territory — i.e. where it serves as a boundary between Miwok and Yokuts — and more gradual in the north, where the Plains and Sierra Miwok territories meet (Barrett 1908a:337). With regard to climate, Barrett claims that the temperature ranges of the San Joaquin Valley plains and the lower foothills are quite similar.

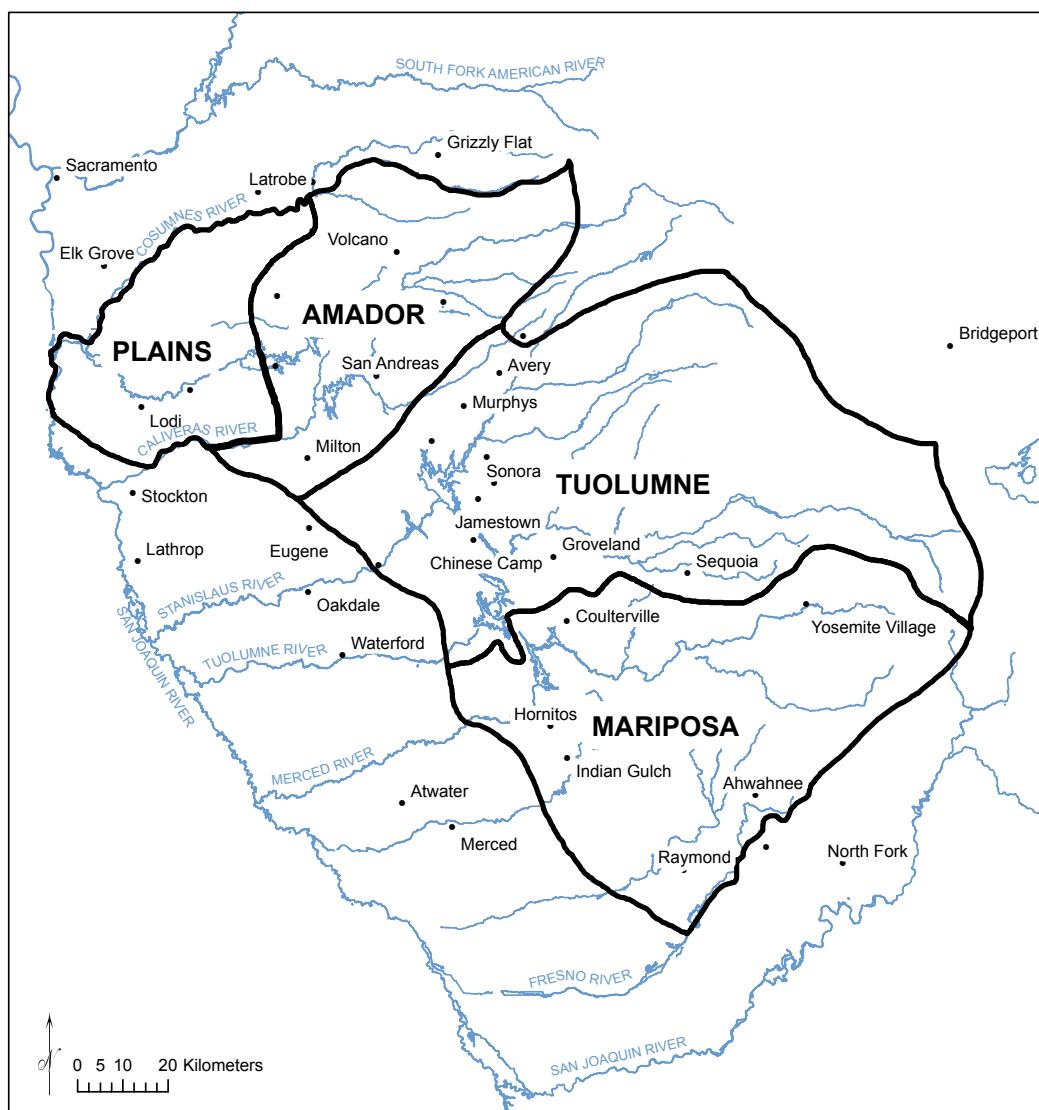


Figure 2.6: Barrett's Miwok geography (after Barrett 1908a: 369)

In describing the cultural divisions within the Eastern Miwok area, Barrett describes two general cultural zones, which extend into neighboring languages. These two divisions — a northern and a southern area — are defined primarily in terms of material culture, and specifically by their acorn-mush stirring tools, baby cradle styles, and the formation of baskets on grass versus rod frames (Barrett 1908a:338-339). He places an admittedly

indistinct boundary for these two culture areas between the Stanislaus and Tuolumne Rivers. The northern culture area corresponds to more general Northern California types, with paddle stirrers for acorn mush like the Achomawi or Pomo and cradle styles that are also similar to Northern California cultures far to the north and east of Miwok territory (Kroeber 1925:447). In the southern area, Yokuts-type basket shapes, cradles, and looped-stick mush stirrers suggest the borrowing of material culture elements from these Central California neighbors (Kroeber 1925).

Barrett's exploration of the territories of the individual language groups — the Northern Sierra Miwok, Central Sierra Miwok, Southern Sierra Miwok, and Plains Miwok — contains more detail than other published sources, evidently constructed in part from several consultants' descriptions of boundaries (see Figure 2.6). With the exception of the discrepancy between Barrett's boundaries for Plains Miwok and those put forth by Merriam, these detailed boundary descriptions are generally consistent with other sources, especially Kroeber (1925). Barrett notes explicitly that the western boundary of Miwok, where it abuts the Yokuts territory, is controversial among ethnographers and Miwok consultants alike. It is likely that this border was blurred by bilingualism and close cultural contact. The conversion of the San Joaquin Valley to agricultural use early in the history of California makes it even more difficult to determine whether the Miwok territory ever extended into the plains. For the current study, the lack of any linguistic data from the disputed portions of the San Joaquin Valley creates an effective western boundary for Sierra Miwok that is consistent with Barrett and Kroeber's designations — that is, the western edge of the foothills.

Plains Miwok was contacted as early as 1806 by Spanish explorers led by Moraga. This linguistic group was affected early and profoundly by the missionary movement in California. As a result, the Plains Miwok inhabitants that remained in the region when scholars like Merriam, Barrett, and Kroeber began their research occupied a reduced territory and were likely displaced from the settlement sites that existed before missionary and gold rush contact. Bennyhoff, unsatisfied with the imprecise and conflicting interpretations of the Plains Miwok boundaries that had resulted from work with post-mission period consultants, turned to the records of the early explorers and historians and archaeological evidence to create a more accurate description of the Plains Miwok geography (Bennyhoff 1977). Bennyhoff provides a detailed description of the plains, marshland, and river channels that comprised the general region of Plains Miwok settlement. Important features in this landscape are the Victor Alluvial Plain, a strip in the center of the Plains Miwok territory covered in grasses and occasional oak trees; the Arroyo Seco Pediment on the eastern side of the territory with its low, rolling hills; the delta marshlands and tidal plains; and the riparian zones around the river channels themselves, particularly the Sacramento River. Bennyhoff provides elaborate descriptions of each of the Plains Miwok sites for which there is documentary evidence, including such details as location, marriage records, a summary of missionization, and extensive notes on the confusions that result from inconsistent descriptions of these sites by various researchers. Unfortunately, most of the sites throughout Plains Miwok territory that were painstakingly researched by Bennyhoff cannot be investigated from a linguistic perspective, since their inhabitants were removed to missions or otherwise departed from the region before their language could be documented.

The internal diversity of Eastern Miwok has been discussed by many of the linguists that

have done documentary work on its component languages. From the sum of these works a picture emerges of relatively even dialect diversity throughout the Eastern Miwok territory — multiple dialects, that is, within each of the four languages. In general, however, the dialect information offered in published works focuses on only the language or languages that were central to the dictionary or grammar writer’s descriptive work, and none of the Miwok scholars claims to have exhaustively catalogued the dialect divisions within their language of focus. The network of dialects that actually existed in the Plains and Sierra Miwok languages is therefore illuminated in a very spotty fashion. This general sense of moderate internal differentiation within each of the Eastern Miwok languages cannot be translated into anything that describes the dialects of Eastern Miwok and the relationships between them without further work.

Barrett’s (1908a) discussion of the classification of the Miwok languages, while focused on language-level rather than dialect-level differentiation, pays close attention to the issue of internal diversity. A key component of his discussion of the relationships between the Eastern Miwok languages is a discussion of percentage of stems from a standard word list held in common between languages. Of the four languages, Plains Miwok stands out as the most distantly related, with only 60 percent of the word list items cognate to forms in any of the other languages (Barrett 1908a:357). Northern and Central Sierra Miwok, or the “Amador” and “Tuolumne” dialects as he calls them, appear to be far more closely related, sharing 80 percent of the stems in the list. Figures for the relationship of Southern Sierra Miwok to the other languages are not listed in the article. Barrett (1908a), Merriam (1907), and Kroeber (1906) all make vague comments about the probable existence of sub-dialects, but these sub-dialects are all dismissed by the authors as being inconsequential. The convention of classifying the family as four distinct languages seems to be based on the findings of Barrett (1908a) and Kroeber (1906), and with so little work done to understand variation among the regional varieties of Eastern Miwok it is possible that this convention glosses over a more complex picture of dialect variation.

Each of the writers of the mid-20th century dictionaries of Eastern Miwok adheres to the standard four-language internal classification of Eastern Miwok and discusses dialect variation within these languages primarily in terms of lexical differences and some sound correspondences. Broadbent, for example, notes some divergence between forms from Yosemite, Mariposa, and the forms produced by individuals from the southernmost region of Southern Sierra Miwok territory (1960). Broadbent gives few examples of the features that distinguish the Southern Sierra Miwok dialects, mostly revolving around the spirant series of phonemes. This makes it even more puzzling that, as she reports, Mariposa speakers claim to understand Yosemite speakers only with great difficulty (Broadbent 1960:8). Callaghan makes similar statements about dialect diversity in her *Plains Miwok Dictionary* (1984) and *Northern Sierra Miwok Dictionary* (1987). For Plains Miwok she notes some variation between Lockeford and Jackson Valley speakers, and for Northern Sierra Miwok she draws a distinction between Camanche, Fiddletown, and Ione dialects. These dialect level distinctions are not supported by discussion of the phonological differences between the varieties, but forms that appear in the dictionaries and occur in only one of the dialects are marked for their dialect of origin in the dictionary entries. Like Callaghan, Freeland and Broadbent note the existence of multiple dialects, and also like Callaghan, they use a special notation system to

alert readers to dialect-specific forms in the dictionary. However, they are also equally silent on the issue of phonological correspondences between the dialects.

Discussion of dialect diversity is not limited to the academic literature. Members of the small community of contemporary speakers and semi-speakers of Sierra Miwok languages have commented during language revitalization meetings in 2008 through 2010 about the forms used in other towns, particularly towns in the Northern Sierra Miwok-Central Sierra Miwok boundary area. Members of the Northern Sierra Miwok and Central Sierra Miwok heritage communities who maintain some limited knowledge of the language find the issue of dialects somewhat confounding, as their recollected pronunciations do not always mirror the documented forms for the four main languages, and their tribal affiliations and linguistic lineages do not correspond neatly. It is quite possible that the issue of dialect diversity has become more muddled in the late 20th century as the last fluent native speakers have moved between tribal communities. However, as complicated as it is, the contemporary language situation seems likely to have arisen from a complex dialect network within Sierra Miwok.

In summary, the Eastern Miwok language was historically spoken in a region extending from the Cosumnes River to Fresno Creek. In pre-contact times, the population density of Sierra Miwok was probably fairly even, but the Plains Miwok area was settled more densely. Depopulation and forced removal devastated the populations in the northwestern end of Eastern Miwok's territory and had a lesser effect on populations toward the eastern and southern edges of the territory. The classification of the languages in this region has been fairly stable throughout the history of study, and splits the family up into 5 languages: Saclan, Plains Miwok, Northern Sierra Miwok, Central Sierra Miwok, and Southern Sierra Miwok. Saclan is occasionally omitted from classifications because it was obliterated early by Spanish settlement and little documentation remains. An unknown amount of smaller scale dialect diversity is recognized but not analyzed by most of the scholars who have worked on the family. Finally, Eastern Miwok culture shows influence from two areas: the Northern California culture area and the Central California culture area. A divide in material culture items shows that these two culture patterns meet between the Tuolumne and Stanislaus rivers.

2.1.2 Archival Resources

Archival resources are of vital importance in making sense of the dialect diversity within Eastern Miwok that existed in the early twentieth century. The disappearance of dialects that has accompanied the 20th century shift to English dominance leaves little contemporary evidence of the historical dialect geography within this language family. While there are certainly limitations to the archival description of Eastern Miwok dialects, these data nevertheless provide the best available picture of the linguistic diversity within this family.

In all, the archival documentation of Eastern Miwok spans roughly a century, from 1877 (Powers 1877) to the late 20th century (Cross 1982). Within this window the bulk of the materials were gathered between 1900 and 1925, with a second wave of documentation in the 1950s and 1960s. Focusing on these two periods of productive and geographically extensive documentation of Eastern Miwok, this study uses data collected between 1900 and 1982, with

the bulk of the data gathered in the first decade of the 20th century.³ The geographic coverage of this documentation extends to all corners of the Sierra Miwok settlement area, and word list data sufficient for this study are available for 25 different locations within Eastern Miwok. The greatest deficiency in this sample is its coverage of Plains Miwok. Although the Plains Miwok territory extends westward to the mouth of the Sacramento River, the geographic location from which data was available (Ione) is located at the eastern edge of Plains Miwok territory. The distribution of the current sample reflects the depopulation of the western portions of Plains Miwok territory and the entire range of Saclan early in history, preventing the documentation and study of most western Plains and Saclan dialects.

The data for the current analysis are word lists that can be mapped to specific geographic locations. In most cases, biographical information about informants was included in researchers' field notes. Barrett's notes contain detailed locations and brief biographical sketches for many of his consultants, and Kroeber and Broadbent provide similarly detailed information about their consultants. Word lists collected from consultants for whom this information was available were mapped to the location deemed most linguistically relevant for those consultants. For most word lists, birthplaces or places where a consultant was raised or simply "from" were selected as the relevant location. Word lists accompanied by little or no information about the consultant were mapped to the collection location. Several of the datasets from Dixon, Merriam, and Tozzer fall in this category. The duplication of data collection at most of these sites is expected to offset any error introduced by using collection locations as linguistically relevant locations. The exception to this is Sonora, a location for which the only available data was collected by Dixon from an unnamed consultant about whom nothing more is known.

The researchers' transcription conventions vary as much as their biographical data collection. While systematic differences in the orthographies used by these researchers do not pose any problem for analysis, variation in the amount of detail recorded is more likely to interfere with comparative analysis. In particular, variable sensitivity in the perception of vowel quality, coronal consonants, and segment length are likely to interfere with measures of phonological dissimilarity. Researchers' ability to hear and record differences in vowels (particularly the high vowels *y* and *u*) and consonants (especially the "front" and "back" *t* and *s* sounds) seems to vary. Littlejohn, Merriam, and Tozzer, in particular, do not indicate any differences between the "front" and "back" consonants, and across all researchers there is a moderate amount of vowel quality variation that does not appear to be dialect-based and is probably unavoidable noise in a signal recorded by researchers with varying auditory sensitivity and different assumptions about Miwok phonology. Similar noise exists in the

³Several important collections of Eastern Miwok documentary materials were not included in this study. Lucy S. Freeland's notes from the 1920s and 1930s primarily focused on Central Sierra Miwok, while undoubtedly a rich resource, are missing. Records indicate that these notes were located at the Survey of California and Other Indian Languages in the late 1950s when Broadbent used them to compile the *Central Sierra Miwok Dictionary with Texts* (Freeland and Broadbent 1960), but they were never catalogued by the Survey or any other archive. It is possible that they are contained in the large portion of the Jaime de Angulo collection at the University of California Santa Cruz that is currently unprocessed and inaccessible. Only a small subset of Freeland's notes is currently available at that UC Santa Cruz's Special Collections Library. Other collections such as the Plains and Northern Sierra Miwok notes gathered by Catherine Callaghan in the mid-twentieth century and Suzanne Wash in the late-twentieth century are also currently unavailable.

Table 2.1: Eastern Miwok documentation by location

Location	Date	Consultants	Researchers
Ahwahnee	1906	2 M	Barrett
Angels Camp	1904	1 M	Kroeber
Avery	1906	1 M, 1 F	Kroeber
Bald Rock	1906	1 M	Barrett
Big Valley	1906	1 M	Barrett
Buena Vista	1903, 1906	1 M	Merriam, Barrett
Bull Creek	1906	1 M	Barrett
Camanche	1960	1 F	Callaghan
Chicken Ranch	1906	1 M	Barrett
Elliott Corner	1957	2 F	Broadbent
Groveland	1906, ca. 1922	1M, 3 F	Barrett, Freeland
Ione	1958	1 M	Callaghan
Jackson	1900, 1906	1 F, 1 ?	Tozzer, Barrett
Jamestown	ca. 1922	1 M	Freeland
Knights Landing	1906	2 M	Barrett
Lockeford	1903	1 M	Merriam
Mariposa	1906, 1957	1 M, 2 F	Barrett, Broadbent
Merced Falls	1906	1 M	Barrett
Murphys	1906	1 F	Barrett
Pleasant Valley	1919, ca. 1925	1 M, 1 ?	Merriam, Littlejohn
Sonora	1903	1 ?	Dixon
Tuolumne	1919, 1982	1 F, 1 ?	Merriam, Cross and Shipley
Wawona	1906	1 M, 1F	Barrett
West Point	1903, 1906, 1908	1 M, 1 ?	Merriam, Barrett, Kroeber
Yosemite Village	1904	1 M, 1F	Kroeber

transcription of segment length. All of the researchers whose notes were used for this project recorded consonant length at least some of the time, but the consistency with which they did so is highly variable. Vowel length is even more inconsistently recorded. To minimize the effect of inconsistent transcription of segment length, aggregate measures of phonological dissimilarity were calculated based on the data as recorded and also after discarding information about segment length (i.e. with all sounds represented as short/non-geminate).

2.1.3 Eastern Miwok Dialectology

In spite of this study's use of archival data instead of the contemporary dialect surveys that form the empirical basis for most dialectology work, this investigation of Eastern Miwok dialect geography still embraces the goals and methods that are central to the field of dialectology. For example, among its other goals, this study explores the primary boundaries

between the several Eastern Miwok varieties and investigates the degree to which subregions of Eastern Miwok territory resemble dialect areas as opposed to continua. Where newer methods are used, specifically dialectometry methods for measuring linguistic differences and geographic methods for modeling distance, these are employed to better understand the strength of contact between locations and the impact this has had on dialect patterns. This study probes deeper than most dialect studies into the question of how environmental factors influence linguistic change, yet the fundamental approach to dialect geography is quite traditional.

Bloomfield notes in his discussion of dialectology that the strength of contact between locations is reflected in the isogloss networks that occur between these locations (1933:341). Within a more detailed discussion of dialect geography, he draws attention to two notable patterns: dialect areas defined by dense bundles of isoglosses, and continua which arise through cumulative changes between two distant points. These types of spatial patterns in dialectology have continued to be important in the study of dialect geography through the twentieth century and into the era of quantitative dialect studies (e.g. Heeringa and Nerbonne 2005). Both of these patterns can result from the process of dialect leveling. Dialect areas represent the canonical outcome of dialect leveling, where the spread of some set of features throughout a geographic area is essentially complete, with relic forms or other dialect forms found outside that area. Continua generally also represent the spread of linguistic features. However, continua arise when features spread to different spatial extents or when the “lexical diffusion” of sound changes is more complete near the source of a change than in the periphery of its spread (Bloomfield 1933:328; Chambers and Trudgill 1980:160).

The opposition between dialect areas and dialect continua is often discussed in terms of the “traveler puzzle” (Chambers and Trudgill 1980; Heeringa and Nerbonne 2005). Would a traveler passing through the region notice abrupt transitions between speech varieties on his journey, or would he find, as he travels from each locale to the next, a rather similar new variety in each place yet a very different variety at his destination than the speech at his point of origin? The descriptions of the major boundaries published by Kroeber, Barrett, and Merriam would seem to provide some anecdotal evidence that these researchers, in their travels, encountered perceptually salient boundaries (Merriam 1907; Barrett 1908a; Kroeber 1925), but these representations of Eastern Miwok do not paint a complete picture of the dialect geography.

Several questions arise when looking at Eastern Miwok more closely through a dialectological lens. First, how well defined are the traditional geographic boundaries proposed by scholars like Kroeber and Merriam, and do the boundaries of early maps accurately reflect the locations of linguistic divisions? Secondly, do the patterns revealed by isogloss mapping and distance metrics allow us to identify areas of leveling, boundaries to feature diffusion, and more subtle effects of social network structure on dialect feature spread? Finally, do the environmental variables that we expect to exert influence over human contact help us to account for the linguistic variation in this region? Answering these questions requires an approach that is adapted to the available data and to what we know about the circumstances of Eastern Miwok linguistic development.

The traditional procedure of mapping the isolines that delineate the spatial extents of linguistic features is of some use in this investigation. The emergence of boundaries in the

form of isogloss bundles illuminates the spatial extent of distinct dialect areas. However, the interpretation of more subtle and diffuse isogloss patterns is difficult in a language group like Eastern Miwok. Bloomfield describes a process of distinguishing “great isoglosses” from “petty” isoglosses, with the former representing the boundaries of large geographic areas and being generally associated with “large event[s] in linguistic history” (Bloomfield 1933:341). However, the formative events in Eastern Miwok linguistic history and social history more generally are not well understood, and the limited data we have includes many isoglosses that appear to divide the region into broad areas. Interpretation of isogloss patterns is made even more difficult by the nature of the network of Eastern Miwok settlements. We might expect a variable pattern of linguistic contact between locations, but we cannot necessarily identify centers of social prestige which might be associated with these large historical linguistic shifts.

Not only do Eastern Miwok archival data differ from typical dialect survey data in the ways mentioned in Section 2.1.2, but the social and political organization of Eastern Miwok speaker communities was also historically quite different from the subjects of most dialect geography work. Whereas the European languages that were central to the development of dialectology are spoken in large populations with great urban centers, agricultural hinterlands, and a good deal of political organization, Eastern Miwok was spoken in a relatively small hunter gatherer population with little hierarchical organization to its social and political systems and no evidence of pre-contact urban centers. The social importance of various nodes in a network of European settlements may be estimated based on demographic or historical facts, but centers of Eastern Miwok culture are more difficult to identify. We do know that Eastern Miwok settlement was organized in very small patrilineal groups or *nenas*⁴ that maintained permanent settlements at ancestral sites but shared surrounding resource areas with other small settlements (Gifford 1926; Barrett and Gifford 1933). A system of two Miwok moieties is also well documented (Gifford 1916), and though this system required exogamous marriage practices, exogamy would certainly have been practiced by the patrilineal settlement groups of Eastern Miwok anyway due to the small size and familial relationships of these communities. In addition to women moving to their husbands’ ancestral settlement sites for marriage, contact between members of different moieties in nearby settlements seems to have been fairly frequent. For example, the *pota* ceremony described by Gifford necessarily included neighbors from outside the hosts’ moiety (1926:390).

Gifford describes the political hierarchy of these small Miwok communities as fairly flat, with small settlements lead by a “captain” or patriarch and each community’s identity and governance determined quite independently. Several scholars note that rather than identifying other Miwok communities as co-members of some larger social group, most Eastern Miwok people identified with their familial group and viewed neighbors essentially as separate communities that happened to speak the same language (Merriam 1904; Barrett 1908a; Gifford 1926). Levy’s description of the Eastern Miwok culture in the Smithsonian Handbook describes slightly larger “tribelets” composed of multiple lineage groups (Levy 1978:398) and Merriam describes a system of principal villages and tributary villages (Merriam 1907:342). Gifford casts some doubt on whether these larger groupings existed before

⁴*Nena* is a Miwok term for familial lineage.

Euroamerican contact, claiming that the emergence of “main” villages and multi-settlement political groups happened only in response to pressure from Spanish settlers and gold rush miners (1926). This view is supported by Maniery’s anthropological analysis of settlement at Murphys Rancheria, a site to which local Miwok people resettled after Euroamerican land use made nearby historical settlement sites inhospitable (Maniery 1983, 1987).

Regardless of whether the small settlements historically grouped into larger political units, we do know that Eastern Miwok permanent settlement sites were clustered in areas with access to water, frequently along river valleys. At least some Eastern Miwok people migrated with their lineage-based communities to higher elevations in the summer to hunt, forage, and trade with Mono or Northern Paiute speakers (Barrett and Gifford 1933; Levy 1978; Chartkoff 1990; Montague 2010). Much of the literature on temporary summer migrations is based on archaeological evidence of temporary camps at high elevations (Chartkoff 1990; Montague 2010), and it is not entirely clear whether Eastern Miwok communities were uniformly involved in seasonal migration or whether these summer migrations were undertaken by only some subset of Miwok communities, although Gifford suggests that seasonal migration was nearly universal (Gifford 1926:391). Seasonal trading with Northern Paiute and Mono speakers from the Eastern Sierra brought salt, obsidian, rabbit skin blankets, and pine nuts into Miwok communities. In return Miwok people traded acorns and certain types of seeds, as well as *Olivella* and *haliotis* shells obtained through trade with coastal Ohlone groups (Barrett and Gifford 1933; Davis 1974; Levy 1978; Chartkoff 1990).

The limitations imposed by the absence of historical records predating the arrival of Euroamericans and the loose political associations among settlements make Eastern Miwok a good candidate for the use of quantitative methods for identifying important patterns in dialect geography. As Nerbonne notes, aggregating large sets of linguistic data and using quantitative methods to analyze trends in this data allows us to characterize spatial variation in language without succumbing to the difficulties of interpreting non-bundled isoglosses or ignoring complexities and exceptions in feature distributions (2009). The solution to the problem of studying dialect geography in a language group whose historical and linguistic documentation are limited is to supplement the identification of isogloss bundles with a quantitative analysis of aggregate linguistic differences and to use geographic analysis to explore the dynamics of contact.

2.2 Eastern Miwok Dialect Features

Several Eastern Miwok dialect features have been noted in earlier publications, either as features of the “languages” that make up this family or as regional variants. Catherine Callaghan has used data from the four primary varieties of Eastern Miwok that are well documented to reconstruct lexical and grammatical forms for Proto-Eastern Miwok and Proto-Miwok (Callaghan 1979a,b, 1985, 1994, 2003), and together with Broadbent she reconstructed the phonological systems of these ancestor languages (Callaghan and Broadbent 1960). Callaghan and Broadbent reconstruct a six-vowel system (**i, e, a, o, u, y*) in Proto-Miwok that remained unchanged in Proto-Eastern Miwok and Proto-Sierra Miwok. The only vowel change they propose in Eastern Miwok is a change of **y* to *ə* in Plains Miwok when

stressed or when the stressed vowel is ə , remaining y elsewhere. The consonant inventory for Proto-Miwok, Proto-Eastern Miwok, and Proto-Sierra Miwok is similarly stable as reconstructed by Callaghan and Broadbent. They propose changes to only two of the reconstructed sibilant consonants ($*\text{š}$ and $*\text{ʒ}$) in the daughter languages of Eastern Miwok ($*\text{š}$ reflexes: M_P s , except h morpheme-finally; M_NS ʒ ; M_CS ʒ ; M_SS h ; $*\text{ʒ}$ reflexes: M_P s , h ; M_NS ʒ ; M_CS s ; M_SS s) and a change from Proto-Eastern Miwok $*\eta$ to n in just Plains Miwok. The phonological differentiation that Broadbent and Callaghan describe within Eastern Miwok is minimal, but the reflexes of the reconstructed sibilant consonants, the $*y$ vowel, and other regionally variable sounds will be discussed below in Section 2.2.1.

Table 2.2: Sound changes proposed in Callaghan and Broadbent’s reconstruction of Proto-Miwok (1960:302)

Proto-Miwok	Proto-Eastern Miwok	M_P	M_NS	M_CS	M_SS
$*\text{š}$	$*\text{š}$	s , h	ʒ	ʒ	h
$*\text{ʒ}$	$*\text{š}$	s , h	ʒ	s	s
$*\eta$	$*\eta$	n	η	η	η
$*y$	$*y$	y , ə	y	y	y

There is considerably more variation in Eastern Miwok lexical items. Some of the meanings for which lexical differences occur at the regional level, among speakers of varieties that would fall under the same traditional language classification (e.g. Northern Sierra Miwok, Central Sierra Miwok, etc.), have been noted in published dictionaries (Freeland and Broadbent 1960; Broadbent 1964; Callaghan 1984, 1987). Many of these meanings did not appear in a large enough number of location-linked word lists to be used in this study. However, even among meanings that appear on standard word lists there is a significant amount of lexical variation. These regional lexical differences are discussed in Section 2.2.2 below.

2.2.1 Phonological Changes

The phonological changes that occur within Eastern Miwok are few, and the patterns they exhibit are somewhat difficult to interpret due to the unknown amount of error in the source data. One phonological element that is unlikely to be misheard or mis-transcribed is the contrast between n and η . Broadbent and Callaghan reconstruct a velar nasal in Proto-Miwok and Proto-Eastern Miwok, which remains unchanged in the Sierra Miwok varieties (1960). They note that this sound has changed to an alveolar nasal in Plains Miwok, and Ione, the sole Plains Miwok variety represented in this study, shows that this change has occurred uniformly. In locations surrounding Ione we find some variation between η and n across lexical items. As shown in Figure 2.7, we see a very limited pattern of lexical diffusion, with n occurring in the forms for *egg* and *knee* in the location nearest Ione, and in the form for *egg* but not *knee* in Lockeford, another nearby location. In varieties farther from Ione, η occurs without exception in all forms which historically contained the velar nasal. The change to n has not permeated far into Sierra Miwok, but the pattern of diffusion is one of item-by-item change in the lexicon.

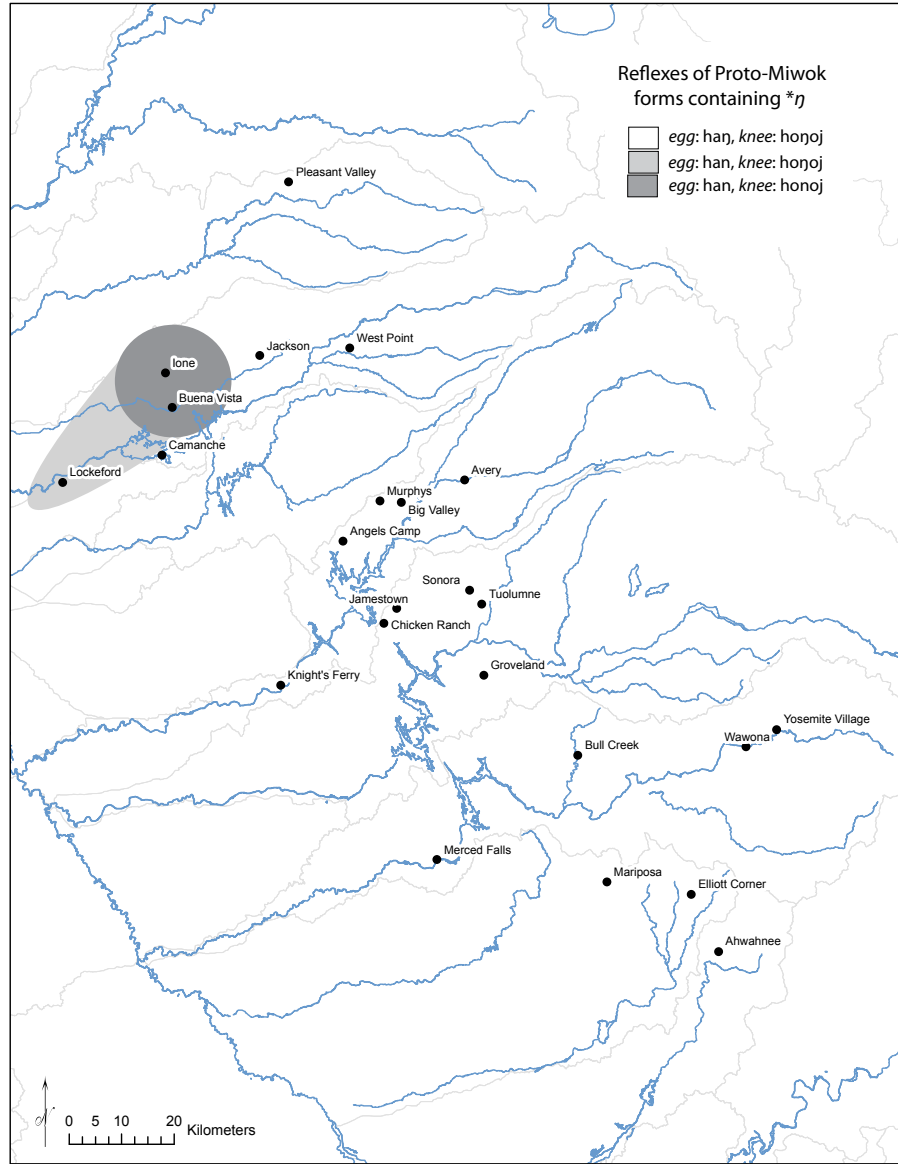


Figure 2.7: Diffusion of $*\eta \rightarrow n$ change in Eastern Miwok

Of the three sibilant sounds Broadbent and Callaghan reconstruct for Proto-Eastern Miwok, only one of them shows much variation in the reflexes that occur in individual Eastern Miwok varieties. Isoglosses representing the reflexes of Proto-Eastern Miwok $*\check{s}$ in this study's sample varieties are shown in Figure 2.8. As Broadbent and Callaghan would predict, the varieties that are within the traditional Southern Sierra Miwok area contain h in these words and the Plains Miwok variety and one location near to it contain either s or h . Between these two areas we generally find a fricative that is transcribed as s , with the exception of a few pockets where $\check{s}$ is found and Groveland, where both s and $\check{s}$ occur. The Southern

Sierra Miwok change to *h* has also permeated to some items. The isoglosses here correspond generally with the pattern identified by Broadbent and Callaghan. The occurrence of *s* and sometimes *š* where Broadbent and Callaghan would predict *š* introduces the question of how documentary scholars perceived and transcribed these sibilant sounds. Because of the uncertainty regarding the transcription of sibilants it is impossible to determine with any confidence whether the pockets where we find *š* represent true linguistic differences or simply perceptual or transcription inconsistencies.

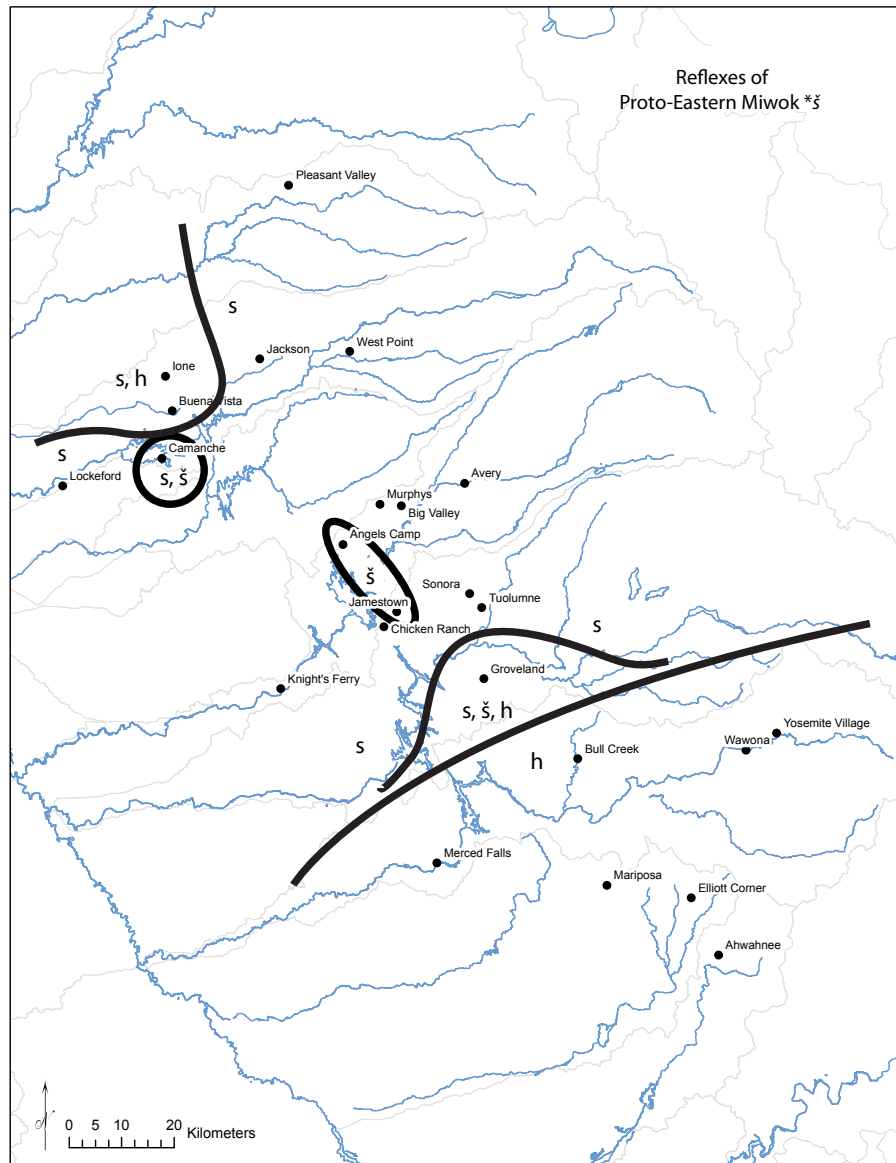


Figure 2.8: Reflexes of Proto-Eastern Miwok *š

The reconstructed Proto-Eastern Miwok sound **y* is claimed by Broadbent and Callaghan to have been retained unchanged in all of the Eastern Miwok varieties. However we find considerable variation in the reflexes of this proto-sound in the varieties investigated in this

study. Like the sibilants discussed above, the distinction between a high central vowel and a high back vowel and the influence of English orthography on the transcription of these sounds have the potential to create inconsistencies in the word list data. However, in spite of the unknown amount of error in this data, there is evidence that a change from **y* to *u* has begun in some locations. Figure 2.9 maps the occurrences of *y* and *u* in forms for which Broadbent and Callaghan have reconstructed the high central vowel in Proto-Eastern Miwok.

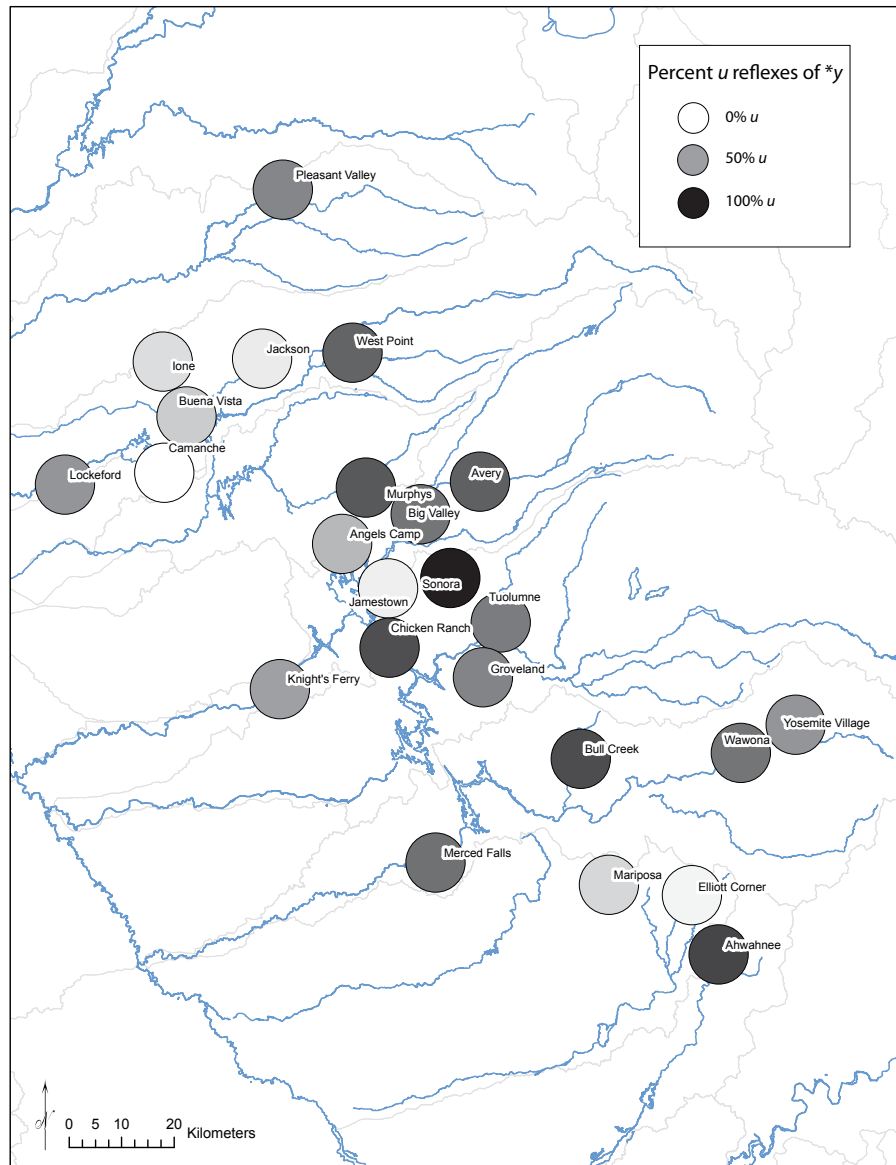


Figure 2.9: Percent occurrence of *u* as a reflex of Proto-Eastern Miwok **y* in 24 forms, represented by percent gray

Though the pattern is not clear cut, we find the vowel *u* more than *y* in most of the easternmost locations, with western locations retaining *y* in a greater percentage of forms.

This pattern is more pronounced at the northern end of the study area, and seems to suggest that the change from *y* to *u* originated in some upper-elevation location and then spread principally through sites at similar elevations, especially in Northern and Central Sierra Miwok regions.

2.2.2 Lexical Changes

The word list data includes 99 items for which multiple non-cognate forms occur. Major isogloss bundles are represented in Figure 2.10, which represents basic patterns in the data schematically, using lines of different weights. The isoglosses representing this full picture of lexical variation are presented through the maps in Appendix A.

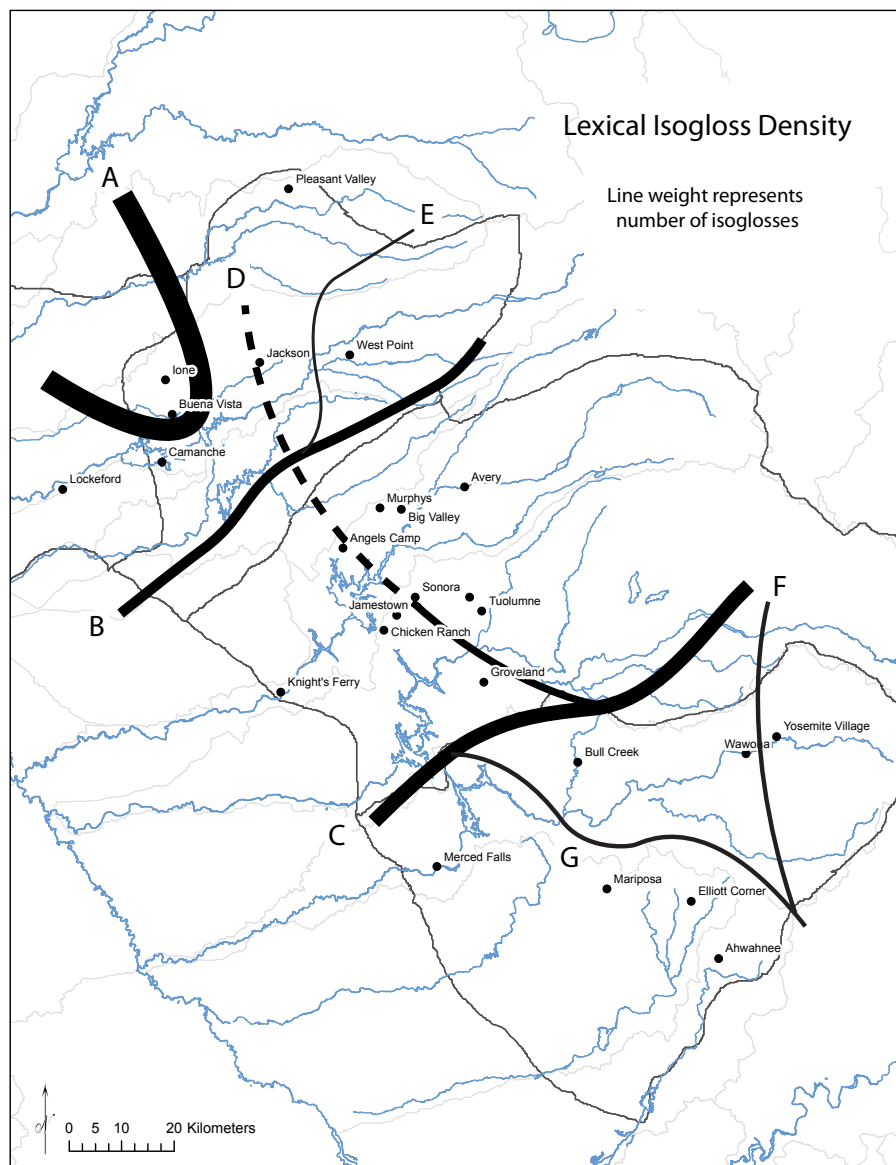


Figure 2.10: Lexical isogloss density in Eastern Miwok

Although the data set used for this dialect study includes only one location that is identified as Plains Miwok, isogloss mapping shows significant support for the claim that Plains Miwok is the most divergent of the Eastern Miwok languages. Thirty-six isoglosses are associated with forms found only in Ione and Buena Vista. Another six isoglosses fall near this bundle, with the Plains Miwok form also found in Camanche, Lockeford, Jackson, or more than one of these nearby locations. An additional five forms are found only in Ione. The forms whose isoglosses fall on or near line A are listed in Tables 2.3, 2.4, and 2.5.

Table 2.3: Lexical forms represented by Eastern Miwok isogloss bundle A

Meaning	Northwest Form	Southeast Form
<i>bow</i>	tanukka	kyčča
<i>chief</i>	čeka	hajapo
<i>dance</i>	lemma	kalape
<i>ear</i>	šolooŋo	ʔookosy
<i>eat</i>	čammak	ywy
<i>eye</i>	welaj	syntu
<i>friend</i>	ota	sake
<i>girl</i>	umu	ossati / lupu
<i>good</i>	welwel	kyči
<i>hair</i>	ʔolo	hanna / jyse
<i>infant</i>	okii	hikimme
<i>jump</i>	toxsi	tujaany
<i>kill</i>	hetak	jynna
<i>large</i>	teme	ytytti
<i>man</i>	šawse	naŋa
<i>mouth</i>	lype	awwo
<i>old woman</i>	ytyjja	onosso
<i>pestle</i>	hopa	kawaači
<i>rain</i>	hooma	nykka
<i>run</i>	taike	hywwate
<i>sand</i>	huuma	wiskala
<i>see</i>	čiisi	heteeyj
<i>shoot</i>	okne	tukke
<i>sing</i>	hytyk	myyli
<i>smoke</i>	kali	hakiisy
<i>south</i>	jakwit	čymmeč
<i>ten</i>	ekkuke	naača
<i>thunder</i>	lillik	timmelele
<i>up</i>	newit	liile
<i>valley</i>	waʔaača	pyylaju
<i>west</i>	ečaawit	olowwin
<i>white</i>	puʔuʔtu	kelelli
<i>white man</i>	uuten	allini
<i>white oak</i>	siwek	molla
<i>wind</i>	wyllyli	hena
<i>wood</i>	tymaj	sysu

Table 2.4: Lexical forms whose boundaries fall near Eastern Miwok isogloss bundle A

Meaning	West or North Form	East or South Form
<i>bluejay</i>	saisi	tajismy
<i>earth</i>	joṭṭok	walli
<i>fish</i>	puu	lapiisaj
<i>owl</i>	wičiki	tukkuli
<i>quail</i>	nykyte	hekkeekke
<i>yellowjacket</i>	ṣuṣṣu	melḥaj

Table 2.5: Lexical forms found only in Ione

Meaning	Ione Form	West or South Form
<i>green</i>	čititti	čooko
<i>house</i>	yčy	koča
<i>old man</i>	hojja	yjana
<i>small</i>	ititti	iccipitti
<i>tree</i> (generic)	linna	leeka / alawa

The traditional Northern Sierra Miwok – Central Sierra Miwok boundary is also represented by a set of isoglosses, although the bundle here contains fewer items than the aforementioned Plains – Northern Sierra Miwok boundary bundle. The items included in this bundle are listed in Tables 2.6 and 2.7. Although it is not as dense as the Plains Miwok isogloss bundle, this set of differences is substantial, and this boundary should be understood as a major division. However, the certainty with which it has traditionally been referred to as a language boundary must be tempered, as the lexical differentiation between Northern and Central Sierra Miwok is far more modest than differences that set apart Plains and Southern Sierra Miwok. Only fifteen isoglosses fall on or near the bundle represented by line B, compared to 42 for line A and 27 for line C.

As the conventional boundaries would suggest, a number of lexical items are realized by different forms on either side of a boundary that runs between Groveland on the north and Merced Falls, Bull Creek, and Yosemite on the south side. Thirteen forms are associated with isoglosses that fall exactly on the line represented by isogloss bundle C in Figure 2.10, and another 14 isoglosses fall very near this line, but either fall east of Merced Falls or north of Groveland. The forms that contribute to this isogloss bundle are listed in Tables 2.8 and 2.9, and detailed isogloss maps are found in Appendix A.

This isogloss bundle (C) runs roughly between the Tuolumne and Stanislaus Rivers, which is roughly the location where Barrett (1908a) and Kroeber (1925) noted a shift in material culture items like baby baskets and acorn mush stirrers. The sharp divide between Southern Sierra Miwok and its northerly sisters may reflect a situation in which Southern Sierra Miwok speakers were in close contact with Yokuts and Mono neighbors to the south, and thus less reliant on an inter-Miwok contact network for economic and social purposes. This sort of scenario would explain the coinciding linguistic and material culture boundaries that separate Southern Sierra Miwok from Central Sierra Miwok.

Table 2.6: Lexical forms represented by Eastern Miwok isogloss bundle B

Meaning	North Form	South Form
<i>arm</i>	tymmaly	woŋotu
<i>arrow</i>	jačči	paipy
<i>fingers</i>	ukkusu	janisu
<i>friend</i>	sake	moʔe
<i>knife</i>	kuči	kajaji
<i>meat</i>	pičema	hukku
<i>medicine</i>	wene	hykisu
<i>nose</i>	hukku	nito
<i>salmon</i>	tukuunu	kosym
<i>small</i>	iccipitti	tyniči
<i>tree</i>	leeka / alawa	laama
<i>white oak</i>	molla	leeka

Table 2.7: Lexical forms whose boundaries fall near Eastern Miwok isogloss bundle B

Meaning	North Form	South Form
<i>bad</i>	sajje	yssyʔy
<i>neck</i>	toopa	lolla
<i>white man</i>	allini	čjejy

Table 2.8: Lexical forms represented by Eastern Miwok isogloss bundle C

Meaning	Northwest Form	Southeast Form
<i>bow</i>	kyčča / solloky	jawe
<i>doctor</i>	koijapi	tyjuk
<i>green</i>	čooko	čititti
<i>hair</i>	hanna / jyse	hisok
<i>large</i>	yttytti	ojani
<i>mosquito</i>	yjyykysy	čulu
<i>nine</i>	woʔe	eliwa
<i>old man</i>	ujatti	humelečču
<i>rattlesnake</i>	wakkaali	lawwaati
<i>rib</i>	wima	alaka
<i>seven</i>	kenekkaky	titawa
<i>valley</i>	pyylaju	aiji
<i>white</i>	kelelli	pasassi

Table 2.9: Lexical forms whose boundaries fall near isogloss bundle C

Meaning	Northwest Form	Southeast Form
<i>arm</i>	woŋotu	paačan
<i>arrow</i>	paipy	mučkyl
<i>basket</i>	pulakka	alymma
<i>black</i>	kululli	tuhuhi / šatati
<i>door</i>	ukuja	kata
<i>fingers</i>	janisu	laslaski / tissu
<i>good</i>	kyči	čutu
<i>grasshopper</i>	koočo	aŋyt / patati
<i>kill</i>	jynna	johe
<i>knife</i>	kajaji	nokočo / hope
<i>medicine</i>	hysiku	losa
<i>neck</i>	lolla	hypyt
<i>small</i>	tyniči	činimi
<i>wind</i>	hena	kannyma / pukyja

While most of the lexical isoglosses in this region either fall on or near these major boundaries or delineate regionalisms used in one or two locations, one small isogloss bundle cuts across the Central and Northern Sierra Miwok area at an angle roughly perpendicular to the phonological bundle discussed above. These isoglosses, represented by line D in Figure 2.10 run roughly parallel to the crest of the Sierra Nevada, and in close proximity to the ecological boundary between foothill woodlands and montane forest. The occurrence of these lexical isoglosses along this northwest-to-southeast trend and at a location that roughly corresponds to the ecological break between open foothill woodlands and denser montane forests resembles the spatial pattern in the spread of the $y \rightarrow u$ change discussed in Section 2.2.1. The isoglosses represented by line D bundle less tightly, particularly on the northwestern end, but they coalesce in a line running between Tuolumne on the north and Chicken Ranch and Groveland to the south. The forms associated with this division are listed in Table 2.10.

Many other lexical isoglosses are found in the Eastern Miwok study area, including three other small, but cohesive isogloss bundles. The forms associated with these bundles, represented by lines E, F, and G in Figure 2.10, are listed in Tables 2.11, 2.12, and 2.13.

Table 2.10: Lexical forms represented by Eastern Miwok isogloss bundle D

Meaning	West Form	East Form
<i>bluejay</i>	taismy	kaikaja
<i>fish</i>	ewu / hyjyty	kosumu / lapiisaj
<i>hair</i>	jyse	hanna
<i>owl</i>	sukkumi	tukkuli
<i>shoot</i>	kopte / tukke	maʔta

Table 2.11: Lexical forms represented by Eastern Miwok isogloss bundle E

Meaning	Northwest Form	Southeast Form
<i>fingernails</i>	tissy	sala
<i>foot</i>	kolo	hatte
<i>hand</i>	ukkusu	tissy

Table 2.12: Lexical forms represented by Eastern Miwok isogloss bundle F

Meaning	West Form	East Form
<i>arm</i>	paačan	tawa
<i>black</i>	tuhuhi / kululli	šatati
<i>head</i>	hana	huku
<i>pipe</i>	paumma	toʔisa
<i>rain</i>	nykka	ymyča

Table 2.13: Lexical forms represented by Eastern Miwok isogloss bundle G

Meaning	North Form	South Form
<i>deer</i>	hika	uwyja
<i>fire</i>	wyke	hyjy
<i>head</i>	hana	huku

2.2.3 Summary

In the lexical isoglosses that represent this study's word list data, dense bundles occur at precisely the locations where early maps have placed the Central Sierra Miwok – Southern Sierra Miwok boundary and to a somewhat lesser extent the Northern Sierra Miwok – Central Sierra Miwok boundary. The phonological isoglosses representing sibilant and nasal sounds also occur along the boundaries mapped by early twentieth century scholars. The boundary between Plains Miwok and Northern Sierra Miwok is somewhat less distinct than would be expected given this language's subgrouping as a sister to Sierra Miwok, and while this indicates some Plains Miwok influence on nearby Northern Sierra Miwok speech, it is unclear exactly how this is affected by the limited Plains Miwok data and the close ties between western Northern Sierra Miwok groups and the Plains Miwok speakers who remained in the valley and foothills in the 20th century.

The isoglosses representing the reflexes of Proto-Eastern Miwok **y* diverge even more notably from the traditional boundary locations, particularly in the area traditionally associated with Plains, Northern, and Central Sierra Miwok. Here, the isoglosses suggest a change of **y* to *u* in the upper elevations, with evidence that the change has begun to spread across upper and middle elevations, while the diffusion of this change is limited in most of the lower elevation locations. Given that the known Miwok settlement sites occur along rivers which run basically perpendicular to the crest of the Sierra, one might expect linguistic features to be more likely to spread along these rivers than across the mountain slopes. However the fact that several lexical isoglosses also run parallel to the Sierra crest at an elevation that is correlated with ecological boundaries suggests that perhaps the environmental adaptation required for settlement in upper elevation locations led to greater contact across slopes of similar elevations than one might otherwise expect.

2.3 Quantitative Analysis and Hypothesis Testing

The phonological and lexical isoglosses that demarcate linguistic differences among the Eastern Miwok languages align in a few notable bundles, but also include more diffuse patterns, particularly in the Northern/Central Sierra Miwok area. Interpreting these patterns is complicated by a general scarcity of information about the historic and late prehistoric social organization and contact within Eastern Miwok. Unlike dialect geography studies in industrialized societies, where one often finds cumulative linguistic differences radiating from economic and cultural centers and historical information to aid in the identification of leveling and relics, the interpretation of Eastern Miwok feature distributions requires other ways of understanding the social circumstances of linguistic change. The approach introduced here employs aggregate measures of linguistic dissimilarity to develop a more complete understanding of Eastern Miwok dialect geography than isogloss mapping alone can reveal. Quantifying linguistic distance also makes it possible to assess the relationships between environmental variables and the spread of linguistic material by comparing quantified linguistic distances between pairs of locations with cost-modeled geographic distances.

2.3.1 Aggregate Measures of Linguistic Dissimilarity

This study uses two aggregate measures of linguistic dissimilarity: Levenshtein distance as a measure of phonological dissimilarity, and percent non-cognate vocabulary as a measure of lexical dissimilarity. These metrics are frequently used in the field of dialectometry to abstract away from vast quantities of data and identify general patterns in the linguistic relationships between regional dialects. Separate measures are used here to describe the lexical and phonological differences between local varieties so that the spatial patterns in phonological and lexical leveling can be investigated independently. Given the minimal phonological differentiation among Eastern Miwok varieties and the phonological data's susceptibility to transcription-related errors, the separation of phonological and lexical distance also aids in identifying meaningful patterns despite some noise in the data.

The phonological dissimilarity metric called Levenshtein distance calculates the difference between two forms based on a string edit distance algorithm. Essentially, this metric counts the minimal number of segment-by-segment changes that would be required to change the form in one variety to the corresponding form in another language. This metric can be refined to reflect theories of sound change by assigning different weights to different types of change a segment may undergo. For example, the deletion of a segment might be considered more or less costly than replacing a segment with another sound. Another common refinement of the Levenshtein distance algorithm defines segments by their phonological features and segment replacements incur different weights based on the associated changes in phonological features (e.g. Heeringa 2004). All of the procedures for weighting Levenshtein distance involve some theoretical assumptions, and this study uses a fairly simple and theory-neutral version of the Levenshtein distance algorithm. Here insertions and deletions are weighted in such a way that they are half as costly as replacing a segment with a different segment, and all replacements are treated as equally costly.

For this study, then, the Levenshtein distance between the forms *kenata* and *kynatys*⁵ would be calculated as shown in below.

kenata	
kynata	+1
kynaty	+1
kynatys	+0.5
Total	2.5

The algorithm used here is the default algorithm of the RuG/L04 program developed by Peter Kleiweg of Rijksuniversiteit Groningen (Kleiweg 2012). For every pair of locations in the study a Levenshtein distance can be calculated for each item in the word list. This study uses only cognate forms to calculate Levenshtein distance, as lexical differences were captured in a separate metric. In instances where one of the two languages being compared had a non-cognate or missing form, that item is excluded from the calculation. The Levenshtein distances for each item in the word list are averaged to produce a mean Levenshtein distance between the linguistic varieties of every pair of locations. It is these mean Levenshtein dis-

⁵These are the forms for the meaning 'acorn mush' attested in West Point and Big Valley respectively.

tance measures that are used to populate a distance matrix summarizing the phonological distance between every two locations in the sample.

The use of edit distance algorithms to measure linguistic differentiation is a relatively recent development, yet these measures have been widely adopted in dialectometry and have proven to be useful tools. The use of string-edit algorithms to quantify linguistic differences was introduced by Kessler (1995) in his study of Irish Gaelic. Kessler compared the results generated by several variations of the Levenshtein distance algorithm to isogloss patterns and found that algorithms that weighted substitutions according to their phonetic features were not better at identifying dialect areas than unweighted algorithms. Recognizing this, this study used minimal weighting, stipulating only that insertions and deletions should carry half the weight of segment replacements, as described above. Kessler also reported that calculating edit distance only for cognate forms did not produce better results than calculating Levenshtein distance across items that matched in meaning only (1995). This study used only cognate forms, since this resulted in mathematical values that were straightforward, clearly related to the underlying words, and entered into other components of the analysis more simply. Furthermore, this decision is justified by our complementary use of a separate metric to evaluate word list differences on a purely lexical basis. Levenshtein distance has since become a common tool in dialectometry, and has been employed to investigate regional variation in many case studies (e.g. Gooskens 2004; Nerbonne and Kleiweg 2006; Valls et al. 2010). This is an active corner of innovation within the field of dialectometry. However, sophisticated new phonetic weighting systems cannot be implemented in this study due to the nature of the source data, and a simple version of the metric is sufficient for this study's purposes.

While dialectometric studies of closely related regional dialects typically measure only phonological distance, this study also employs a measure of lexical distance. Following the tradition established by Séguy (1973), lexical distance is calculated as a percentage of the word list for which the forms in two varieties are not discernibly cognates. Because this method uses a summary statistic to measure the differences across the entire word list no further calculations are needed to arrive at a single number representing the lexical distance between two locations.

Measuring the differences between the speech at every pair of locations using these two metrics yields a phonological distance matrix and a lexical distance matrix. The values in these matrices, which represent the differences in the much larger and more complex word list dataset, can be used to conduct a number of further analyses. Multidimensional scaling is used to examine the distances between linguistic data points in a way that is visually analogous to geographic mapping. Clustering identifies links between varieties as well as the relative strength of these links. Finally, the distance matrices are used to compare linguistic dissimilarity to geographic distance to explore the spatial variability of the linguistic data and to test hypotheses about the role of environmental factors in conditioning this variability.

2.3.2 Relating Linguistic Differences to Geography

One of the questions this study asks is whether knowledge about the physical environment can be used to help interpret the spatial patterns that occur in Eastern Miwok dialect data. As Bloomfield notes, linguistic boundaries emerge in response to “some line of weakness in

the network of oral communication” (Bloomfield 1933:328). Dialect studies frequently use historical knowledge to relate the spatial boundaries of dialect features with these areas of weak contact, but the limitations of the Eastern Miwok historical record make this difficult. Yet weaknesses in networks of social contact are frequently associated with elements of the physical landscape, and we might expect this to be particularly true in a study area defined by rugged terrain, minimal infrastructure, and a culture that makes very direct use of natural resources.

To assess the role of physical environment factors in conditioning social contact and thereby constraining the spread of linguistic features, this study correlates the linguistic distances discussed in Section 2.3.1 with measures of geographic distance that are weighted according to the cost of travel across different physical landscape elements. This process builds upon a common procedure in dialectometry – assessing the strength of correlations between linguistic dissimilarity and spatial distance – but uses cost distance models to represent several hypotheses about physical factors that might affect the degree of social contact between locations. By comparing the strength of the correlations between linguistic distance and cost-weighted models of physical distance, the explanatory value of these physical factors can be assessed.

Both of the procedures described in this section are supported by scholarly precedents. Correlating distance measures is a staple in dialectometry and cost distance modeling is used for similar network connectivity questions in archaeology. Combining them is a logical way to adapt the standard dialectometry approach to the historical circumstances in this study area. The correlation of linguistic and geographic distances has been argued for by Nerbonne and others working in dialectometry as a way to investigate dialect distributions (e.g. Gooskens 2004; Nerbonne and Heeringa 2007; Nerbonne 2008). Gooskens implements a procedure very much like the one proposed in this study, correlating linguistic and perceptual distances with contemporary and historical travel times between locations to understand the dynamics of dialect variation in Norway (2004). This study found historical travel times to be a better predictor of linguistic distance than straight-line distance or contemporary travel times, and attributed this to the role of Norway’s central mountains as a barrier to contact. Heeringa used data from historical travel guides to reconstruct travel times between locations. This is not possible for pre-contact California, so cost distance modeling is used instead.

Cost distance modeling is used for a wide variety of applications in Geography and Geographic Information Science, from mapping wildlife habitats (e.g. Singleton et al. 2002; Ray et al. 2002; Janin et al. 2009) to planning urban development (e.g. Allen and Lu 2003; Caruso et al. 2005; Braimoh and Onishi 2007). Archaeologists have also used this procedure to model connectivity between sites in an effort to understand the distribution of material culture and identify prehistoric social boundaries (van Leusen 2002; Hare 2004; Wood and Wood 2006; Livingood 2009; Nolan and Cook 2012). The use of cost distance models here is quite similar to their archaeological application. In investigating spatial patterns in both linguistic variation and material culture, cost distance modeling estimates the connectivity of the underlying prehistoric social networks based on known environmental variables.

2.3.3 Distance Models

To determine whether elements of the physical environment have acted as barriers to linguistic contact and diffusion of dialect features the linguistic distances described in Section 2.3.1 are correlated with geographic distances that represent hypotheses about these environmental variables. The null hypothesis, that geographic variables do not impact the spread of linguistic changes, is represented by a Euclidean measure of spatial distance, which is measured along a straight line between locations. The cost distance models are calculated using a far more complex process.

Cost distance is a measure of the cumulative cost of traveling the least costly path between two locations, where the cost of travel is based on a “friction surface”. In a Geographic Information Systems (GIS) context, the “friction surface” is a raster data layer whose cell values are determined by the cost of traveling across that cell. These cost rasters are generally derived from other raster datasets that contain information about relevant geographic features. A wildlife corridor model, for example, might be based upon the theory that major highways are a significant barrier to travel and that the cost of travel through urban land is greater than the cost of travel through forests. The cost raster layer representing this model would be calculated using individual raster datasets containing information about highways, urban land cover, and forests, assigning costs based on the theorized impact of each of these model parameters on travel. These datasets representing individual model parameters can then be overlaid upon one another, and the total cost of crossing each cell can be calculated using an algebraic formula. This hypothetical wildlife corridor model would generate a cost surface in which cells representing forested lands would have low values, cells representing urban areas would have high values, and cells representing highways through urban areas would have the highest values. The creation of cost surface rasters is not limited to adding up factors – the formulas used to derive these datasets can accommodate very simple or very complex models through the use of algebraic expressions, logical operators, and conditional statements.

Once the cost raster representing a particular model of geographic distance has been created, GIS tools are used to identify the total accumulated cost of traveling the least costly path between a source and a destination. This computationally intense procedure involves the further creation of a *cost distance* raster, which the path distance algorithm uses to identify the least costly path from every cell back to the source cell and add up the accumulated costs for each cell. The value of the destination cell in this cost distance raster represents the accumulated costs of travel between that destination and the source cell based on the modeled cost surface. The path distance procedure used in this study computes least cost distance using a cost surface and also uses a digital elevation model (DEM) to adjust the costs of crossing cells based on a trigonometric calculation of the actual surface distance of each cell.

The cost distances in this study were calculated using models based on four physical environment parameters: elevation, vegetation, surface water, and watershed boundaries. Each of these physical elements has been identified as a potential conditioning factor in the development of social networks within the Sierra Nevada and foothills. Elevation, highlighted by Gooskens as a contributor to Norwegian dialect differentiation and frequently used to

model human networks, contributes to the cost of travel based on the increased energy costs associated with moving across steeper slopes (Gooskens 2004; van Leusen 2002; Wood and Wood 2006; Nolan and Cook 2012). The relationship between slope and travel can be modeled in terms of metabolic costs, which is a useful approach when modeling the transport of loads of known weights (Mickelson 2003; Rees 2003). However, the complexity of these models requires detailed inputs that are unavailable here, and such complex metabolic modeling is not necessary for the purposes of this study. Instead, the difficulty of crossing terrain is modeled based on the relative speed at which humans travel over paths with different slopes. This hiking time approach to elevation-related travel costs is appropriate for investigating contact without discriminating between different types of journeys and the burdens carried on these loads.

The primary mode of transportation in the Sierra Nevada and foothills before Euroamerican contact was foot travel. This mode of travel is modeled using Tobler’s Hiking Function (Example 2.1), an algorithm that calculates the velocity of human foot travel based on slope (Tobler 1993; Pingel 2010).⁶ This model predicts that a traveler on perfectly flat ground would travel at around 5 km/hr, and the speed of travel would decrease exponentially with increasing slope, leading to a velocity just over 4 km/hr at 3 degrees slope. In addition to the exponential curve, a conservative upper limit of 50 degrees slope was used to rule out travel across impassable cliff faces and canyon walls.

$$(2.1) \quad V = 6e^{-3.5 |s + .05|}$$

The elevation-based cost surface was derived from a 30 meter DEM layer obtained from the U.S. government’s National Elevation Dataset (Gesh et al. 2002; Gesh 2007). A percent slope raster was calculated using the DEM layer. Using Tobler’s Hiking Function to establish velocity and a 30 meter slope raster, a cost raster was created whose cell values represent the time required for travel across each cell.

Another, vegetation-based cost surface was created to capture the hypothesis that vegetative land cover, and specifically chaparral and dense forest understories, could impede travel and thus constrain contact. While the elevation model used in this study is relatively conservative in its assignment of cost, the vegetation cost model is probably exaggerated due to the reduction in vegetation-related costs when traveling on trails or deer paths. It is unlikely that people traveled in a trackless terrain. However, because the trail system between Eastern Miwok locations is largely unknown, the calculation of vegetation-based costs included very little bare ground.

The vegetation cost surface was derived from CALVEG Historical 1977 vegetation type vector datasets (Remote Sensing Lab 2010). These datasets were converted to 30 meter resolution raster datasets, and walking time coefficients were assigned to each cell based upon detailed descriptions of each vegetation type represented in the GIS data layer (Remote Sensing Lab 2009). The detailed descriptions of the vegetation alliances represented in the CALVEG datasets were used to determine the character of the surfaces created by these vegetation types, and costs were assigned to cells in the friction surface raster based on coefficients for human walking measured by Giovani and Goldman (1971). The coefficients,

⁶In this function, e represents an exponential function and s represents percent slope.

listed in Table 2.14, represent ratios of the time for a human to cross a particular surface type versus flat travel on a treadmill. These coefficients are multiplied with a standard hiking time

Table 2.14: Human walking coefficients from Giovani and Goldman (1971)

Surface	Coefficient	Landcover Types
Flat Treadmill	1.0	
Dirt Road	1.1	Bare Rock
Light Brush	1.2	Grasslands, Low Brush, Low Understories
Heavy Brush	1.5	Chaparral, Dense Understories
Swampy Bog	1.8	
Loose Sand	2.1	

of 0.002 hours per 30 meter raster cell to obtain hiking times for a single-factor vegetation-based cost model. For multi-factor models the coefficients are multiplied with the hiking time values of the other cost surface(s).

Vegetation is sensitive to climate change and human disturbance and there were undoubtedly changes in the distribution of various vegetation alliances in the Sierra and foothills during the historic period and late prehistoric period. This study uses the earliest available complete vegetation dataset for the study region to minimize the impact of late-twentieth century environmental alterations and rapid changes in human land use in this period. More sophisticated control for historical vegetation changes is impractical given the amount of guesswork involved and uncertainty that would be introduced.

The third basic cost surface used in this study is a surface water-based cost layer. This layer represents the hypothesis that the natural lakes and major rivers of the study area serve as boundaries to travel and linguistic contact. Several large rivers flow down from the high Sierra toward the western edge of the study area. These streams are typically steep and swift in their upper reaches and high volume in their lower reaches, and crossing them would certainly require strategic crossing site selection. The natural lakes that are found in the alpine slopes and the lower foothills are considered absolute barriers to travel for obvious reasons. The cost surface associated with the surface water barrier hypothesis assigns cost values only to cells that represent major streams⁷ or natural lakes.

The cost surface representing hypothesized costs of crossing surface water is weighted only in raster cells that overlap with large stream or natural lake features in the National Hydrography Dataset surface water area shapefile (NHDArea) (U.S. Geological Survey 1999). Costs for stream crossings were assigned based on Balstrøm's finding that humans can walk through shallow water at a rate of 1.25 meters per second (2002). The conservative model used in this study assigned a cost coefficient of 4 to all stream cells whose slope was below 2%. Variable cost coefficients were applied to steeper stream cells based on stream channel morphologies that would be expected in various slope ranges, as shown in Table 2.15. Cells associated with cascades and cells associated with lakes were assigned a coefficient of 9,999, making them effective barriers to travel. In reality the difficulty of crossing streams would

⁷For the purposes of this study the major streams in the study area are: Cosumnes River, Mokelumne River, Calaveras River, Stanislaus River, Tuolumne River, Chowchilla River, Fresno River, and Merced River.

Table 2.15: Surface water crossing coefficients

Stream Morphology	Slope Range	Coefficient
Riffle Pool	0% – 2%	4
Rapids	2.1% – 5%	40
Step Pool	5.1% – 20%	400
Cascade	20%	9999
Lake	N/A	9999

depend on a number of factors, including slope, depth, velocity, channel morphology, channel bed materials, and the slope of stream banks. However, many of these geomorphological properties are related – for example, a steep stream would be expected to flow at a higher velocity than an otherwise similar low-slope stream. The slope-based approach captures the general trend that streams become exponentially more difficult to cross as velocity, discharge, or other hazards increase. As with the vegetation cost raster, the travel time cost surface associated with this model multiplies these coefficients by a standard (flat) hiking time of 0.002 hours per 30 meter raster cell to obtain costs in hours. When combined with elevation costs, the coefficients are multiplied by the elevation-based time costs.

The final factor considered in this study is watershed boundaries. Unlike the three factors discussed above, watershed boundaries are not themselves expected to serve as boundaries to linguistic contact. Instead, the organization of settlement sites along river valleys leads to a hypothesis that social networks in Sierra Miwok are largely constrained by drainage networks (Cook 1955). We might expect decisions about local journeys to be affected by the costs of these journeys – these relatively short-distance trips that are unlikely to be motivated by strong economic incentives are more likely to be made frequently if the cost of travel is low. Longer distance trips are likely to be infrequent and motivated by large economic payoffs such as trade for important resources, and so these journeys outside of the immediate social network are more likely to be undertaken in spite of travel costs. Following this logic, the watershed boundaries from the National Hydrography Dataset were used to determine where to apply the cost weighting of other cost surface rasters (U.S. Geological Survey 1999). Because elevation is expected to be the geographic factor most likely to contribute to dialect patterns, this dataset’s weighting is used to compute travel costs in the watershed model. For each point of origin, travel within the same watershed is weighted according to the description of the elevation-based model above. Travel outside the boundaries of the watershed of origin is assigned a uniform cost.

Additional multi-factor models were created by multiplying the costs and coefficients of the models described above. With the exception of the watershed model, which was not combined with other models due to processing constraints, all possible combinations of factors were modeled. The resulting cost distance surfaces represent models based on elevation-and-vegetation, elevation-and-surface water, vegetation-and-surface water, and elevation-vegetation-and-surface water.

Each of the cost distance models described above yields a set of cost-weighted distances that quantifies time to travel between each pair of locations. These distance matrices are the geographic counterpart to the linguistic distance matrices discussed in Section 2.3.1. The measures in the linguistic distance matrices were correlated with the corresponding measures of geographic distance in the cost distance matrices to evaluate the strength of the relationship between these factors. The R^2 statistic is used to quantify the amount of linguistic dissimilarity that is predicted by each of these models of geographic distance. Correlations between each linguistic distance matrix and each geographic distance matrix are conducted for the full set of locations as well as for sub-regions. Comparing linguistic and geographic measures at multiple scales allows for the identification of broad, family-wide patterns as well as the possibility that sub-regions within the study area may pattern more like continua or dialect areas, and may demonstrate greater or lesser sensitivity to these environmental factors.

2.4 Quantitative Analysis Results

2.4.1 Linguistic Distance

Phonological distances were calculated using both the raw dataset and a version of the dataset in which segment lengths were neutralized. Results confirm the expectation that the length-neutralized dataset is less noisy; henceforth “phonological (Levenshtein) distance” will refer to the distance measures derived from this dataset. Results based on the unedited phonological data are reported in Appendix B.

The measures in the phonological and lexical distance matrices were used to conduct clustering using Ward’s Method, and the results of this clustering were plotted as the dendrograms in Figure 2.11 and Figure 2.12.

The first order branching in the phonological clustering dendrogram (Figure 2.11) does not correspond to the traditional divisions between Miwok varieties. The top branch includes Southern Sierra Miwok and Central Sierra Miwok locations, while the lower branch includes Central Sierra Miwok, Northern Sierra Miwok, and Plains Miwok locations. The second order branching in the top portion of the dendrogram separates what is essentially a Southern Sierra Miwok cluster from a Central Sierra Miwok cluster. The exception to this is Groveland, which is geographically near to the traditional Southern-Central Sierra Miwok border and here clusters with Merced Falls and other Southern Sierra Miwok locations. In the lower portion of the dendrogram we find clusters that bear even less resemblance to the traditional designations. The upper group of the second order branch here includes one Central Sierra Miwok locations (Angels Camp), grouping with a far western variety of Northern Sierra Miwok (Pleasant Valley) and a far eastern variety (West Point), and then combining with a group containing most of the remaining Northern Sierra Miwok locations. The bottom branch contains a set of locations that are distributed from the northwest corner of the study area to the central east area and include locations that are traditionally classified as Plains Miwok, Northern Sierra Miwok and Central Sierra Miwok. The low degree of geographic coherence in the bottom primary branch of this dendrogram is reflected in the length of the

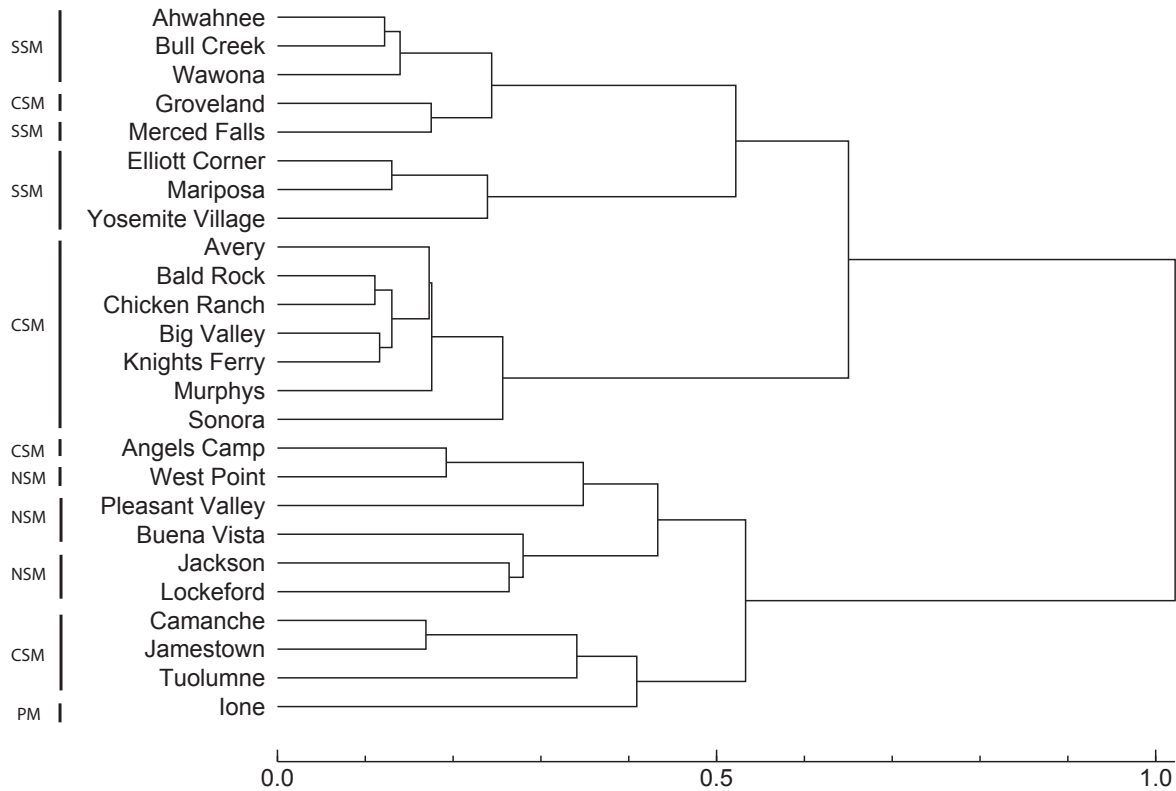


Figure 2.11: Clustering of Eastern Miwok varieties based on phonological (Levenshtein) distances

branches, which are longer on average than the low-level branching in the upper portion of the dendrogram. This dendrogram largely reflects the noisy nature of the phonological data used in this study. However, it is interesting to note that the southern end of the study area seems to be clustered more closely on the basis of phonological distances and these clusters show a greater level of spatial coherence in this sub-region. The central and northern portions of the study area show a far less tidy pattern.

The lexical dendrogram in Figure 2.12 shows stronger relationships between locations and demonstrates relationships that more closely resemble the traditional classification of Eastern Miwok varieties. The primary branching in this dendrogram separates a Southern Sierra Miwok cluster from the remainder of the varieties. This primary cleavage is unsurprising given the patterns that emerged through isogloss mapping. The Southern Sierra Miwok cluster shows two closely related subgroups. The top subgroup (Ahwahnee through Merced Falls) represents the westernmost/lowest elevation Southern Sierra Miwok locations, while the lower subgroup (Bull Creek through Wawona) contains the easternmost/highest elevation Southern Sierra Miwok locations. This subdivision within Southern Sierra Miwok was not obvious from isogloss mapping.

In this lexical dendrogram, the lower primary branch splits into a Central Sierra Miwok

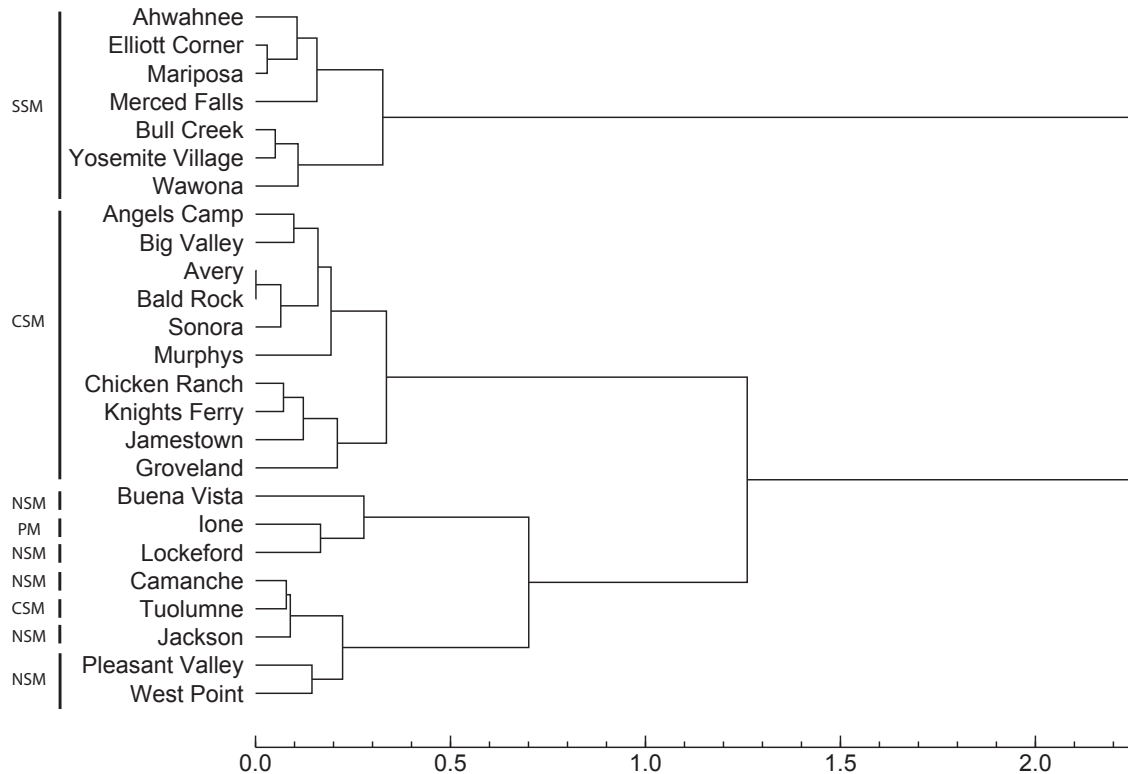


Figure 2.12: Clustering of Eastern Miwok varieties based on lexical distances

cluster and a Northern Sierra Miwok/Plains Miwok cluster. The two main sub-clusters in the Central Sierra Miwok branch do not correspond to geographic sub-regions. The Northern Sierra Miwok/Plains Miwok cluster contains two relatively long branches. The top branch contains Lone, the sole location whose speaker identified as Plains Miwok, as well as Buena Vista and Lockeford, the two nearest geographic neighbors to Lone. The branch below this Plains/western edge cluster contains the remainder of Northern Sierra Miwok as well as Tuolumne, which is located relatively far southeast in the Central Sierra Miwok territory and would be expected to cluster with Central Sierra Miwok. This location's data includes a word list gathered in 1982, which is significantly later than any of the other data used in this study. The unexpected clustering of Tuolumne with the varieties traditionally classified as Northern Sierra Miwok may reflect rapid leveling in the Northern/Central Sierra Miwok area during the late twentieth century as speaker numbers fell dramatically and political reorganization shifted patterns of social contact in this area.

Multidimensional scaling was also used to represent the network of linguistic distances between locations in a format which, like mapping, locates similar data points visually close to one another and dissimilar data points visually distant from one another. Scatter plots representing the linguistic distances between Eastern Miwok locations scaled to three dimensions are shown below in Figure 2.13 and Figure 2.14.

The phonological distance MDS scatter plot in Figure 2.13 shows that the patterns in

Phonological Distance Multidimensional Scaling

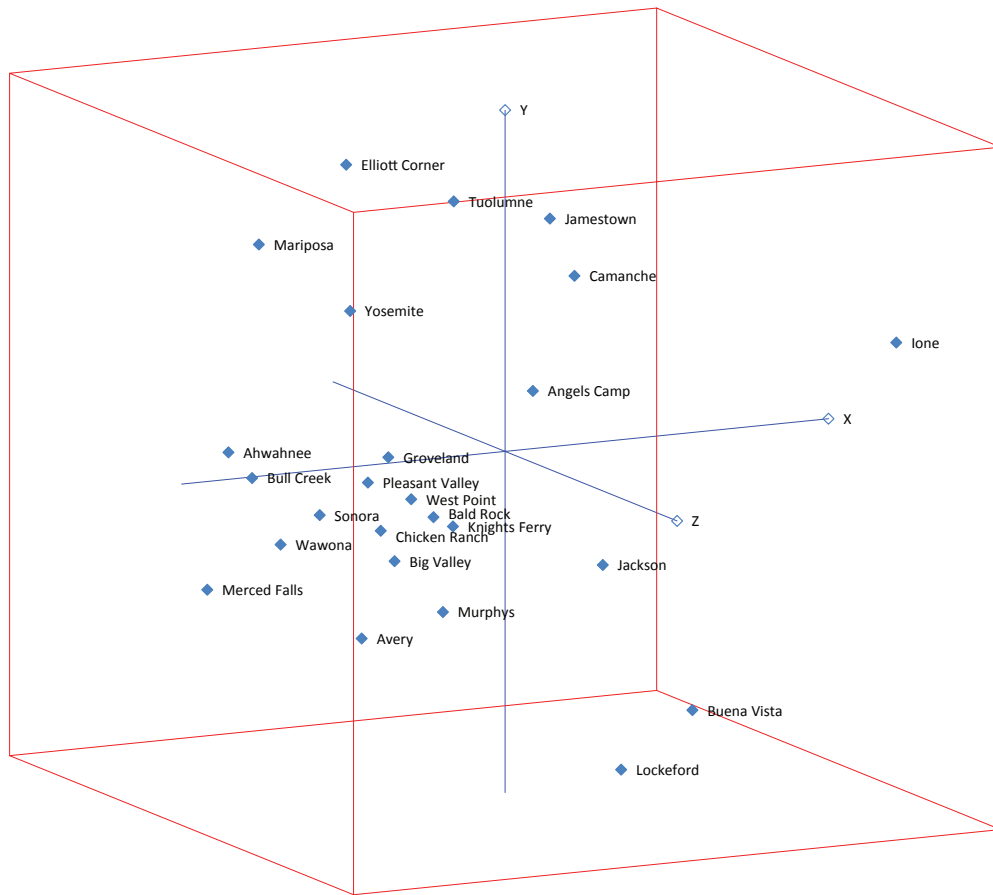


Figure 2.13: 3-Dimensional scaling of Eastern Miwok phonological (Levenshtein) distances

Eastern Miwok phonological feature distribution are generally not very distinct. Ione, the sole location whose informant identified as Plains Miwok, is distant from all other locations as we might expect. Buena Vista and Lockeford, locations which occur very near to the Plains Miwok border and geographically close to Ione, are near to each other but not to the other locations. Jackson, which is also in the general geographic vicinity of Buena Vista and Lockeford, is also located in this lower right front quadrant. Three Southern Sierra Miwok locations appear near to one another in the scatter plot in the top left area, yet these are quite distant from the scatter plot locations of the other Southern Sierra Miwok locations. These three Southern Sierra Miwok locations do not form a geographic sub-area within Southern

Sierra Miwok; the other three Southern Sierra Miwok varieties are found along the left edge of the lower left-hand cluster of points. The points in the top central area of the plot are points in the Central and Northern Sierra Miwok geographic areas, as are the remainder of the points in the lower left-hand region of the plot. Obvious links between this scatter plot and the dendrogram in Figure 2.11 are few. The three Southern Sierra Miwok varieties plotted in the upper left portion of this chart form a dendrogram branch, but the remainder of the dendrogram clusters are not represented on this scatter plots as well-defined groupings.

Lexical Distance Multidimensional Scaling

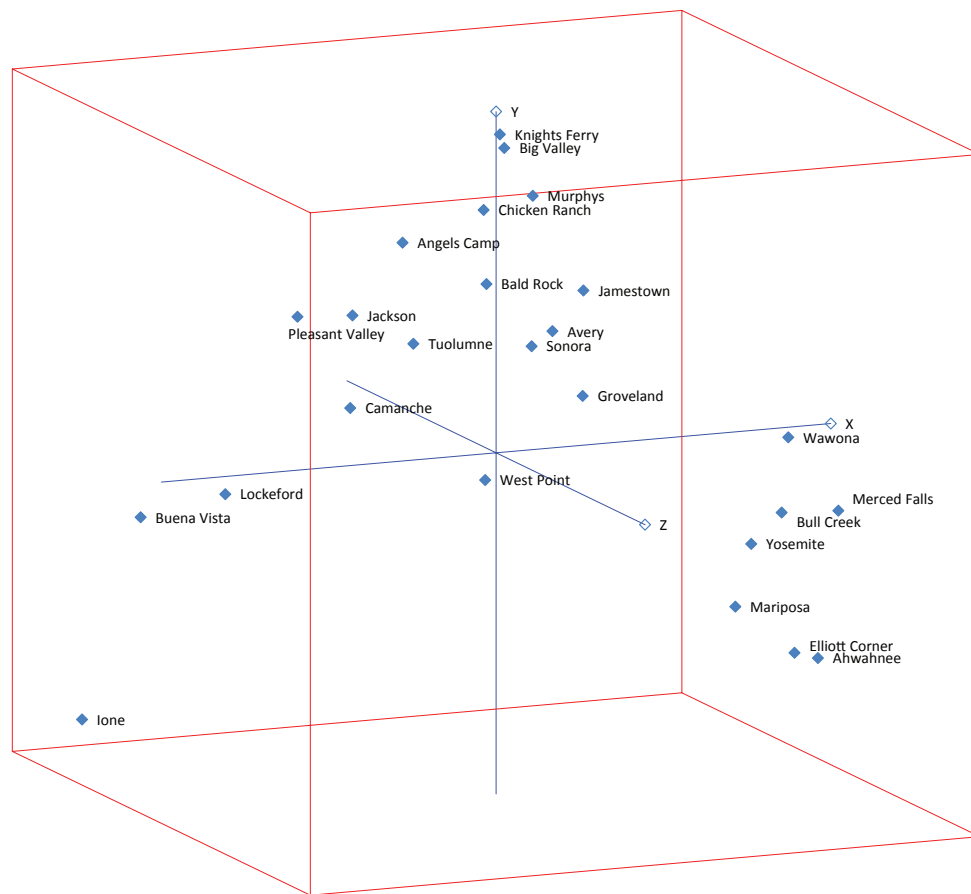


Figure 2.14: 3-Dimensional scaling of Eastern Miwok lexical distances

The lexical distance MDS scatter plot in Figure 2.14 shows a grouping in the lower right front quadrant that corresponds to the traditional Southern Sierra Miwok classification as well as the top first order branch of the dendrogram in Figure 2.12. The Central Sierra Miwok varieties which clustered together in the lexical dendrogram are grouped in the top central

portion of this plot. The Northern Sierra Miwok varieties, which again correspond to both a dendrogram branch and a traditional classification, appear in a fairly distinct group to the left of the Central Sierra Miwok varieties. As in Figure 2.12, Tuolumne appears closer to these Northern Sierra Miwok varieties than we would predict. The lower left portion of the plot shows Ione, the Plains Miwok location, situated quite distant from all other varieties. The two nearest points, Buena Vista and Lockeford, are clustered with Ione in the dendrogram, and their position here slightly closer to the main Northern Sierra Miwok cluster than the sole Plains location supports their traditional classification as Northern Sierra Miwok varieties. The influence of Plains Miwok on these geographic neighbors, however, is also quite evident from these results.

2.4.2 Correlations with Geographic Distances

Table 2.16 contains the R^2 values, or coefficients of determination, that express the strength of the relationships between linguistic distance measures and each of the basic models of geographic distance. The relationships between lexical dissimilarity and geographic distances

Table 2.16: R^2 values for significant correlations between linguistic and geographic distances – all Eastern Miwok locations

	Euclidean	Elevation	Vegetation	Surface Water	Watershed-Constr.
Phonological	0.326	0.242	0.323	0.323	0.294
Lexical	0.552	0.427	0.543	0.541	0.496

for this study area are uniformly stronger than the relationships between phonological dissimilarity and geographic distance. Because the sound changes differentiating the Eastern Miwok varieties are few and the phonological data used in this study is more vulnerable to researcher-based error, this trend is unsurprising. In general the R^2 values presented here demonstrate a slightly weaker correlation between spatial distance and linguistic distance than the $r = .8054$ ($R^2 = 0.65$) reported by Heeringa and Nerbonne (2005:384). But the general range of Pearson’s correlation coefficients (r) for these correlations is very similar to the r values reported by Gooskens for the correlations between linguistic distance and historical travel times in Norway (2004). Furthermore, the finding that geographic distance explains roughly 25 – 55% of the linguistic variation in Eastern Miwok satisfies this study’s expectations. The failure of these correlations to match the strength of the one reported by Heeringa and Nerbonne (2005) is likely a function of several factors. One of these is the variable levels of data consistency in contemporary dialect data and this study’s archival documentary data. Another is the fact that the Dutch locations in that study fell more or less along a straight line, whereas the network of Eastern Miwok locations is physically more similar to a two-dimensional network. The difference between Heeringa and Nerbonne’s result and the ones reported here should not necessarily be interpreted as an indicator that

geography has a lesser impact on linguistic variation in Eastern Miwok than in other dialect study sites.

Comparing the coefficients of determination across geographic models yields some surprising results. The rugged terrain in the western Sierra Nevada slopes and foothills was expected to be a significant factor in explaining the linguistic variation throughout the study area. The R^2 values for the elevation model and the watershed-constrained elevation model are both lower than the R^2 values for the Euclidean model, indicating that for the Eastern Miwok linguistic area as a whole, terrain does not significantly impact the networks of linguistic contact. The R^2 values for the vegetation and surface water models are very similar. Neither of these models adds great costs to the basic Euclidean distance model – water crossings are scarce enough and vegetation is low-impedance enough that these models produce distance measures that are on average only slightly greater than those calculated using basic Euclidean distance. Because the coefficients of determination for these models are lower than those of the Euclidean distance model we can conclude that neither vegetative land cover nor surface water contribute significantly to the patterns of linguistic variation across the entire Eastern Miwok study area. Mantel tests were also used to verify the patterns identified across the whole area, and the results of these tests demonstrated the same relative relationship strengths as the results reported in Table 2.16. They are reported in Appendix B. The multi-factor distance models perform worse than those represented in Table 2.16 when correlated with both phonological and lexical distance. Because of this their coefficients of determination are not reported here. Mantel test results representing the performance of the multi-factor distance models can also be found in Appendix B.

2.4.3 Regional Results

Clusters identified in the Figure 2.12 dendrogram were isolated in order to examine the relationships between geographic distance and linguistic dissimilarity within regions that resemble dialect areas. These smaller-scale patterns are summarized by the correlations in Tables 2.17 through 2.20.

The larger of the first order branches of the dendrogram in Figure 2.12 includes many locations in the region that encompasses the traditional Plans, Northern, and Central Sierra Miwok areas. The distance correlations for this subset of locations exhibits the same trends as identified in the Eastern Miwok dataset as a whole. The correlations between linguistic distances and Euclidean distance are stronger than the correlations between linguistic distance and the cost distance models. As with the overall results, we do not find any support in these results for the hypotheses that physical environment variables impact social contact and dialect geography.

Table 2.17: R^2 values for significant correlations between linguistic and geographic distances
– Central Sierra, Northern Sierra and Plains Miwok

	Euclidean	Elevation	Vegetation	Surface Water	Watershed-Constr.
Phonological	0.433	0.388	0.413	0.431	0.409
Lexical	0.324	0.253	0.313	0.312	0.283

Table 2.18: R^2 values for significant correlations between linguistic and geographic distances
– Southern Sierra Miwok

	Euclidean	Elevation	Vegetation	Surface Water	Watershed-Constr.
Phonological	0.568	0.561	0.571	0.565	0.531
Lexical	0.661	0.685	0.675	0.661	0.635

Table 2.19: R^2 values for significant correlations between linguistic and geographic distances
– Central Sierra Miwok

	Euclidean	Elevation	Vegetation	Surface Water	Watershed-Constr.
Phonological	0.063	0.071	0.059	0.064	0.058
Lexical	0.340	0.331	0.320	0.339	0.161

Table 2.20: R^2 values for significant correlations between linguistic and geographic distances
– Northern Sierra and Plains Miwok

	Euclidean	Elevation	Vegetation	Surface Water	Watershed-Constr.
Phonological	0.049	0.355	0.382	0.376	0.028
Lexical	0.044	0.165	0.174	0.172	0.054

Table 2.18 contains the R^2 values for the linguistic/geographic distance correlations within the Southern Sierra Miwok region identified by a first order branch in the dendrogram in Figure 2.12. As with the overall correlations, the Southern Sierra Miwok results show a stronger correlation between lexical distance and geography than phonological distance and geography, yet for both linguistic measures we find stronger geographic correlations in this sub-region. Unlike the overall results, where the Euclidean measures of geographic distance correlated more closely with linguistic dissimilarity than any of the cost-distance models, within this sub-region the elevation-weighted measures of geographic distance account for the greatest amount of lexical variation and the vegetation-weighted model explains the greatest amount of phonological variation. Additionally, the R^2 values representing the correlations of the elevation model and the surface water model are nearly as closely correlated with phonological distance as the Euclidean model. And the elevation, vegetation, and surface water-based models account for as much as or more of the lexical variation in this sub-region than the Euclidean distance measures. Whereas the overall results did not support the hypothesis that elevation, vegetative land cover, and surface water influenced the connectivity of the Eastern Miwok social network, the results shown in this table suggest that these elements did impact social contact and linguistic feature diffusion in Southern Sierra Miwok.

The results for the varieties grouped in the Central Sierra Miwok-like cluster in Figure 2.12 are strikingly different from the Southern Sierra Miwok group's results. While phonological distance was well correlated with all measures of geographic distance within Southern Sierra Miwok, here we find very poor correlations between phonological differences and geography. Lexical distance in Central Sierra Miwok is also less strongly correlated with geography than we found in Southern Sierra Miwok, but the lexical/geographic distance correlations are still reasonably strong. While none of the geographic distance measures are well correlated with phonological distance in this sub-region, the elevation model's R^2 value is marginally greater than the Euclidean distance R^2 value. We find greater variation in the row representing correlations with lexical distance. Here the watershed-constrained model explains far less of the lexical variation than the other models. While the Euclidean distance measures are better correlated with lexical distances than any of the other models in this region, the coefficients of determination for the elevation, vegetation, and surface water distance models are again quite similar to that of the lexical/Euclidean distance correlation.

The results for the cluster representing Northern Sierra Miwok and Plains Miwok, as clustered in Figure 2.12, are perhaps even more surprising. Again we find much weaker correlations for every pair of linguistic and geographic measures than were found in the Southern Sierra Miwok sub-region and the Plains–Northern–Central Sierra Miwok region. However, here we find a general pattern of stronger correlations between phonological distance and geography than lexical distance and geography. Another striking pattern here is that the correlations between the elevation, vegetation, and surface water-weighted models with both phonological and lexical distance are much stronger than the correlations between Euclidean distance or watershed distance and these linguistic measures. Again this is evidence that at the sub-region scale these elements of the physical environment did influence the social network and through this the development of dialect patterns.

Comparing the relationships between linguistic and geographic distance for the Central

and Northern Sierra Miwok subregions (Tables 2.19 and 2.20) with those associated with the larger region that encompasses both of these subareas (Table 2.17), we find a pattern of much stronger overall links between language and geography in the larger region. The fact that the relationship between linguistic variation and geography is significantly stronger when these areas are combined indicates that while the overall region is organized into spatially defined dialect areas, within these areas a far smaller amount of linguistic variation is organized around geographic patterns.

Mantel tests were not used to verify the patterns found at the regional scale, since the number of language varieties in these small areas creates sample sizes too small for that statistic to be accurately calculated.

2.5 Discussion

The relationships between the geography of Eastern Miwok and the natural environment have been featured in descriptions of this language group since Barrett and Merriam's early work on the classification of California languages (Merriam 1907; Barrett 1908a). These scholars noted the striking coincidence of the family's eastern and western boundaries with the crest of the Sierra Nevada and the edge of the San Joaquin Valley plain. This chapter examines the linguistic diversity within this Eastern Miwok area and evaluates hypotheses regarding environmental factors within the Sierra Nevada mountains and foothills that may have had a more subtle influence on social contact and language change. The quantitative approach used here investigates the impact of the physical environment on the development of the Eastern Miwok dialect network and finds several scale-dependent relationships between environmental factors and the strength of links between local Miwok speech varieties.

The conventional boundaries mapped by Merriam, Kroeber, and Barrett correspond well to the linguistic boundaries identified in this chapter through isogloss mapping and dialectometric distance analysis (Merriam 1907; Barrett 1908a; Kroeber 1925). The traditional Southern Sierra Miwok–Central Sierra Miwok boundary is identified in this study as a strong boundary, and is represented by a lexical isogloss bundle that includes 27 basic vocabulary items. This boundary also generally corresponds to the extent of the spread of an $*\check{s} \rightarrow h$ sound change. The lexical distances calculated in Section 2.3.1 show that Southern Sierra Miwok locations share between 71% and 81% of their vocabulary with Groveland, the nearest location on the Central Sierra Miwok side of this boundary. The vocabulary shared between Southern Sierra Miwok varieties and more distant locations falls to 60% to 70%. The boundary between Southern Sierra Miwok and Central Sierra Miwok is well-defined despite the relative scarcity of phonological changes that separate these areas. On the basis of the evidence uncovered through isogloss mapping and quantitative measures of linguistic relationships, Southern Sierra Miwok appears to be an emerging language in this family.

The standard boundary between Central Sierra Miwok and Northern Sierra Miwok also corresponds to an isogloss bundle, but here the differences between the two linguistic areas delineated by this border are less distinct. The lexical items corresponding to 17 basic meanings form the isogloss bundle running along this traditional border, but no sound changes align with this boundary. The shared basic vocabulary measures between varieties in the

Central Sierra Miwok area and the Northern Sierra Miwok area range widely between 55% and 87%, with the majority in the 80% range. The influence of Plains Miwok on the westernmost locations in the Northern Sierra Miwok area accounts for much of this variation, and the mapping the distance measures shows that the boundary is weakest in the upper elevations, with Avery and West Point sharing 85% of their basic vocabulary. The isogloss bundle discussed in Section 2.2.2 suggests that the regions delineated by the boundary drawn by Barrett and Kroeber are distinct dialect areas. Based on the lack of phonological differentiation, the number of lexical isoglosses that align with this boundary, and the complexity of the linguistic distance measures across it, however, this boundary does not seem to represent as strong a linguistic divide as the Central–Southern Sierra Miwok border. Northern and Central Sierra Miwok may be better described as dialect areas than separate languages.

The Plains Miwok–Northern Sierra Miwok boundary is reported by Barrett to be the strongest of the Eastern Miwok internal boundaries (1908a), and is supported here by an isogloss bundle containing 42 lexical items. Yet a more detailed understanding of variation within and around Plains Miwok is difficult to achieve due to the limitations of the available Plains Miwok data. The percent shared vocabulary between Ione and nearby locations ranges between 48% and 83%. The two locations nearest to Ione share upwards of 74% of their vocabulary with this Plains Miwok variety, but the lexical similarity drops off steeply into the 45% to 65% range for the rest of the Northern Sierra Miwok varieties. The boundary between Plains Miwok and Northern Sierra Miwok corresponds to more phonological differences than the two other boundaries discussed above. Both an $*\eta \rightarrow n$ change and an $*\check{s} \rightarrow h$ change distinguish Plains Miwok, although like the lexical differences these sound changes have diffused into nearby varieties that were identified by the data collectors as Northern Sierra Miwok. Despite evidence of feature spread across the Plains Miwok–Northern Sierra Miwok boundary, the divergence of Plains Miwok from the Sierra Miwok varieties is obvious from the extremely low shared vocabulary values between Ione and all but the nearest locations. The strong presence of Plains Miwok features in western varieties of Northern Sierra Miwok probably reflects a movement of Plains Miwok speakers into these neighbor communities in the wake of Spanish missionization and heavy agricultural settlement throughout the entire historical Plains Miwok region.

Through the quantification of aggregate linguistic dissimilarity and geographic cost distances this study has also uncovered new insights about the dialect geography within these primary Eastern Miwok divisions. Strong correlations between linguistic distance and geographic distance, as reported in Table 2.18, suggest that the variation within Southern Sierra Miwok is somewhat continuum-like in its spatial distribution. The linguistic differences between nearby locations are relatively small, but over greater geographic distances the linguistic dissimilarity is greater. The clustering of locations based on lexical distance, presented in Figure 2.12, shows two branches within Southern Sierra Miwok. Lower elevation sites (Ahwahnee, Elliott Corner, Mariposa, Merced Falls) group together, and upper elevation sites (Bull Creek, Yosemite, Wawona) form the other Southern Sierra Miwok sub-cluster. The correlation of linguistic distances with geographic cost distances produces evidence that elevation and vegetation had some impact on the network of linguistic differences among Southern Sierra Miwok locations beyond the effects of simple Euclidean distance. Together the clustering results and the geographic modeling results suggest that the cost of traveling

over rugged terrain influenced the pattern of contact between Southern Sierra Miwok communities and that uphill and downhill sub-dialects of Southern Sierra Miwok may have been emerging in the early 20th century.

The very low coefficients of determination for the correlations between phonological distance and geographic distance within the Central Sierra Miwok area indicate that there are no geographically-conditioned continuum effects in the phonology of this dialect. Slightly stronger correlations link lexical dissimilarity and geographic distance, and while this is evidence of spatial diffusion of linguistic features, the low R^2 values indicate that the relationships between locations in this region are determined primarily by factors other than geographic distance. This region fails to cluster as a single group based on phonological data, as Figure 2.11 shows, but the dendrogram based on lexical distances includes two sub-groups that suggest a divide between Sonora and Jamestown, running parallel to the Sierra crest. The isoglosses represented by bundles b. and c. in Figure 2.10 do not coincide exactly with the divide suggested by the lexical clustering in Figure 2.12, but do fall generally near the boundary that would divide the dendrogram subgroups. The overall picture of Central Sierra Miwok we are left with is one in which there are some differences between high elevation and lower elevation varieties, but these sub-dialects are phonologically very similar to one another and local variation among regional varieties of both of these sub-dialects is quite high. This pattern is quite interesting given Central Sierra Miwok's location in the center of the study area, and it brings to mind the transition area discussed by Davis and McDavid (1950), where linguistic influence from two or more directions leads to a complex pattern of spatial variation in dialect features. Influence radiating from a single source frequently creates the patterns of accumulated differentiation that we associate with dialect continua, but when these patterns overlap the resulting dialect networks appear a lot less orderly. We do find evidence of influence from both Southern Sierra Miwok and Plains and Northern Sierra Miwok in this region, and the pattern of linguistic similarities in this region may be shaped by the dual influences of Northern Sierra Miwok and Southern Sierra Miwok overlaid on some variation that is inherent to this region.

It is also possible that the low correlation values between geographic distance and linguistic differentiation in the Central Sierra Miwok region are associated with a high rate of migration – either seasonal or permanent – which has created less stability in this region. Differential rates in migration across the Sierra Miwok sub-regions might also help explain why Southern Sierra Miwok, which is separated by a dense isogloss bundle, also shows the strongest correlations between geographic and linguistic distances. A less migratory lifestyle might make sense in this high sierra region, where the wide range of ecological zones between valley floor and valley wall would provide some diversity in resources, and it might also result in enough population stability to explain the differences between this area and the rest of the dialect network.

Northern Sierra Miwok and Plains Miwok are represented by the fewest data points in this study. The small number of sample locations makes it difficult to assess whether geographic patterns that emerge through clustering are significant or merely coincidental. Yet Figure 2.12 shows Camanche and Jackson, both lower elevation varieties, clustering together while Pleasant Valley and West Point, both upper elevation varieties, group together. These two very small groupings form the basic Northern Sierra Miwok cluster in Figure 2.12. The

remaining two Northern Sierra Miwok locations cluster with Plains Miwok on the basis of lexical similarity, and this cluster combines with the aforementioned basic Northern Sierra Miwok branch to make a larger grouping. The branch that connects these groups is rather long, suggesting that the influence of Plains Miwok on the lexicons of the westernmost Northern Sierra Miwok varieties is quite strong. As mentioned above in the discussion of the primary Eastern Miwok language and dialect boundaries, this diffusion of Plains Miwok material into the western edge of Northern Sierra Miwok also includes phonological features. Interestingly, this is the only region where phonological distance is more closely correlated with linguistic variables than lexical distance is. Although the correlations between linguistic and geographic distance are generally weak in this region as a result of the Plains Miwok influence, the correlations linking linguistic distances to elevation, vegetation, and surface water-based cost distances are far stronger within this dialect area than the correlations linking linguistic variation to simple Euclidean distance. For this region all three of these variables appear to play a role in the networks through which linguistic material spreads.

The hypotheses introduced here that elevation, vegetative land cover, and surface water features affect the relative degree of contact between locations are not supported by quantitative analysis at the language family scale. When linguistic dissimilarity and cost distances modeled on these variables are compared across all study sites the resulting correlations are weaker than the relationship between linguistic dissimilarity and basic Euclidean distance. Although the residuals of the correlations show a great amount of variation that does not neatly correspond to the major boundaries mapped by Barrett and Kroeber, the occurrence of discrete dialect areas is known to result in lower language-geography correlation strengths. The absence of elevation, vegetation, and surface water effects at this scale also reflects constraints on human foot travel. While these physical environment factors are likely to be important for routine local travel, where convenience is likely to impact the frequency of contact, physical factors are less likely to influence decisions about whether to make long distance trips. The motivations for long distance travel, the infrequency of these journeys, and the likelihood that these trips would be undertaken by only a subset of the population mean that not only is geography less likely to impact this travel, but also that long distance contact is likely to play a smaller role in linguistic feature diffusion. Because human foot travel was the primary mode of transportation in this region, the scale at which we would expect to see environmental costs influencing social networks is related to the scale of reasonable human walking distance.

At the dialect area scale we do find that distances weighted by the travel costs associated with elevation, vegetative land cover, and surface water features are better correlated with linguistic distance than simple Euclidean distance measures are. In addition to being sensitive to scale, the explanatory value of these factors varies from one dialect area to another. Some of this variation may be related to regional differences in the environmental factors themselves, but in the Central Sierra Miwok area the relationships between language and geography also seem to be outweighed by unexplained linguistic variation. Therefore both the scale of environmental features and more intangible social network properties can affect the relationships between physical environment factors and linguistic patterns. The application of these methods to other historical dialect geography studies could be useful if the impact of scale is kept in mind. However, the modeling of physical environment costs in the contact

network must be re-evaluated and adjusted for each study based upon the local environment and the ways in which people interact with that environment.

The process of aggregating linguistic data to produce distance measures and comparing these measures with cost distance representations of geographic distance is particularly useful in language group like Eastern Miwok, where limited historical and linguistic data, missing forms in word lists, and inconsistent practices in data collection complicate the comparison of individual forms. Condensing both the linguistic data and environmental data into distance metrics makes it possible to identify trends in noisy datasets. In areas where more data is available and non-aggregate quantitative measures of dialect feature occurrence are possible, spatial statistics, including spatial autocorrelation, could add further detail to our understanding of dialect geography.

2.6 Conclusions

The methodology employed in this study introduces a computational procedure for evaluating the relationships between linguistic diversity and geographic barriers to travel. Such a relationship has been imputed for Eastern Miwok for more than 100 years. Until now, there has been no formal statistical approach to this issue. The cost distance models implemented in this study reveal that environmental influences on language contact, at least in this region, are very scale dependent. In pre-contact times, when the primary mode of travel was by foot, the scale at which we see environmental barrier effects is very similar to the scale at which we might expect frequent travel to occur. Thus, there is no evidence for overall patterns of environmental influences on contact throughout the entire Eastern Miwok territory, but within each of the traditional divisions of Eastern Miwok we find patterns within the linguistic network that reflect particular environmental influences. In Southern Sierra Miwok, elevation and vegetation play a small but significant role in shaping the dialect network. In the Plains and Northern Sierra Miwok area, elevation, vegetation and surface water all seem to have played a somewhat more significant role in shaping contact. In Central Sierra Miwok, elevation and surface water appear to have influenced the patterns of linguistic contact through which the dialect network has developed. The scale-dependency of these environmental influences and the variation in their impact throughout the Eastern Miwok region reflect the scale of contact and the overall geographic diversity of the region.

Certain facts about linguistic diversity in Eastern Miwok have also been revealed or corroborated by this study. These findings include support for the traditional boundaries drawn between the primary divisions of Eastern Miwok as well as a more detailed characterization of the variation within each of these divisions. The Central Sierra Miwok–Southern Sierra Miwok border is the strongest boundary identified, followed by the Plains–Northern Sierra Miwok boundary. The Northern Sierra Miwok–Central Sierra Miwok is less distinct, and these two traditional divisions of Eastern Miwok are separated by lexical isoglosses but no sound changes. This boundary is particularly weak in upper elevations. The overall pattern of linguistic similarities and dissimilarities in the region suggests that Plains Miwok and Southern Sierra Miwok were at least emerging languages at the time of documentation, but that Northern Sierra Miwok and Central Sierra Miwok together are better characterized as a

single dialect area, with a weak internal boundary separating slightly more lexically different varieties.

The Sierra Miwok languages all show some evidence of the development of uphill and downhill sub-dialects, as demonstrated by clustering and isogloss patterns. In Southern Sierra Miwok there appears to be an emerging pattern of uphill and downhill dialects, which show up as clusters on dendrograms derived from lexical and phonological distances. This emerging division appears to be related to the influences of elevation and vegetation on contact within the Southern Sierra Miwok region. Northern and Central Sierra Miwok also show evidence, albeit much weaker, for an emerging divide between uphill and downhill dialects. Lexical patterns in Central Sierra Miwok, for example, suggest a divide between Jamestown and Sonora.

These findings represent a significant advance in our understanding of Eastern Miwok. However, the fundamental value of this study is that it demonstrates the validity of the method described above for evaluating environmental influences on the development of dialect diversity.

Chapter 3

Deep Relationships Among California Languages

3.1 Introduction

The geographic area now contained in the state of California coincides roughly with a region of tremendous linguistic diversity. Historical methods in linguistics have illuminated many of the genealogical relationships between the numerous languages of this region, but the deepest levels of interrelatedness that have been proposed are shrouded in uncertainty. In particular, the Hokan group, whose member languages ring the edges of the state, and the Penutian group, which extends through Central and Northern California, have been debated – and remain largely unsubstantiated – since their original proposal.

This study reflects a desire to investigate the validity of these deep language families and to explore the utility and broader applicability of relatively new computational phylogenetic techniques. Specifically, this study employs a quantitative measure of recurrent sound correspondences, a Monte Carlo-style statistical analysis, and a clustering method driven by multilateral comparisons in order to piece together the significant phylogenetic relationships among California languages. This study provides both an assessment of the suitability of this phylogenetic methodology for identifying the evidence of deep hypothetical subgroupings like Hokan and Penutian and an examination of the relationships between California languages.

The suitability of the statistical method for detecting California relationships and the existence of distant California linguistic relationships is addressed in several steps. Section 3.2 introduces the controversial Hokan and Penutian stocks, classifications that have had a broad but contentious impact on the study of California languages – and beyond, into the broader “lumping” and “splitting” debate – over the course of the last century. Section 3.3 discusses the suitability of a methodology proposed by Kessler (1999; 2001) for identifying distant California relationships, and outlines the procedure. Section 3.4 discusses the sample of languages used to represent the California area and outlines some possible outcomes. Results of the test are presented and discussed in section 3.5, and some concluding remarks are offered in section 3.6.

3.2 Hokan and Penutian as Genealogical Classifications

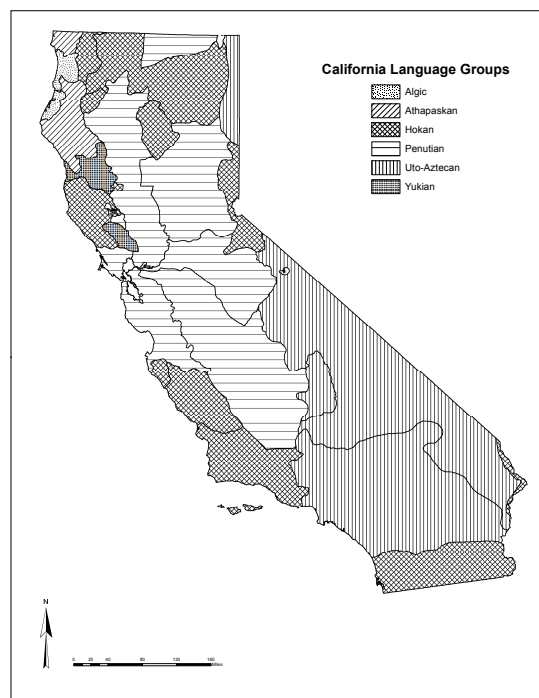


Figure 3.1: California language map showing stock-level groupings

3.2.1 Origins of the Hokan and Penutian Stocks

The terms “Hokan” and “Penutian” are used frequently in Californianist linguistics and have been adopted rather uncritically in certain typological literature, but there is considerable variation in what they are intended to signify. The assumption that the languages included under these designations are genealogically related is embraced by some in the field, while others use the terms to indicate groups that bear some features in common, though not necessarily with the assumption that this is a result of common ancestry. This nomenclature has become familiar and convenient for making reference to California languages, but the pervasive use of “Hokan” and “Penutian” does not indicate that any consensus has been reached regarding the nature of the relationships they denote. The languages that fall under the Hokan and Penutian umbrellas, however, are more or less agreed upon, at least within California. The memberships of the California kernels of the Hokan and Penutian groups are comprised of the following languages and families:

Hokan		Penutian	
Chimariko	Salinan	Maiduan	Wintuan
Esselen	Shasta	Miwokan	Yokutsan
Karuk	Washo	Ohlone	
Palaihnihan	Yana		
Pomoan	Yuman-Cochimí		

When the terms were introduced by Dixon and Kroeber (1913), they were presented explicitly as new genealogical designations. Smaller families had already been identified in California (e.g. Powell 1891), yet Dixon and Kroeber were the first to hypothesize that the diversity of the California linguistic region might boil down to divergence between sub-branches of a couple of very deep language families.

The initial description of the Hokan and Penutian stocks presented only a handful of cognates and a sketch of Penutian case suffixes to support the new family designations (Dixon and Kroeber 1913). A subsequent publication presented their evidence for the classifications in its entirety, including larger sets of potential cognates, sound correspondences, and proposals for the historical processes that might explain how attested forms in daughter languages could have arisen from common ancestors (Dixon and Kroeber 1919). A prominent element of this foundational paper is a table, in which the “stem resemblances” between pairs of California languages were tallied (see Table 3.1).

Table 3.1: Pairwise counts of “stem resemblances” among California languages reported by Dixon and Kroeber (1919:51)

Athabascan	-	Athabascan	Yurok	Wiyot	Lutuami	Karok	Shasta	Chimariko	Yana	Pomo	Yuki	Washo	Wintun	Maidu	Miwok	Costanoan	Yokuts	Esselen	Salinan	Chumash	Shoshonean	Yuman	Aggregate	Stems Represented
Yurok	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
Wiyot	0	4	-	1	1	3	1	2	2	2	2	2	1	1	2	2	2	1	0	1	1	1	13	8
Lutuami	0	1	-	1	1	11	4	9	9	4	6	4	4	7	2	2	6	1	0	2	0	0	34	21
Karok	0	1	3	1	-	9	9	10	-	13	21	5	2	2	2	2	2	3	1	5	0	4	61	13
Shasta	0	0	1	11	9	-	19	13	9	13	15	6	2	8	4	5	9	4	5	5	0	7	141	49
Chimariko	0	1	2	4	9	10	-	9	13	13	15	6	2	4	2	5	4	7	2	5	1	6	109	30
Yana	0	0	2	2	4	13	9	-	15	-	14	2	3	7	3	4	7	6	2	3	3	5	97	27
Pomo	1	1	3	6	6	21	13	15	-	-	14	2	17	8	11	7	5	8	1	2	5	5	151	57
Yuki	0	1	0	4	2	5	6	2	14	-	4	4	14	12	12	9	8	4	0	2	3	3	105	34
Washo	0	0	1	0	3	2	2	3	2	4	-	3	4	4	3	2	3	2	1	4	3	3	45	8
Wintun	0	1	1	4	2	12	8	7	17	14	3	-	23	27	10	17	5	2	5	2	4	164	62	
Maidu	0	0	1	7	2	8	4	3	8	12	4	23	-	18	6	24	3	3	4	4	2	136	58	
Miwok	0	0	2	2	2	5	2	4	11	12	3	27	18	-	27	18	5	5	5	6	3	158	64	
Costanoan	0	0	2	2	2	5	5	7	7	9	2	10	6	27	-	19	8	4	5	5	2	127	43	
Yokuts	0	0	1	6	2	9	4	7	5	8	3	17	24	18	19	-	4	2	8	9	2	148	57	
Esselen	0	1	1	1	3	4	7	6	8	4	2	5	3	3	5	8	4	-	4	8	1	5	78	16
Salinan	0	0	0	0	1	5	2	2	1	0	1	2	3	3	6	4	2	4	-	12	3	4	51	19
Chumash	0	1	1	2	5	5	5	3	2	2	4	5	4	4	5	5	8	12	-	4	6	87	23	
Shoshonean	1	1	1	0	0	0	1	3	5	3	3	3	2	4	6	5	9	1	3	4	-	5	57	24
Yuman	0	1	1	0	4	7	6	5	5	3	3	4	4	2	3	2	2	4	3	6	5	-	66	15
Aggregate	2	13	28	54	61	141	109	97	151	105	45	164	136	158	127	148	78	51	87	57	66	-	-	

This tabular summary of the data from which Dixon and Kroeber identified the Hokan and Penutian stocks appears at a glance to show some quantitative evidence for these proposed genealogical groupings. More careful consideration of what these data represent, however, reveals far more uncertainty than is implied by the counts, sums, and tidy boxes outlining the Hokan and Penutian groups. First, it is unclear from this table how resemblances were identified. It is important to understand both the criteria that were used to distinguish resemblances and what input data was considered in this process. The cognate lists in the article (Dixon and Kroeber 1919:55-69, 105-112) explicate some portion of the evidence summarized in this table, but there is no description of the process by which items were determined to resemble one another. Furthermore, the boxes delineating the Hokan and Penutian groupings in this table represent another layer of subjective interpretation of the data. The pattern represented in this figure, then, is one of apparent relatedness that is subject to a great deal of researcher bias. While this figure is perhaps useful as a summary of the data presented later in the paper, it is not a meaningful tool for assessing quantitatively the significance of the Hokan and Penutian genealogical groupings.

Dixon and Kroeber's discussion of the data in Table 3.1 (1919:49-50) suggests that their aim was to collect sets of lexical items from a sample of California languages in order to identify significant relationships between well-established families. This approach to the question of large genealogical stocks in California is very similar in spirit to the one employed in this chapter. A key difference between the methodologies employed by Dixon and Kroeber (1919) and in this chapter is the notion of significance. While Dixon and Kroeber suggested that the counts of lexical resemblances were themselves indicators of significant relationships, this paper employs a statistical test of significance. The matter of implementing a test for statistical significance in a study of lexical resemblances will be revisited in section 3.3.3. Without such an objective measure to substantiate their conclusions, however, Dixon and Kroeber's paper left the proposed families open to both support and skepticism.

3.2.2 Evidence in Support of Hokan and Penutian as Genealogical Stocks

The Hokan and Penutian groupings quickly became frameworks for investigating the history of California languages. The search for evidence regarding the nature of the Hokan and Penutian stocks gained momentum in the mid-twentieth century. Arguments in support of these groups as genealogical entities introduced many types of data to supplement Dixon and Kroeber's limited early evidence. Several sets of cognates were compiled for each stock (e.g. Sapir 1917; Haas 1963 for Hokan, Pitkin and Shipley 1958 for Penutian). Sapir also identified grammatical features that were characteristic of the Hokan languages (1917) and the Penutian languages (1921b). Numerous pairwise comparisons of languages or small families within a stock were conducted, contributing less directly but no less significantly to the overall Hokan and Penutian debates (e.g. Jacobsen 1958; Olmsted 1956; McLendon 1964; Silver 1964 for Hokan; Broadbent and Pitkin 1964; Callaghan 1997 for Penutian). Eventually, more extensive sound correspondences were noted and lexical reconstructions were undertaken for each of the large stocks (Kaufman 1988 for Hokan; Berman 1983 for Penutian). And though it is generally regarded as more intriguing than convincing, Greenberg's "mass

lexical comparison” of languages in America (1987) also argued for genealogical relationships as the basis for both Hokan and Penutian.

3.2.3 Controversy

Since they were introduced, the Hokan and Penutian hypotheses have met with considerable skepticism. Hokan in particular has been surrounded by doubt. On top of a quiet consensus in the field that the evidence for a genealogical relationship among the Hokan languages is inadequate, caveats regarding the general insufficiency of the evidence – and even the certainty of their own findings – have appeared in papers that argue in favor of the Hokan hypothesis (e.g. Gursky 1974:173, as translated by Campbell 1997:295). Penutian has been more generally viewed as a probable genealogical grouping (Golla 2011:128), though it too was treated with caution when first proposed (see, e.g. Frachtenberg 1918:177), and many researchers who are familiar with the evidence still consider it unsubstantiated (e.g. Shipley 1980; Campbell 1997:317). Attempts to expand the Hokan and Penutian stocks to include languages far afield from their California origins have provoked even more critical opinions of the groupings (see Bolnick et al. (2004) for a summary of criticism of Greenberg (1987) and Voegelin and Voegelin (1973, 1966) for a deconstruction of Sapir’s (1921a) Hokan-Siouan).

The lack of resolution regarding the Hokan and Penutian hypotheses, even when restricted to California, reflects a lack of data. The great temporal distance between documented, modern California languages and their hypothetical Proto-Hokan and Proto-Penutian ancestors would diminish the strength of a genealogical signal considerably. The varying states of documentation of contemporary languages do not make these patterns easy to detect. Documentation for most California languages began at periods when these languages were already severely endangered. Some even passed into extinction before thorough lexical and grammatical descriptions could be compiled. Finding the data to conduct comparative studies of Hokan and Penutian can be a difficult problem in itself. Another obstacle to understanding Hokan and Penutian is that the pairwise comparisons and investigations that have been used to investigate the stocks contain data for only a few languages. These studies are important steps towards understanding the deeper relationships, but the generalizability of their findings to higher levels of the family trees is limited.

When Hokan and Penutian were originally proposed, a relatively limited amount of documentation and analytical work had been done on these languages. By now, however, more of the data relevant to the linguistic history of California is accessible through published papers, dictionaries, grammars, and archives. The analyses that now exist are in place for many languages (e.g. phonemic transcriptions and morphological decomposition of lexical items) further aid the use of that data in comparative studies. Advances in computing also make the investigation of deep genealogical relationships more feasible, since procedures that would have been prohibitively computationally intensive before the digital era can now make use of relatively small amounts of data to identify statistically significant patterns.

3.3 Kessler's R^2 : A Suitable Method for California

The metric that this study uses to test the deep relationships between California languages is Kessler's R^2 , a measure of the recurrence of sound correspondences. The procedure outlined by Kessler (2001) for using this metric in a test of the strength and significance of relationships is a promising tack to take in looking for quantitative evidence regarding the Hokan and Penutian hypotheses. The procedure's reported ability to identify deep relationships and its use of lexical data make it theoretically a good match for California, and this study serves as an exploration of its suitability for the question of distant California relationships as well as a test of those deep groupings.

3.3.1 Reviving Lexicostatistics

With the introduction of new phylogenetic methods, increasingly varied types of data have been used as inputs for statistical historical linguistic investigations. The lexicostatistical methods that emerged in the mid-twentieth century established the importance of vocabulary lists as a source of quantifiable comparative data. Grammatical features have also been incorporated into quantitative historical linguistic studies, primarily in conjunction with cladistic methods adapted from evolutionary biology for use in linguistics. Methodological decisions for a computational phylogenetic study, therefore, require simultaneous consideration of candidate procedures and the data that can serve as their inputs.

The popularity of statistical studies based on lexical data fell sharply in the second half of the twentieth century due to controversy surrounding glottochronology, or lexicostatistical dating of linguistic relationships (as advocated in papers such as Swadesh (1955), Lees (1953) and Sankoff (1972)). The assumptions that must be made to estimate the time depth of linguistic relationships from lexical data have been widely criticized (Bergsland and Vogt 1962; Sjoberg and Sjoberg 1956), but glottochronology is only one sub-branch of lexicostatistics. Many methods that have achieved wider acceptance in modern statistical phylogenetic studies use lexical data but do not attempt to attach dates to family tree branches. Word lists can be used for relatively simple metrics of correspondence, more complicated measures of phonetic similarity, and in the identification of cognates which can serve as characters for cladistic analysis. Each of these uses of lexical data shapes a statistical study in a different way, but all can be employed in tests of relatedness that do not involve controversial chronological components.

The types of grammatical characters commonly used in phylogenetic studies include inflectional and derivational morphemes, phonological features (which can be expressed as the presence or absence of a certain sound change in a language as in Ringe et al. (2002)), and other typologically interesting morphosyntactic elements. These characters are most commonly employed in the tree-building methods that were borrowed from biology into linguistics. In such programs, the character states for each language are generally compared to one another in order to build and evaluate possible family trees.

All of the types of data that are commonly included in phylogenetic studies have both advantages and drawbacks, making different methods most appropriate in slightly different situations (McMahon and McMahon 2005:71-73). Grammatical features like inflectional af-

fixes and phonological features have the advantage of time depth — these characters are often relatively diachronically stable, and can reflect distant stages of a language's development. However, the line of inheritance for these features cannot always be assumed to be direct. Areal features, which are spread across multiple language families in a single region, are one of the defining characteristics of linguistic residual zones (Nichols 1992:21). With its deep history of settlement and geographic density of languages, the California linguistic region is a classic residual zone. Contact-related spread must therefore be considered a plausible mechanism for grammatical feature transmission in this region.

Even if areal features could be reliably identified and removed from character lists, the remaining grammatical features might be difficult to analyze in a meaningful way. Part of the purpose of employing statistical methods in historical linguistics is to assess the significance of the relationships that are found. The probabilities of occurrence of specific grammatical features have not been analyzed in a comprehensive and quantifiable way, nor have the statistical probabilities of co-occurrence of these features. In this regard, linguistics is unlike biology. Genes have a very high probability of being inherited, since mutation of any gene is relatively rare and the chemical and biological constraints on genetics presumably do not disrupt this chain of inheritance. Yet the effects of general human cognition on language are not as well understood, and basic properties of human cognition could have considerable influence on the way languages develop. As a result, linguists can expect to find some homoplasy – the parallel innovation, convergence, and reversal of changes that can create non-genealogical similarities – in grammatical features, which will not only interfere with the determination of genealogical relationships, but will also make it difficult to determine how likely these genealogical features are to occur due only to chance. Confidence in the results of statistical procedures is undermined when the assumptions of the method are not entirely compatible with the input data. Until some of these questions about grammatical development are better understood, cladistic models based on grammatical characters are best used cautiously in tests that involve other types of data.

Another category of data, which encompasses some types of grammatical data and some types of lexical data, is comprised of characters that are derived from traditional historical methodologies. Both sound changes and known cognates fall into this domain. Unlike most grammatical features, whose demonstrations of genealogical relationships are muddled by the possibility of borrowing or homoplasy, historically defined features are clear indicators of common ancestry. While the certainty introduced by these characters is undoubtedly one of their strengths, one important limiting factor is that they duplicate the results of traditional methods and fail to add anything to the pool of data beyond what has already been verified as genealogical. Historically defined characters might be useful for strengthening the results of clustering programs that consider multiple types of characters, but in the absence of other data these characters are not well suited for investigating relationships that are deeper than the temporal range of the comparative method.

Grammatical characters and data derived from traditional historical linguistic methods are both common inputs for cladistic studies, where each language is coded with a character state for each linguistic feature and a clustering algorithm is performed on the resulting data matrix. The results of these character-based cladistic methods are highly dependent on the clustering algorithms and dendrogram selection criteria used. Most of these tree-building

methods produce one single tree that contains all candidate languages (e.g. Felsenstein 2001). That is, the assumption that all “leaf” languages are related at some level is inherent to many character-based clustering methods. These standard clustering methods were, however, used by Dunn et al. (2005) to detect a phylogenetic signal among groups of Papuan languages. Their study did not assume at the onset that all candidate languages were related, but their test of deep relationships by means of a single unrooted parsimony tree was made possible by the unique geography of the archipelago where these languages are found.

In the case of California languages, where not only the composition of genealogical groupings but also the very existence of large genealogical categories is in question, the best representation of historical relationships may involve neither a single family tree nor an obvious spatial pattern. The geography of this region does not provide the same distance-based information that can be used to interpret the single-tree results of cladistic models like the one employed in the Papuan investigation of Dunn et al. (2005). Lexical data, however, prove to be a useful resource for testing the significance of relationships in an area like California, since they can be used with a variety of statistical methods that do not assume a single tree. The relative simplicity of these methods allows for a certain amount of adaptability, whereby a researcher can tailor the statistical procedure to their assumptions and research questions.

The theoretical reasoning for using lexical data in a study of California linguistic relationships is balanced by equally important practical considerations. The indigenous languages of California are all either endangered or extinct, and for some of them the documentation that exists is limited. The amount of grammatical analysis that has been done for these languages is also highly variable, and successive waves of theoretical developments in linguistics during the period of documentation have resulted in analyses that vary in terminology, focus, and comprehensiveness. Word lists are often the first materials gathered in the field, and word lists exist for several California languages that lack extensive and detailed grammatical descriptions. Furthermore, the creation of standard word lists by Morris Swadesh (1952; 1955) and the Survey of California and Other Indian Languages has resulted in the collection of many of the same lexical items for a wide variety of languages in this region. Whereas the coverage and reliability of grammatical descriptions varies widely among California languages, Swadesh or Survey word lists are available for most of these languages.

3.3.2 Methodology

The metric developed by Kessler (2001) is a promising statistic for measuring the strength of the relationships between languages by comparing word lists. This measure, called R^2 by Kessler,¹ quantifies the amount of recurrence of sound correspondences between a pair of languages. The principle that is at the heart of Kessler’s R^2 is drawn from traditional historical methods. Specifically, the use of recurrent sound correspondences as an indicator of historical relationships in this computational method echoes the use of systematic sound correspondences in traditional historical linguistics to identify sound changes.

In a nutshell, Kessler’s R^2 method operates on a contingency table, whose values quantify the correspondences between initial sounds in pairs of languages. Sounds from each of the

¹This metric will be referred to henceforth as “Kessler’s R^2 ” to distinguish it from the coefficient of determination statistic, conventionalized in the field of statistics as R^2 .

two languages under consideration are represented along the X and Y axes respectively. The contingency table is populated with counts of the frequency of correspondence² of each possible pair of sounds. Because pairs that occur only once are not indicative of language relatedness, but rather the simple fact that each word in language 1 must correspond to a word in language 2, each non-zero cell must be reduced by one so that all non-zero values represent linguistically significant pairs. The values in each cell are squared in order to weight correspondences that exhibit greater recurrence, and to facilitate numerical analyses of their significance. The sum of the values in all cells of the table is Kessler’s R^2 for the associated pair of languages. The process can be repeated for every pair of languages in the sample and can be iterated in a Monte Carlo style significance test, which uses random re-sampling of the dataset to ensure the statistical robustness of the results. The method yields a set of relationship strength values with associated p-values representing their statistical significance.

The input data for this procedure is a set of word lists that contains basic lexical items in each language that correspond to a standard set of meanings. The construction of the meaning list is a delicate task – meanings must be cross-linguistically salient, resistant to replacement or change over time, and faithful to the Saussurian arbitrariness hypothesis (e.g. word list items should not be meanings that are frequently represented with sound symbolism). For this study, items from the Swadesh 200 meaning list (Swadesh 1955) were used. The use of this well-established list capitalizes on the work already done by Swadesh to control for the stability of meanings, but the Swadesh 200 list is not a perfect fit for the languages in this study. Further discussion of the word list data can be found in section 3.4.2.

Though it is possible to align and compare multiple segments for each word, this paper only uses the initial segment of each item, following Kessler’s observation (2001) that initial segments produce results as least as satisfactory as those derived from multiple segment matching. Initial segments are particularly perceptually salient, and because of this are more likely than other parts of the word to be preserved over time. The salience of initial segments plays a role in phonetics and phonological theory. Onsets and root-initial syllables are often resistant to phonological processes like assimilation and neutralization due to their salience (Beckman 1998), and it has been hypothesized that the preservation of onset information affects the timing of articulatory gestures (Chitoran et al. 2002). Support for the importance of word-initial sounds can also be found in psycholinguistics, where the importance of word onsets for lexical recognition has been demonstrated experimentally (e.g. Gow et al. 1996). Since word-initial sounds are perceptually important, we can expect them to be at least as diachronically stable as any other segment in a word. Using the initial segments of list items as the domain in which to search for sound correspondences, then, should produce similar results to the more complex process of mapping out word-internal correspondences and is certainly a more economical approach.

After reducing each item in the word list to a single segment, the first step of the process is to construct a contingency table for each pair of languages. Each possible initial sound for one of these languages is listed on the X axis, with all possible initial sounds for the other language on the Y axis. The internal cells of this table are populated by comparing the

²e.g. if “t” in language 1 corresponds to “d” in language 2 seven times, the cell in the contingency table corresponding to X = t, Y = d would contain the value 7

word list data for the pair of languages under consideration, one meaning at a time. Each time a particular pair of segments occurs, the corresponding cell in the contingency table is incremented by one. At the end of this procedure, the contingency table will contain counts of occurrences of sound pairs, and the sum of all cells will equal the number of items in the meaning list.

These counts already contain information about the recurrence of sound correspondences and, therefore, about the strength of the relationship between the current pair of languages. Higher levels of sound correspondence, and thus larger numbers in the contingency table cells, are expected in languages that are related. This expectation is based upon the way languages develop. Daughter languages of some parent language which have undergone phonological change will contain reflexes of the parent language’s sounds. These will show a regular, repeated pattern of correspondence that reflects the environments in which sound changes occurred. Languages that are more distantly related, and have thus undergone more sound changes, also contain regular, recurrent sound correspondences, but as several generations of sound changes, each with their own environments, are layered upon one another, the pattern of sound correspondences in the descendant languages will become patchier, and the number of times a given sound correspondence will recur will decrease. Recurrence counts are thus expected to be greater for shallow relationships than for distant ones.

The remainder of Kessler’s R^2 procedure serves to capture the relevant information about the distribution of items throughout the table in a single metric. Consider the sample distributions in table 3.2, adapted from Kessler (2001:149).

	a	b	c
A	11	11	11
B	11	12	11
C	11	11	11

(a) Scenario A

	a	b	c
A	21	6	6
B	6	22	6
C	6	6	21

(b) Scenario B

Table 3.2: Contingency tables representing different distributions of 100 items

In each of these constructed contingency tables, 100 items are distributed across a nine-cell table. The first of these tables contains a roughly even distribution across the cells, while the pattern in the second table shows more recurrent correspondences between A and a, B and b, and C and c. The amount of recurrence of sound correspondences in scenario B is much greater than in scenario A, though the sum of all of the cells for each table would yield the same number. To capture the amount of recurrence in a single value that can differentiate between these distributions, we must diminish the influence of less-recurrent sound correspondences and accentuate the recurrent sound correspondences.

The next step in calculating Kessler’s R^2 is to subtract 1 from all cells that contain a non-zero value. This eliminates sound correspondences that occur only once. Since these non-recurrent sound correspondences are statistically as likely to occur by chance as through a historical relationship, they cannot be counted as evidence for relatedness. Subtracting 1 will thus eliminate all data that does not represent a recurrent sound correspondence and thus cannot be counted as evidence for relatedness. This subtraction will have little impact on the

large numbers that represent frequent recurrence, and because the same amount is subtracted for all recurrence counts in every contingency table, its effect on the overall outcome of the test is negligible. The sample distributions introduced in table 3.2 do not contain any non-recurrent correspondences which this step would eliminate, but the subtraction step is performed anyway. The results of this step are shown in table 3.3.

	a	b	c
A	10	10	10
B	10	11	10
C	10	10	10

(a) Scenario A

	a	b	c
A	20	5	5
B	5	21	5
C	5	5	20

(b) Scenario B

Table 3.3: Contingency tables with 1 subtracted from each non-zero cell to eliminate non-recurring correspondences

To give weight to correspondences that exhibit greater recurrence, the value in each cell is squared. The highly recurrent sound correspondences that are magnified the most by this step are the types of patterns we expect to see between two related languages. Weighting more recurrent correspondences by squaring their count values also allows the procedure to measure the strength of relationships, since the sound correspondences in closely related languages, upon which fewer sound changes have operated, are expected to recur more frequently than those in more distantly related languages that have undergone more changes since they diverged. The results of this weighting step are shown below in table 3.4 for the example data. After squaring the cell values, the differences between scenario A and scenario B are amplified, and the sums of each table's cells are no longer equal.

	a	b	c
A	100	100	100
B	100	121	100
C	100	100	100

(a) Scenario A

	a	b	c
A	400	25	25
B	25	441	25
C	25	25	400

(b) Scenario B

Table 3.4: Contingency tables with recurrence values squared

At this point the relationships that are expressed by these tables can be summarized in simpler terms. Kessler's R^2 , which serves as a simple numerical representation of the strength of the relationship, is calculated by summing all of the squared recurrence values in the table for a particular pair of languages. For the two sample data distributions this produces the following Kessler's R^2 values:

- Scenario A Kessler's $R^2 = 921$
- Scenario B Kessler's $R^2 = 1391$

The differences between these two distributions that were impressionistically obvious in the original contingency tables (table 3.2) are represented quantitatively by these two Kessler's R^2 values. The higher level of recurrence in Scenario B is manifested by its higher Kessler's R^2 value.

The computation of Kessler's R^2 targets information that is very similar to that which Dixon and Kroeber intended to employ in their 1919 paper (see Table 3.1). However, where Dixon and Kroeber relied on raw counts of impressionistic similarities between languages, Kessler's R^2 uses a well-defined, objective procedure for comparing language data and quantifies the trends it identifies with a more sophisticated statistic.

3.3.3 Statistical Significance

A measure of statistical significance for Kessler's R^2 values can be computed through a Monte Carlo simulation (Kessler 2001). Kessler's R^2 was constructed to measure a linguistically important characteristic of meaning list datasets (i.e. sound correspondence recurrence) that conventional statistics are not well suited for (i.e. sound correspondence recurrence). Unlike standard inferential statistics (e.g. χ^2), Kessler's R^2 is not associated with a well-understood probability distribution, so an independent measure of statistical significance must be used to test the null hypothesis (that a given Kessler's R^2 value would occur in the absence of a historical relationship between languages). The Monte Carlo portion of the methodology generates a random sample of Kessler's R^2 values for a dataset that can be used to identify the Kessler's R^2 values that are greater than would be expected due to chance.

The Monte Carlo simulation for this statistic involves rearranging the word list data randomly and recomputing Kessler's R^2 . When this procedure of reorganization and retesting is iterated thousands of times, the result is a very large set of Kessler's R^2 values that represent chance pairings of words in the word list. By comparing the actual results to this sample that is representative of chance, a p value can be computed. The concept of shuffling data to test the probability of lexicostatistical findings was, in fact, introduced much earlier by Robert Oswalt (1970), who proposed a procedure of shifting word list data (still in its original order) one line at a time and recomputing the relevant statistic for each rearrangement of the data. Limited by the technology of the mid-twentieth century, Oswalt's permutation procedure produced only as many rearrangements of the data as there were meanings in the input word list, resulting in a "chance" sample too small to yield acceptably precise p values. More importantly, shifting the entire word list without scrambling the data did not result in a real "chance" outcome for any of these iterations. Modern computing has made it possible to reorganize data in a truly random manner and to compute the statistic quickly enough that it is possible to carry out the thousands of iterations necessary for calculating meaningful p values.

The Monte Carlo portion of this methodology begins with a random reorganization of the word list for one of the languages under consideration, as demonstrated in table 3.5. In table (a), the data is arranged in its original order. The "a" in the first row of data is the first segment of the word in language 1 with meaning 1, and the "B" in the same row is the first segment of the word with the same meaning. In the second row, "c" is the first segment of the language 1 word and "A" is the first segment of the language 2 word for meaning 2.

The data in the rest of the table similarly represent the initial segments of words in these two languages, sorted by meaning. In table (b), the data for language 2 has been shuffled, so that it no longer arranged by meaning. Notice that the data for language 1 is the same in table (b) as in table (a). The data for language 2 in table (b), on the other hand, has the same number of “A”, “B” and “C” segments as in table (a), but these segments have been randomly reordered.

Meaning	Lang. 1	Lang. 2	Meaning	Lang. 1	Lang. 2
1	a	B	1	a	C
2	c	A	2	c	B
3	a	C	3	a	B
1	b	B	1	b	A
2	c	B	2	c	A
3	b	A	3	b	B
1	b	B	1	b	C
2	c	C	2	c	A
3	b	C	3	b	B
10	a	A	10	a	C

(a) Original Data
(b) Shuffled Data

Table 3.5: Random rearrangement of simulated word list data

The reshuffled data are next used to populate a table of sound correspondence counts and compute Kessler’s R^2 , just as was done for the original, unscrambled data. For this paper, this process of randomly reordering the word list data and recomputing the Kessler’s R^2 value was repeated 10,000 times for each pair of languages. As these 10,000 random Kessler’s R^2 values were computed, they were each compared to the actual Kessler’s R^2 value for the unscrambled data. A tally was kept of the number of times the random Kessler’s R^2 value exceeded the actual Kessler’s R^2 value, and the p value was calculated by dividing this tally by the number of iterations in the Monte Carlo test. In this study, 10,000 random rearrangements of the data were used in the Monte Carlo simulations, so the p value for a particular comparison between two languages is thus equal to the number of times the actual Kessler’s R^2 exceeded the Kessler’s R^2 value for any round of the Monte Carlo test, divided by 10,000. If 5,412 of the Monte Carlo-generated Kessler’s R^2 values exceed the actual value, the p value for that test is 0.5412. If only 98 of the Kessler’s R^2 values in the random sample exceed the actual Kessler’s R^2 value, the p value for the test will be 0.0098. Thus, smaller p values correspond with greater levels of statistical significance, and the p value for a statistically significant correspondence will be smaller – frequently several orders of magnitude smaller – than the p value for a language pair that exhibits no significant relationship.

3.3.4 Comparing Kessler’s R^2 across Languages with Different Phonologies

The methodology outlined up to this point has followed the prescription of Kessler (2001) quite closely. However, differences between the languages of California and the Indo-European languages upon which Kessler tested this method require this paper to depart here from the procedure he outlines. The axes of a contingency table represent the phonemic inventories of the languages in a study, causing the overall size of the contingency table to vary with the size of the languages’ phonemic inventories. Because each word list contains the same number of items, the same total number of sound correspondences will be distributed throughout each of the contingency tables, which vary in size. This becomes problematic in a region like California, where differences in the size of the phonemic inventories of languages can be striking. For example, a contrast between plain, aspirated, and glottalized stops occurs in many languages of California (Golla 2011), resulting in consonant inventories that can be several times the size they would be without these contrasts. The other end of the spectrum of phonemic inventory sizes is also found in California. The Uto-Aztecan languages on the eastern side of the Sierra Nevada do not distinguish between these three different types of stop consonants, and their phonemic inventories are generally much smaller than, for example, the Palaihnihan languages which do have this phonemic contrast.

As table 3.6, below, demonstrates, when an equal number of items are distributed evenly in tables of different sizes, different Kessler’s R^2 values will be computed for each table. With 100 items distributed across each table and a maximally even distribution in each case, the Kessler’s R^2 value for the smaller table is several times as great as the result for the larger table. This difference in the Kessler’s R^2 values is an artifact of the differences in phonemic inventory size, rather than an indication of different strengths of relationships. In order to compare the strength of relationships in a sample of languages with very different phonemic inventories, these Kessler’s R^2 values must be normalized in a way that filters out the effects of phonemic inventory size.

	a	b	c		a	b	c	d	e
A	11	11	11	A	4	4	4	4	4
B	11	12	11	B	4	4	4	4	4
C	11	11	11	C	4	4	4	4	4
				D	4	4	4	4	4
				E	4	4	4	4	4

(a) Smaller phonemic inventory (Kessler’s $R^2 = 921$)

(b) Larger phonemic inventory (Kessler’s $R^2 = 225$)

Table 3.6: Even distributions of 100 items in contingency tables of different sizes

One of the strengths of the simple Monte Carlo methodology Kessler recommends for testing the significance of the relationships measured by his R^2 metric is that the word list data used for this process acts simultaneously as a lexical sample and a sample of the patterns of the language’s phonology. Instead of independently calculating a set of probabilities of

various sounds occurring word-initially, the Monte Carlo procedure outlined in the previous section captures the size of a language’s phonological inventory and serves as a sample of the frequency with which its phonemes occur in the initial position. Because the Monte Carlo simulation accounts for the size and variation of the phonemic inventory, the chance measures of Kessler’s R^2 it generates can be used to normalize the actual Kessler’s R^2 data.

One simple way to do this is to express the ultimate measures of relationship strength as ratios. The contingency table for the actual Kessler’s R^2 calculation will be equal in size to those used for each round of the Monte Carlo simulation. The effects of phonological inventory size on the calculation of Kessler’s R^2 will be very similar for the actual data and each Monte Carlo iteration’s scrambled data. To control for the effects of phonemic inventory size, the actual Kessler’s R^2 value can be divided by the mean of the Kessler’s R^2 values generated in the Monte Carlo test.

During each iteration of a Monte Carlo simulation, the contingency table will be populated by the same number of “A”, “B”, “C”, and “D”, and “a”, “b”, “c”, and “d” segments that are in the original data in the original dataset (in this case, table 3.6). Thus, even though the Kessler’s R^2 values that are calculated in each round of the Monte Carlo simulation will be different, they will all be constrained by the number and distribution of initial segments that are in the dataset. When the data in table 3.6 is rearranged, it will result in contingency tables like those in table 3.7.

	a	b	c		a	b	c	d	e
A	9	14	10	A	7	4	3	2	4
B	12	11	11	B	3	6	4	6	1
C	12	9	12	C	4	2	6	4	4
				D	3	4	5	3	5
				E	3	4	2	5	6

(a) Smaller phonemic inventory (Kessler’s $R^2 = 941$)

(b) Larger phonemic inventory (Kessler’s $R^2 = 279$)

Table 3.7: Contingency tables for randomly reorganized simulated data

The Kessler’s R^2 values for these tables are 941 for the smaller phonemic inventory and 279 for the larger phonemic inventory. Rescrambling the data and recalculating Kessler’s R^2 9,999 more times would give us 10,000 different and random but similarly constrained Kessler’s R^2 values, from which a mean random Kessler’s R^2 value could be computed. Suppose that the mean Kessler’s R^2 values that resulted from a Monte Carlo simulation for these data were 1018 and 251. The normalized recurrence values would be 0.905 (= 921/1018 or actual/simulated mean) for the smaller phonemic inventory condition and 0.896 (= 225/251 or actual/simulated mean) for the larger phonemic inventory condition. These ratios (i.e. 0.905 and 0.896) are more similar than the raw Kessler’s R^2 values (i.e. 921 and 225) for these tables. Because this method is stochastic in nature, however, slightly different values will be computed any time the test is run, and comparable datasets will not always result in exactly equal normalized recurrence values. The use of a very large number of iterations in the Monte Carlo test will minimize the variation in the results.

3.3.5 Multilateral Clustering

When the above procedure is repeated for each pair of languages, the output is a set of relationship strength values and corresponding p values for every pair of languages. Table 3.8 is an example of what this output might look like, where L1, L2, L3, and L4 are the languages being tested and an asterisk (*) represents a statistically significant p value (i.e. $p < 0.005$).

	L1	L2	L3	L4
L1		5.936*	1.127*	0.693
L2	5.936*		0.991	1.168
L3	1.127*	0.991		2.727*
L4	0.693	1.168	2.727*	

Table 3.8: Sample table of normalized recurrence metrics

This table measures the strength of the pairwise relationships between languages and indicates their statistical significance. A clustering method must be applied to these data to build phylogenetic trees from them. One option, which will be discussed in section 3.5.3, is to use the significant recurrence values from such a table as a similarity matrix for a standard average linking clustering method. In such a method, the relationships identified in a single round of binary comparisons are used to infer the structure of the tree. Another option is to collapse the dataset as relationships are built up, treating clusters as single languages and repeating the Kessler’s R^2 /Monte Carlo procedure every time a language is incorporated into a cluster, until no significant relationships are found. This latter method is advocated by Kessler (1999; 2001). He argues that as a way of detecting distant relationships, this approach achieves the power that Greenberg (1987) ascribed to multilateral comparison, and it also institutes a rigorous control for statistical significance.

To begin the clustering procedure, the strongest binary relationship that is statistically significant at an acceptable level³ (for this investigation, $p < 0.005$) is identified, and the two languages that bear this relationship are united to form the first cluster. Clustering creates a parent node in the dendrogram and requires that the two-dimensional array describing the relationship between these two languages be collapsed into a one-dimensional (i.e. 1 column) vector of unique sound pairs. The data vectors that represent clusters of languages have actually collapsed multiple dimensions of data down to one dimension, and are treated just like any other language’s word list for the next round of comparisons. Whereas the data for any individual language contains a string of individual initial segments (e.g. A, B, C, etc.), the data for the new cluster will contain multiple initial segments (one per language) in each

³The significance level of $p < 0.005$ was chosen somewhat arbitrarily, as a sort of Bonferroni correction for the problem that a significant relationship is more likely to be found by conducting multiple comparisons (e.g. comparing Shasta to each of the languages in the sample) than a single comparison (e.g. comparing Shasta only to Mono) and as the groups are built this becomes more and more likely to lead to false positives. It should be noted that running the test with a less stringent significance level of $p < 0.05$ differs from the results reported in section 3.5 only in that it produces a spurious relationship between Maiduan and Diegueño instead of Cocopa and Diegueño in the tenth round of clustering.

position in the meaning list (e.g. A & a, B & a, C & a, etc.). This data, along with the word list data for any other clusters and all remaining unclustered languages serves as the input for the next round of Kessler’s R^2 and Monte Carlo procedures.

To illustrate, if the languages represented by table 3.9 (repeated from 3.2b) were found to have the most significant relationship in a round, the contingency table for the resulting cluster’s comparison to a third language might look like table 3.10, where the row headings are one-dimensional arrays that consist of combinations of initial segments from languages 1 and 2. The value of 7 in table 3.10b, column 1, row 1, for example, means that in 7 instances, initial segment x in language 1 corresponds with A in language 2 and a in language 3.

	a	b	c
A	21	6	6
B	6	22	6
C	6	6	21

Table 3.9: Contingency table for simulated languages 1 and 2

	x	y	z
A,a	7	9	5
A,b	1	3	2
A,c	4	1	1
B,a	5	0	1
B,b	9	4	9
B,c	2	3	1
C,a	4	1	1
C,b	1	2	3
C,c	4	8	9

Table 3.10: Contingency table for a simulated comparison between a cluster (languages 1 and 2) and language 3

This contingency table would be used to calculate a Kessler’s R^2 value, and then the word list for language 3 would be randomly rearranged during the Monte Carlo simulation to test significance and normalize the recurrence value. Each round of clustering continues by grouping the languages or clusters with the strongest significant relationship together. Every time a language is added to a cluster, the dataset for that cluster will be expanded so that the array in each position of the word list contains the initial segment of each of the languages in that cluster. This process continues, building up family trees round by round, until no relationships can be found with a p value less than 0.005.

When this program of clustering and retesting data exhausts the statistically significant relationships, it will return a dendrogram or set of dendrograms that represent the relationships between the input languages, provided that any significant relationships have been

found. If no relationships have a p value under 0.005, no dendrograms will be returned. If significant relationships are found between some, but not all, of the input languages, the output dendrogram(s) will contain only those languages involved in significant relationships; all other languages will be left unclustered and considered unrelated.

3.4 Applying this method to California

The methodology outlined in section 3.3 is a promising tool for investigating the relationships between California languages for two primary reasons. First, the lexical data it employs can be gathered from published and archival sources for many California languages. Secondly, the clustering component of the method compares the data of all languages of a given cluster to each of the other candidate languages or clusters. This gives the procedure the potential to identify the sort of distant relationships that have been hypothesized to exist in California.

3.4.1 The Language Sample

Ideally, an investigation of the deep relationships between California languages would consider data from every language in the California region for which there is documentation. The selection of a subgroup of California languages to serve as a sample for this study was motivated by both matters of research design and practical considerations. Because Kessler's R^2 method as implemented in this chapter has only been tested on Indo-European languages,⁴ this paper serves as a test of the procedure as well as a test of the linguistic relationships in the region. For this exploratory investigation, the advantage of considering every documented language of the region was made prohibitively impractical by the difficulty and time consuming nature of extracting lexical data from archival resources.

To carry out the dual purpose of testing the applicability of Kessler's R^2 method in this region and investigating the remote relationships that may exist between the languages of the Hokan and Penutian stocks, this study considered a sample of 21 languages from the California region and the American Southwest. The sample was constructed to include languages from the Hokan and Penutian stocks and the Uto-Aztecan family.

The Hokan stock is represented by multiple members of the Palaihnihan, Pomoan, and Yuman families and other individual Northern California Hokan languages. This group includes geographically and genealogically diverse members of the Hokan category. Penutian is represented by Miwok and Maiduan languages and the individual language Wintu. This sample is intended to be both broad, in that it contains members of three of the primary California Penutian families, and deep, in that it contains multiple members of some of these families. However, it is admittedly limited in both of these dimensions as a result of the relatively small number of Penutian languages for which reliable data could be readily compiled. The third major group represented in the sample is Uto-Aztecan, a family which

⁴The general framework of this method, including the Monte Carlo and multilateral clustering procedures, was utilized by Kessler and Lehtonen (2006) in an investigation of the Indo-Uralic question, but instead of using Kessler's R^2 as the measure of relatedness that study used a measure of similarity based on place of articulation of initial segments.

extends into California east and south of the Sierra Nevada and also stretches throughout much of the Great Basin and into Mexico. The general structure of the Uto-Aztecan family tree has been established through traditional historical methods and is widely accepted by historical linguists who work on American languages(e.g. Campbell 1997; Mithun 1999; Hill 2001). These languages provide a means of assessing the function of the statistical program, by comparing its results with known Uto-Aztecan relationships. Uto-Aztecan languages from outside California were included in order to provide multiple layers of branching and more temporal depth in this “control” group’s family tree.

Languages included in the test sample

Hokan	Penutian	Uto-Aztecan
Karuk language ⁵	<i>Miwok</i>	Hopi language ²¹
Yana language ⁶	– Lake Miwok ¹⁶	<i>Takic</i>
<i>Shasta-Palaihnihan</i> ⁷	– Northern Sierra Miwok ¹⁷	– Luiseño ²²
– Shasta ⁸	<i>Maiduan</i>	<i>Numic</i>
– <i>Palaihnihan</i>	– Maidu ¹⁸	– <i>Southern Numic</i>
– – Achumawi ⁹	– Nisenan ¹⁹	– – Kawaiisu ²³
– – Atsugewi ¹⁰	<i>Wintuan</i>	– – Ute ²⁴
<i>Pomoan</i>	– Wintu ²⁰	– <i>Central Numic</i>
– Kashaya ¹¹		– – Tümpisa Shoshone ²⁵
– Northern Pomo ¹²		– <i>Western Numic</i>
– Southeastern Pomo ¹³		– – Mono ²⁶
<i>Yuman</i>		
– Cocopa ¹⁴		
– Diegueño ¹⁵		

⁵Bright and Gehr (2008); Bright (1957)

⁶Sapir and Swadesh (1960)

⁷Shasta-Palaihnihan is a controversial grouping.

⁸Olmsted and Bright (1959); de Angulo and Freeland (n.d.); Silver (1961)

⁹Olmsted (1966)

¹⁰Olmsted (1984)

¹¹Oswalt (1957, n.d., 1967)

¹²Oswalt (n.d., 1967); Vihman (1970)

¹³Moshinsky (1965); Haynie and Kelsey (2007)

¹⁴Crawford (1989)

¹⁵Couro and Hutcheson (1973)

¹⁶Callaghan (1965)

¹⁷Callaghan (1987)

¹⁸Shipley (1963)

¹⁹Uldall and Shipley (1966); Sawyer (n.d.)

²⁰Pitkin (1985)

²¹Hill (1998)

²²Bright (1968)

²³Zigmond et al. (1990)

²⁴The Southern Ute Tribe (1979)

²⁵Daley (1989)

²⁶Lamb (1958); Bethel et al. (1993)

3.4.2 Word List Data

The list of meanings that was gathered for each language is derived from the Swadesh 200 meaning list (Swadesh 1952). Though Swadesh lists have become a common tool for lexical elicitation in the field, not all of the field notes and dictionaries that are available for California languages include all of the meanings in the 100 or 200 item Swadesh lists. Other standardized lists, such as the Survey of California and Other Indian Languages' topical word list²⁷, were more common in the mid-twentieth century documentation for this region. Thus to maintain a list length of at least 100 items and to take advantage of the work done by Swadesh to control the diachronic stability and semantic universality of word list items, the 200 word list was chosen as a starting point and the list was edited based upon the availability and consistency of data.

During the process of gathering word list data from published dictionaries and archived field notes, the list was pared down to include only 102 meanings. Meanings were excluded from the list for several reasons. Most of the items that were discarded are meanings for which no basic term could be found in at least one of the languages. In many cases these words were absent altogether from the documentation. For example, words for *vomit* and *louse* could not be found for multiple languages. It is possible that elicitation and/or publication of these terms was avoided by some researchers because they were deemed to be vulgar or impolite. Other missing words include meanings like *road*, which may have been unavailable for cultural reasons, and *egg*, whose absence is probably a simple coincidence. Another category of meanings that were discarded is those for which no general term exists. The meaning *grass*, for example, could be expressed by a number of terms for specific grass species rather than a single general term to refer to a superordinate category of all grasses. These items were removed from the list.

Borrowings and multiple words derived from the same root also reduced the size of the word list. Obvious borrowings, like the Spanish *semilla* for *seed*, were removed from the list. Borrowings of this type were relatively rare and generally reflected the early influence of Spanish on California's indigenous languages. More frequently, words were excluded because they were derived from the same root as another item in the list. The natural opposition between *moon* and *sun* and the clear semantic connection between meanings *water* and *wet* predispose these meanings to being expressed with derivationally related words. In these situations, only one instance of the root was kept in the list. The choice of which item to discard was based primarily on whether one of the items was problematic in a greater number of languages than the other, a consideration which led to essentially arbitrary choices between meanings and allowed the maximum number of non-redundant items to be retained.

Once the 102 item word lists were compiled for all 21 languages in the sample, the next step as prescribed by Kessler (1999; 2001) was to isolate the smallest morphological unit with the intended meaning. Since only the initial segment of each item is used to calculate Kessler's R^2 , only prefixes were removed in this step. All other information beyond the initial segment is discarded before the construction of contingency tables in any case, so no extra effort

²⁷A short California word list that was frequently used for data collection in this era is available online at <http://linguistics.berkeley.edu/survey/resources/fieldwork-tools.php>, but the longer Topical Word List that was also used is now only found in archival collections.

was undertaken to remove suffixes. For the Palaihnihan and Shasta languages, this process was informed by morphological descriptions (de Angulo and Freeland 1930, n.d.; Olmsted 1961). For the Uto-Aztecan languages, instrumental prefixes were removed in consultation with scholars of Numic and Uto-Aztecan languages (M. Houser and M. Toosarvandani, p.c. 2008). Yuman language data were also reviewed in consultation with a scholar of that family (L. Hinton, p.c. 2008). After editing the data in this way, numerical codes were assigned to represent the first segment of each item. Representing the data in numerical form allows it to be manipulated easily in the Matlab program that was written to perform the Kessler's R^2 , Monte Carlo, and clustering procedures. Each consonant in a language's phonemic inventory was assigned a unique code. Vowel quality was ignored, and all vowels in a particular language were assigned the same numerical code.

3.4.3 Predictions

Several possible outcomes are conceivable for this test for relationships among the 21 languages. One conceivable outcome is the production of dendrograms that correspond to the accepted subgroupings for Uto-Aztecan and the identification of significant, high-level relationships between Hokan or Penutian languages. This would constitute strong evidence for deep relationships corresponding to Hokan and/or Penutian.

Another possibility is the identification of the expected relationships among Uto-Aztecan languages but the failure of the program to cluster any of the Hokan and/or Penutian sub-families into higher order stocks. Since the reproduction of expected Uto-Aztecan groupings would indicate that the program is functioning correctly, this outcome would suggest that either no stock-level relationships exist in Hokan and/or Penutian or that the test is unable to detect relationships as deep as ancestral Proto-Hokan or Proto-Penutian would have to be.

A third possibility is that the test might fail to cluster the Uto-Aztecan languages in a way that is consistent with existing reconstructions of the Uto-Aztecan family tree. Regardless of whether high-level relationships between Hokan or Penutian languages were found, this outcome would indicate that the program is not functioning as intended.

3.5 Results

When the statistical program was run on the numerically coded word list data for the 21 languages in the California sample, 11 rounds of testing and clustering were performed. Six individual dendrograms were produced in this process, and four of the languages failed to be incorporated into any clusters.

3.5.1 Clustering

At the end of each round of testing in the procedure one language or cluster of languages was added to another language or cluster, forming a larger grouping and a parent node on a dendrogram. The round by round results of the clustering are shown in table 3.11. Each of these relationships corresponded to the greatest normalized recurrence value (abbreviated NRV in the table below for convenience) identified for its respective round and all of these were found to be significant at the $p < 0.005$ level. Atsugewi, Karuk, Yana, and Wintu were left unclustered at the completion of the test.

Round	Languages Grouped	NRV
1	Maidu + Nisenan	6.008
2	Ute + Kawaiisu	5.934
3	Tümpisa Shoshone + (Ute + Kawaiisu)	4.570
4	Mono + (Tümpisa Shoshone + (Ute + Kawaiisu))	6.892
5	Northern Pomo + Southeastern Pomo	2.8579
6	Kashaya + (Northern Pomo + Southeastern Pomo)	6.711
7	Luiŝeño + (Mono + (Tümpisa Shoshone + (Ute + Kawaiisu)))	2.310
8	Hopi + (Luiŝeño + (Mono + (Tümpisa Shoshone + (Ute + Kawaiisu))))	5.221
9	Lake Miwok + Northern Sierra Miwok	2.203
10	Cocopa + Diegueño	1.716
11	Shasta + Achumawi	1.511

Table 3.11: Round by round building of California phylogenetic groupings

These results can be represented visually in the dendrograms on page 80.

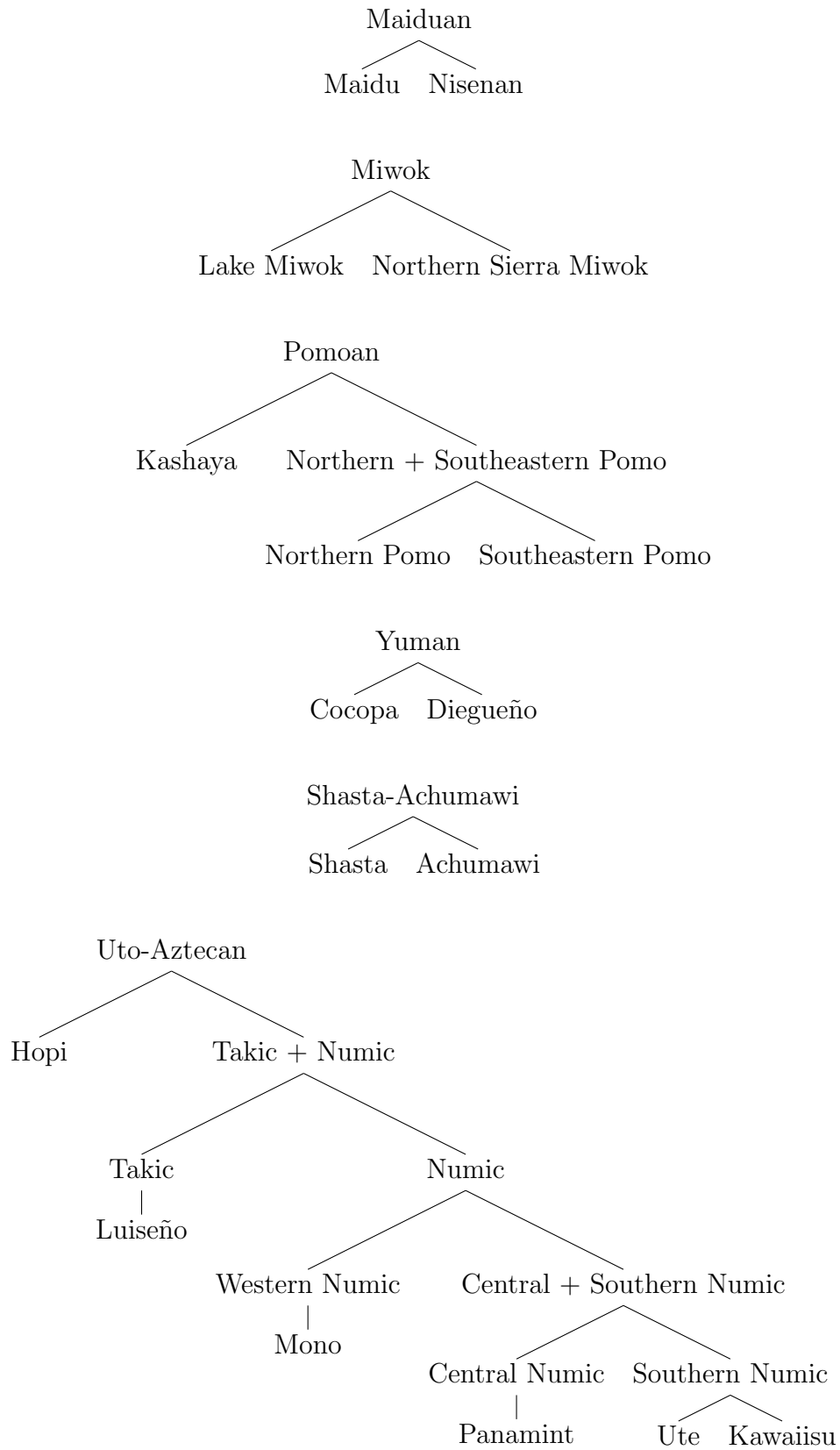


Figure 3.2: Dendrogram representation of identified California language relationships

It is clear from the results for the Uto-Aztecan test group that the general operation of the statistical program is sound. All of the Uto-Aztecan languages in the sample have been clustered into a single group. Although this test identifies only the likelihood of a relationship, and so the subgroupings represented in these dendrograms should not be interpreted as diagnoses of genealogical subgrouping, the branching of the dendrogram generated for these Uto-Aztecan languages is consistent with what is known about the relationships within this family. Ute and Kawaiisu, members of the Southern Numic subfamily, form a small unit. These are clustered with Central Numic's Tümpisa Shoshone and Western Numic's Mono to create a larger Numic category. Numic is then clustered with Luiseño, a representative of the separate Takic branch of Uto-Aztecan. Historical linguists do not consider Takic and Numic together to form a sub-branch of Uto-Aztecan, but a somewhat artificial sub-level is forced in the output of this program, since the method is only capable of producing binary branching. Because the California Numic languages represented in this sample are geographically closer to Luiseño than to Hopi, and thus more likely to have been in contact with Luiseño, the clustering of Luiseño with Numic before the addition of Hopi to this group is not an unexpected outcome for Uto-Aztecan.

Looking at table 3.11 on page 79, another expectation regarding the program's function is borne out. As predicted by Kessler (1999; 2001), the power of this program to identify relationships increases as larger groups are built up. In the fifth round of the test, Southeastern Pomo and Northern Pomo are merged into a cluster, with a normalized recurrence metric of 2.858. In the same round, a slightly weaker relationship is found between Northern Pomo and Kashaya, with a recurrence metric of 2.729. The relationship between Kashaya and Southeastern Pomo is not found to be statistically significant in this round. The subsequent round of testing, in which the data for Northern Pomo and Southeastern Pomo are combined and treated like a single language, groups Kashaya with the Northern/Southeastern Pomo cluster based on a recurrence metric of 6.711. The apparent strength of the relationship between Kashaya and the cluster of other Pomoan languages is greater than the strength of Kashaya's relationship with either of the individual languages that comprise that cluster. The relationship is also found to be significant, even though a significant binary relationship between Kashaya and Southeastern Pomo was not detected in earlier rounds. Most of the family trees in the output are either too shallow or too strong to show evidence of this increased sensitivity of larger clusters to subtle evidence of relationships. Hopi, however, also exhibits this tendency to bear a stronger relationship to a cluster than to its component languages and the ability to be clustered with a group that contains languages with which it does not show a significant pairwise relationship.

3.5.2 Unexpected Findings

The ability of this program to identify the relationships between Uto-Aztecan languages and the lack of support for the Hokan and Penutian stocks are not very surprising results, given linguists' acceptance of Uto-Aztecan and the uncertainty of Hokan and Penutian. Some of the smaller groupings in the output of the program were less anticipated.

Oswalt's exploration of the relationships between Pomoan languages, which is itself based primarily on a lexicostatistical study (1964:412), groups both Kashaya and Northern Pomo

into the Western Pomo subgroup, while Southeastern Pomo forms its own branch, descended directly from Proto-Pomo. This subgrouping has become the most widely-accepted analysis of the Pomoan family’s internal relationships. The order in which Southeastern Pomo and Northern Pomo were grouped with Kashaya in this study, however, is the opposite of what Oswalt’s Pomo family tree would predict. It fits much better with Kroeber’s classification of the Pomoan languages (1925:227), which groups Kashaya and Southern Pomo under one node and the five other Pomoan languages – including Southeastern Pomo and Northern Pomo – directly under a sister node. Kroeber’s groupings are presented without any specific discussion of the evidence for the branching structure, and is contradicted not only by Oswalt’s work, but also by Halpern’s proposals for the internal structure of the family (1964). The Pomoan grouping produced by this program should not necessarily be considered evidence for Kroeber’s otherwise unsubstantiated Pomoan classification, however.

As noted previously, significant Northern Pomo-Kashaya and Northern Pomo-Southeastern Pomo relationships were both found in the round during which Pomoan grouping was initiated. The difference between the measured magnitudes of these relationships (2.729 and 2.858 respectively) by which the Northern-Southeastern Pomo relationship was selected first for clustering is relatively small. This difference could reflect a sound correspondence recurring just a few times more in one language than the other. Had fewer or different items been culled from the original 200 meaning list, the outcome for Pomoan might have been more consistent with Oswalt’s classification. Additionally, the three Pomoan languages in this study sprout from different parts of the Pomoan family tree, but the four other languages that flesh out the branching of that tree are absent from this test. Kashaya, like Northern Pomo, is part of the Western branch, but unlike Northern Pomo Oswalt places it under a Southern Group sub-node of that Western branch, along with its sister Southern Pomo. Including Kashaya’s Southern Group sisters in the study could also conceivably have affected the outcome. Until a more complete set of Pomoan languages is tested with this method, the results should be treated with caution.

Another surprising result is the failure of Atsugewi to group with Achumawi, either as a binary cluster or as part of the Shasta-Achumawi group that is created in the final round of the program. The relationship between these two languages has not otherwise been cast into any doubt. Again, looking at the full set of round by round results sheds some light on the issue. The eleventh and final round of the program, which grouped Achumawi with Shasta, produced a recurrence metric value for the Atsugewi–Achumawi relationship only slightly lower than the selected Achumawi-Shasta value (1.129 to be exact). However, the corresponding p value that was generated through Monte Carlo resampling (0.183) was too great for this relationship to be considered significant. In 1,828 of the 10,000 randomly resampled iterations of Kessler’s R^2 test, the R^2 value for the Achumawi-Atsugewi relationship was exceeded by the recalculated value. In other words, it is not a lack of recurrent sound correspondences that prevents Achumawi and Atsugewi from being grouped together. Instead, it is the recurrence of sound correspondences in the “chance” resamplings of the data that disqualify the Achumawi-Atsugewi relationship for the clustering procedure. Some characteristic of the datasets must bias the data toward highly recurrent correspondences between sounds, even when randomly rearranged.

The problem stems from the domination of the Atsugewi, Achumawi, and Shasta word

lists by only one or two of their possible initial sounds. Out of 102 meanings, Achumawi contains 57 vowel-initial roots, including most of its verbs. Shasta contains 44 items that begin with a glottal stop and 39 vowel-initial roots. Verb roots in Shasta are also very likely to begin with a vowel. Atsugewi's imbalance is more moderate, with 30 vowel-initial roots and no particular tendency for verb roots to be vowel-initial. These unbalanced distributions of sounds will bias the data, even when randomly shuffled, toward containing highly recurrent sound correspondences. This predisposition undermines the significance of the Achumawi-Atsugewi relationship, since the correspondences that are created by chance are often as highly recurrent as those in the original data.

Random rearrangements of data in the Shasta and Achumawi comparison must also have created highly recurrent sound correspondences, but the significance of this relationship is salvaged by the tendency in both of these languages for verb roots to begin with a vowel. Many of the vowels in the data for these two languages will be aligned in positions that correspond to verb meanings in the original list. This pattern is strong enough to cause greater recurrence in the original dataset than in the vast majority of the randomly resampled versions. The fact that so many verb roots in both Shasta and Achumawi begin in vowels is a possibly coincidental reflection of the morphological patterns of each language. Both of these languages form verbs using obligatory prefixes, and many of the verb prefixes in each of these languages consist of one or more consonants (de Angulo and Freeland 1930, n.d.). Atsugewi's verb prefixes, on the other hand, more frequently contain vowels (Olmsted 1961).

To prevent this type of false negative in situations where a small number of sounds occur initially in most of the list items, future implementations of this methodology should consider coding all phonemically differentiated vowels with different codes. Type II errors like this may be harder to avoid when dealing with languages that have small phoneme inventories and restrictive phonotactics. In such cases considering more than one segment of each item might improve the results.

The final result that may not have been entirely expected is the fact that the relationship between Shasta and Achumawi was found to be significant. The relationship between Shasta and the Palaihnihan languages has been a topic of debate throughout the twentieth century. Dixon proposed a genealogical Shasta-Achumawi stock early in the history of California linguistics (1905). Then Olmsted's phonological investigations in the 1950's concluded that the evidence for grouping of Shasta with Palaihnihan was insufficient (Olmsted 1956, 1957, 1959). Haas responded with arguments for the Shasta-Palaihnihan group (1963), and Kaufman also looked favorably on the hypothesis (1988). The wider community does not seem entirely convinced by the arguments that have been put forth in support of the Shasta-Palaihnihan group. Though the significance level issues discussed above interfere with the creation of a tidy Shasta-Palaihnihan family tree, the clustering of Shasta with Achumawi in this test does constitute evidence for a link between Shasta and Palaihnihan. Future extensions of this project that use an expanded set of vowels for data coding should present clearer results for Shasta-Palaihnihan, but even this preliminary evidence warrants another look at Haas and Olmsted's arguments.

3.5.3 Evaluating the Results

One of the primary goals outlined in Kessler’s (1999) dissertation, in which the basic procedure used in this study was developed, was to implement a statistically controlled version of Greenbergian multilateral comparison. The question of how well the multilateral component of the procedure has worked is difficult to assess by looking only at its results. But a comparison between the results of this study and the output of a clustering algorithm that operates only on binary relationship measures can help in this evaluation.

At the end of the first round of recurrence calculations and Monte Carlo simulations, the program used for this study returns normalized recurrence values for all possible pairs of languages in the sample. After removing any values that are not statistically significant to the chosen $p < 0.005$ level, these values essentially comprise a similarity matrix. This can be converted into a distance matrix by calculating the inverse of each similarity measure and multiplying each of these values by 100. A standard average linkage method for clustering is a good point of comparison. Not only is it a relatively simple binary branching method, but it also parallels Kessler’s implementation of multilateral comparison in averaging the relationship values between a candidate language and all component languages of any node it is compared with (Johnson 2008). One difference between the methods is that Kessler’s procedure aborts under certain conditions without forcing the creation of a single tree that contains all languages, while the average linking method used to produce figure 3.3 requires all input languages to be included in a single output tree.

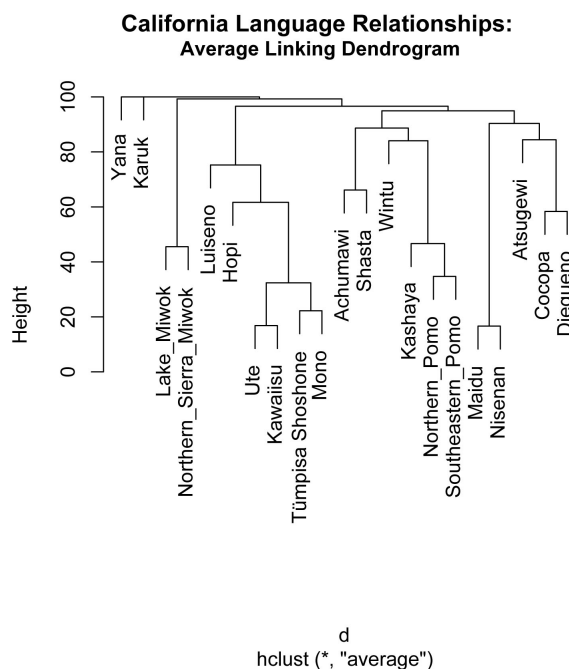


Figure 3.3: Average linking clustering of California languages

Aside from the eventual linking of all smaller families into a single group, the lower level clusters indicated by this dendrogram are quite similar to the clusters produced by Kessler’s method. The clustering of Wintu with the Pomoan languages and the creation of a node containing Atsugewi, Cocopa, and Diegueño, however, are problematic. And although

this procedure forces higher level groupings to be constructed, the Hokan and Penutian stocks do not emerge as branches. The obviously spurious low-level relationships like the Atsugewi-Yuman cluster together with the absence of plausible branchings in the higher level relationships that are forced demonstrate that Kessler’s clustering procedure is indeed better suited to the task of using the recurrence measures to identify layered relationships. Furthermore, the absence of Hokan or Penutian-like groupings in this maximally deep tree supports the original method’s finding of no Hokan or Penutian stock-level relationships.

3.6 Conclusions

The results of this test confirm that the program is able to identify phylogenetic relationships. Furthermore, because the method uses the data from all languages of a cluster to retest the statistic during each round of the clustering procedure, it is able to identify increasingly subtle patterns as larger groupings are built (see section 3.5.1, page 79), making it more sensitive to deep relationships than algorithms that are based solely on pairwise comparisons. The other striking result of the test is that all of the generated dendrograms represent relationships for which there is already evidence from traditional historical methods. None of the identified relationships are obviously spurious, but neither do the results include any evidence for deeper relationships like Hokan and Penutian. In fact, even if statistical significance were disregarded, the measures of relationship strength reported for the final round of clustering show no particular Hokan or Penutian patterns (see Round 11 in Appendix I). The unlikely relationships between Miwok and Shasta, Miwok and Atsugewi, and Pomoan and Uto-Aztecan have some of the highest recurrence metrics of the non-significant relationships. Meanwhile the “Hokan” Pomoan-Yana and Pomoan-Yuman relationships and the “Penutian” Maiduan-Wintu and Maiduan-Miwok relationships have some of the lowest recurrence values in the table. Both in its measure of relationship strength and in its evaluation of statistical significance, this test of linguistic relationships among California languages offers no support for Hokan or Penutian.

The heavy reliance of this work on documentation and analysis done by a number of researchers makes it vulnerable to a several sources of error. The extraction of data from dictionaries and field notes brings with it the assumption that the authors and collectors of these materials were experts, whose morphological and phonological analyses are correct. Similarly, the faithfulness of transcriptions and the adherence to phonemic representations by authors who describe their transcriptions as phonemic is also assumed. A more insidious potential source of error is the task of selecting the most basic root for each meaning in each language. When a meaning was represented by only one word in a particular source, that word was assumed to be the most basic term. When multiple forms were listed for a single meaning, the most basic form was chosen by looking up each candidate form and reading any available dictionary entries and notes. This choice involves a certain amount of subjectivity, but the potential introduction of random, non-systematic error by this process was counterbalanced by the advantage of not shortening the word list by discarding these items.

The length and composition of the word list used in this study is a substantial limitation.

In order to maintain a length of at least 100 items, the minimum number required to produce reliable results, some items were kept in the list that might otherwise have been discarded. Meanings for which multiple dictionary entries exist are one category of data that could introduce error. Meanings whose exact semantic sense is difficult to control are another type of item that would ideally be removed. Any items for which there were obvious, unavoidable cross-linguistic inconsistencies in meaning were removed from the list, but it is possible that some semantic inconsistencies escaped notice.

These little sources of human error may introduce noise into the data but are unlikely to cause broad systematic problems with the procedure. More disconcerting is the observance of a false negative result in the Palaihnihan family. Recoding the data to preserve vowel quality information is a potential solution for the Atsugewi/Achumawi dilemma, but the occurrence of this error indicates that despite efforts to reduce the interference of phonological differences in the identification of relationships, the significance testing procedure is affected by morphophonological patterns that skew the distribution of initial sounds in a meaning list. The obvious way to diminish the possibility that a single morphological or phonological pattern could create a false positive is to compare a larger portion of each word. An ideal solution would eliminate false negatives in a cross-linguistically applicable way, without adding the cumbersome process of matching multiple segments to the already computationally intensive Monte Carlo and round by round retesting that this methodology involves.

One measure that might help to prevent false negatives is the use of a longer and more carefully controlled word list. The process of whittling Swadesh's 200 item list to 102 meanings in this study produced a relatively arbitrary set of data. Controlling the balance between items that are generally nominal and those that are more verb-like could dampen the impact of morphological artifacts somewhat. Increasing the number of items in the list would also make the results of the study more robust all around. The subtle differences in the strengths of the Northern Pomo-Kashaya relationship and the Northern Pomo-Southeastern Pomo relationship would be more accurately represented in a study based on a longer word list.

Fortunately, the false negative result obtained for Atsugewi and Achumawi does not cast any particular doubt on the negative results obtained at the stock level. As clusters are built up, the data that represents them becomes more complex, suppressing idiosyncrasies in the data of component languages. The fact that none of the families that were identified were clustered together into deeper groupings than those that are already well established cannot be discounted. The Penutian sample, which should be completely independent of Shasta and Palaihnihan, produced no deeper groupings than did the Hokan languages.

The absence of any Hokan or Penutian stock-level relationships in the results of this investigation is suggestive, but not conclusive. The 21 languages in the study provide an adequate sample for testing the applicability of this method to the question of deep relationships in California, but they may not provide a sufficient basis for an argument against the Hokan and Penutian hypotheses. This method targets meaningful relationships and resists grouping individual languages with families that they are not members of. This is one of the method's strengths, but it also limits its ability to construct deep groupings when only a limited number of component languages and low-level clusters are available. The sensitivity of larger clusters to subtle trends in sound correspondences among the lexical data of their

members suggests that with a larger input sample from which bigger family groups could be built, deeper relationships would be more likely to emerge.

One specific deficiency in the language sample used for this test is that it does not contain many members of the Yok-Utian group, a sub-category of Penutian for which considerable support and acceptance has been found (Whistler and Golla 1986; Callaghan 1997, 2001). A more complete language sample might make the construction of Utian and its parent node Yok-Utian possible. Quantitative support for this middle layer of the Penutian family tree would itself constitute an interesting result, but it is also possible that the size of this cluster would make it possible to identify even deeper relationships. The current sample includes only Northern Sierra Miwok and Lake Miwok as representatives of Yok-Utian. Adding Yokuts and Ohlone data will be a priority for subsequent work on this topic.

This investigation has demonstrated that Kessler's method for testing relationships between languages is a useful tool for examining the deep connections among California languages, and more importantly it has outlined a number of modifications to the general procedure that can improve the results that are obtained. Importantly, the expression of recurrence values as ratios allows the procedure to operate across languages with phonemic inventories as varied as those of the California languages. The results of this study also point toward expanding the categories of vowels represented in the encoding system, using more than one segment in the representation of word list items, including as many languages as possible in the test, and using a long, well thought out word list as ways of achieving better results.

As a test of deep linguistic relationships in California, this study provides some surprisingly tantalizing negative results – while a larger sample of languages might produce more definitive results, the findings of this chapter certainly deepen the shadow of doubt that surrounds Hokan and Penutian. The absence of any clustering above the level of well-established small families in the Hokan and Penutian samples is a challenge to the genealogical status of these groups. This study also opens up obvious avenues for continued work on the topic, including methodological refinements necessary to handle the sort of morphophonological complications typified by the Shasta-Palaihnihan relationship as well as a more complete test of the Yok-Utian subgroup of Penutian, which is supported by sound evidence and should be within the temporal range of this methodology.

Chapter 4

The Geography of the Northern California Linguistic Area

4.1 Introduction

Northern California is commonly considered to be a linguistic area, having been classified as such by Haas (1976) and associated with several areal traits in Sherzer's North American survey (1976). Since this era, Northern California has made many appearances on lists of linguistic areas and has been treated uncontroversially as a region of areal diffusion. But the proposals put forth by both Haas and Sherzer provide limited evidence for this linguistic area and recommend at least in general terms that further areal linguistic research be conducted in this region. Since that era, more careful work has been done on the Northwestern California subregion (Conathan 2004), the Clear Lake subregion (Mithun 1999), and California as a whole (Golla 2011), but the Northern California linguistic area has been accepted without much further comment on whether areal patterns really correspond to that scale and location (e.g. in Campbell 1997; Mithun 1999).

As many have noted, California's deep history of settlement, great diversity, and small settlement sizes make it a difficult region in which to distinguish traces of historically deep genealogical relationships from contact-related borrowings (Haas 1976; Golla 2011). There is no doubt that linguistic traits have spread across genealogical boundaries in the Northern California region. Moreover, it is less clear whether it is most appropriate to classify this exact region as a linguistic area, or to instead think of it in terms of smaller regions of more intense areal diffusion or an extension of some of western North America's larger linguistic areas.

This study uses mapping and geographic analysis to re-evaluate the linguistic traits that have been used to support the idea of a Northern California linguistic area. This method identifies features more likely to have diffused geographically (versus genealogical or chance occurrence) and examines their distributions to determine whether the spatial patterns associated with proposed areal features support the notion of a Northern California linguistic area. While many of the features associated with Northern California do show geographic organization in their distributions that suggest they may have spread areally, the cumulative patterns that emerge from these individual feature distributions point to greater intensity of

feature diffusion on the western edge of the Northern California area, where the Northwestern California and Clear Lake micro-areas are located, and fewer feature spreads on the eastern side of the area. As a result, Northern California appears less like a cohesive region of areal feature diffusion and more like a collection of smaller diffusion zones that is overlapped by the fuzzy boundary of the Great Basin area to the east and may be linked to the Northwest Coast linguistic area to the north.

4.1.1 Northern California in Areal Linguistics

The idea that areal feature spread might be a significant contributor to language change and linguistic resemblances among Northern California languages is not unreasonable given the sociolinguistic character of the region. As Golla notes in his discussion of California as a sociolinguistic area, the sociopolitical organization of people into “tribelets” throughout much of California led to a situation in which languages and linguistic boundaries were not functionally tied to political entities, since the primary political units in this region were very small communities that spoke the same languages as other nearby but politically independent groups (2011:1-2). This organization of people into small groups and independence of language from political identities resulted in languages being associated instead with places. Though this may have tended to preserve languages and linguistic boundaries it also led to a situation where multilingualism was commonplace. Known patterns of trade, intermarriage, and cultural similarity contribute further to a sociolinguistic situation across much of Northern California where the conditions would seem to support the spread of linguistic traits across unrelated languages. So despite the great genealogical diversity of Northern California’s languages the region has the depth of settlement, the mobility of people, evidence of language shift, and sociolinguistic unity that would fit well with at least some theories of how linguistic areas arise. However, an important factor in the wide geographic spread of features across a linguistic area is the geographic scale of contact. Conathan (2004:4) notes that in Northwestern California linguistic contact is a very local phenomenon, and this is certainly true for many other parts of the greater Northern California area. The restricted spatial scale of contact in much of Northern California is atypical of classic linguistic areas. A further complication in considering the areal status of Northern California is the depth of settlement and size of language families. This region is home to many isolates and very small established language families, and at the same time very deep genealogical relationships have been proposed among the region’s languages (e.g. the Hokan and Penutian stocks). Distinguishing areal feature spread from very distant genealogical signals is a complex task in a region like Northern California, and the limited data for some languages only compounds the problem.

Questions of areal resemblances feature in discussions of California linguistics as early as the beginning of the 20th century, when classificatory work on this region’s languages began. Boas, whose interest in the diffusion of linguistic traits is well known, took note early on of the Pacific Coast of North America as a region where diffusion of grammatical features has led to similarities in neighboring, but genealogically unrelated, languages (Boas 1929). Dixon and Kroeber examined the relationships specifically among California languages with a pointed interest in similarities that were not explained by established genealogical

relationships (Dixon and Kroeber 1903). This work eventually gave rise to the Hokan and Penutian stock proposals (Dixon and Kroeber 1919), but in developing these theories of genealogical relatedness Dixon and Kroeber also paid close attention to the possible influence of areal feature spreads on California languages. Several papers investigating the occurrence of individual grammatical features in California languages were published in this era as a result of this interest in understanding both the genealogical and the areal factors that have influenced the development of California's indigenous languages (e.g. Dixon 1906; Dixon and Kroeber 1907; Kroeber 1909, 1907a).

The first explicit claim that Northern California is a linguistic area was put forth by Mary Haas in 1976. Her article in the proceedings of the First Conference on Hokan Languages formalized the idea that despite its great linguistic diversity, Northern California is defined by several notable linguistic features that have spread across genealogical boundaries in this region. Haas's work on the Northern California linguistic area made good use of limited descriptions of California languages, examining both phonological and morphological features of the region. However, the paper centers on the impressionistic identification of similarities among languages within a specific geographic region and offers few details regarding the histories of most individual features or the probabilities of the observed patterns arising for non-areal reasons. Haas describes the phonological inventories of a sample of Northern California languages, which includes members of each of the small, well-established language families of the region, but does not include every member of each family (e.g. Yukian is represented by Wappo but not Northern Yukian). Based on her sample of 16 languages Haas notes that several phonological features commonly associated with the Northwest Coast are present in many California languages, including "back velar" consonants (q , x) and voiceless fricated or glottalized lateral consonants ($ʈ$ and $ʂ'$). Other features, such as back/retroflex apical stops ($ʈ$) and a distinction between l and r , are described as occurring in a more restricted Northern California region. In her examination of morphological features, Haas considers a slightly larger sample of Northern California languages that includes multiple languages from Algic, Athabaskan, and some of the smaller families such as Palaihnihan, Wintun, Miwok, and Yukian. She notes that diminutive consonant symbolism is common in Northern California as well as in several Oregon languages and n - and m -initial first and second person pronouns also seem to define the area, though they also extend outside of the region of interest. Haas's discussion of numeral systems, however, is restricted to the Northern California area without much commentary on how the region's patterns compare with its surroundings.

Also in 1976, California was one of the regions used by Sherzer to frame his discussion of North American areal linguistics. Sherzer's North American study surveyed phonological and morphological features in languages grouped according to Driver's (1961) culture areas in order to investigate the distribution of areal features in this continent. While Sherzer's work was far broader than Haas's and outlined areal trends in the entire North American continent on a region-by-region basis, the list of areal traits he tracked in each region included many of the same linguistic features Haas noted as Northern California traits. Sherzer's work was an important contribution to areal linguistics, outlining the distribution of linguistic traits across North America and beginning to tease apart features associated with genealogical groups from those for which geographic spread is a more likely mode of inheritance in some

areas. Sherzer's study also estimated the scale of areal feature spreads, noting sets of "regional areal traits", "central areal traits", and "whole areal traits" for each of the regions in his study, as well as features associated with specific genealogical groups.

Sherzer's interest in the entirety of North America rather than in a specific region imposed a measure of uniformity on his treatments of individual areas. For California, as well as each of the other regions of North America, he noted whether a standard set of phonological and morphological traits was attested in each of the languages (e.g. Yurok, Yana) and small families (e.g. Ohlone) for which data was available. The discussion of each region includes summaries of the apparent areal patterns at the three aforementioned scales as well as further discussion regarding the probable directions of feature spread across geographically neighboring languages.

For California, Sherzer notes a great overall level of diversity, with no feature present in all the languages of one family/stock while at the same time absent from any language outside that family/stock (1976:126). This general pattern corresponds to a rather striking level of diversity within most of the small language families of California, which is particularly notable when compared to less diverse regions like the neighboring Great Basin. Sherzer finds very few "whole areal traits" that are present throughout the entire California region – pronominal plurals, in fact, are the only feature he identifies as a California "whole areal trait". However, in the Northern California sub-region he identifies several "regional areal traits": a consonant system with three series of stops, a k/q contrast, a k/k^w contrast, an $s/\text{ʃ}$ contrast, an r/l contrast, voiced and voiceless velar fricatives (x and ɣ), a labiovelar fricative (x^w), a lateral fricative and glottalized lateral affricate (ɬ and ɬ^w), a numeral classifier system, and prefixation of nominal person markers, verbal subject markers, and tense/aspect markers. Sherzer notes that several of these traits are also attested in the Northwest Coast and Plateau regions.

Several specific directional spreads of traits within Northern California are also proposed by Sherzer based on his areal feature survey. He hypothesizes that glottalized stops spread into Yurok and Lake Miwok from neighboring languages, and that Wintu adopted a k/q contrast from neighboring languages. He further proposes that the development of a k/k^w contrast and a glottalized lateral (ɬ^w) in Yurok¹ and Wiyot was contact-induced, that the development of a glottalized lateral affricate in Lake Miwok resulted from contact with Wintun, and that the velar nasal (ŋ) spread from Uto-Aztecan into some neighboring Hokan and Penutian languages. With regard to morphological traits, he suggests that a nominal case system developed in some Hokan and Uto-Aztecan languages as a result of contact with Penutian, that Yuki's inclusive/exclusive plural distinction and Yana's evidential suffixes arose as a result of contact, and that the Maiduan languages developed instrumental verbal prefixes through the influence of neighboring languages. These proposals are made carefully and with a disclaimer that more analysis is necessary to better understand each of these apparent geographic feature spreads, as well as to better distinguish true areal traits from traces of very deep genealogical relationships.

More recently, Northern California received some mention in Campbell's overview of the historical linguistics of American languages (1997). Campbell defines the Northern Califor-

¹Though Sherzer analyzed Yurok's glottalized fricatives as unitary phonemes, they are probably better analyzed as sequences of sounds

nia linguistic area by the languages that fall within it, and the region he spells out has a northern border that is nearly coincident with the northern border of the state, except where Shasta and Klamath-Modoc extend into southern Oregon². The eastern edge of Campbell's Northern California area is also near the state's eastern border, separating Northern California's Klamath-Modoc, Palaihnihan, Maiduan, and Miwok languages from the Great Basin's Washo and Numic languages. The southern border of the region is mostly defined by the southern boundaries of the Miwok languages, extending as far south as Southern Sierra Miwok in the east, but limited to the area north of the Golden Gate of San Francisco Bay in the west.

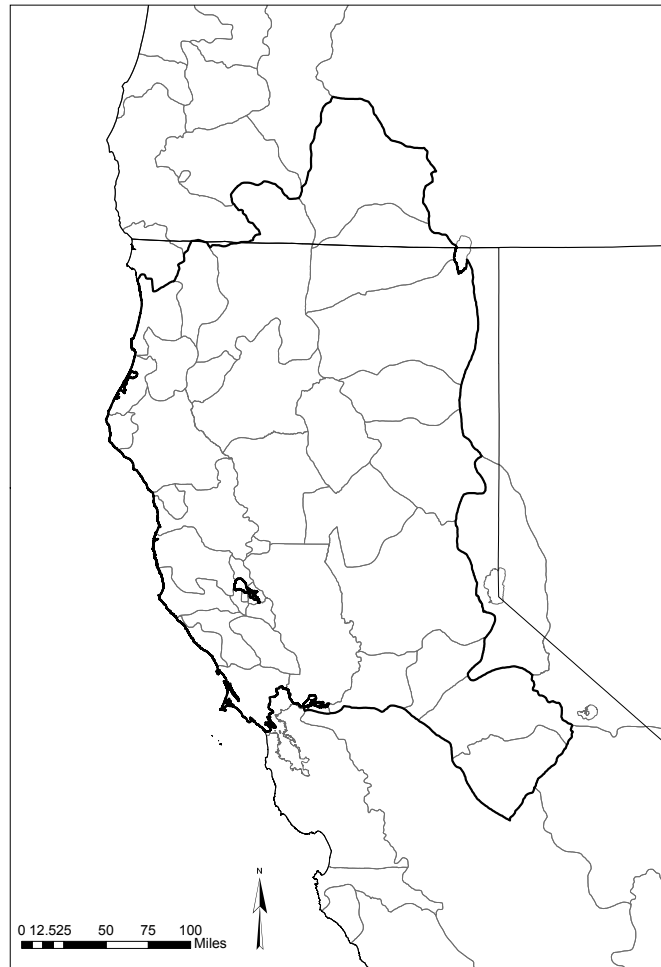


Figure 4.1: Campbell's Northern California area

²The Tolowa-Chetco language is also spoken on both sides of the California/Oregon border. Campbell does not explicitly place Tolowa in either the Northern California area or the Northwest Coast area, and so its extension into Oregon is not considered part of Campbell's Northern California area (1997). However, in spite of being classified as an Oregon Athabaskan language (Golla 2011), Tolowa was in close contact with both Yurok and Karuk and was included in Conathan's Northwestern California linguistic area (2004). Though it is apparently not part of Campbell's conception of the Northern California area, there is enough evidence to include Tolowa in this study's Northern California area.

Campbell's discussion of Northern California as a linguistic area primarily summarizes the findings of Dixon and Kroeber (1903) and Haas (1976), but he provides an extraordinarily concrete description of the extent of the region and interesting commentary about the features found in Northern California and their attestation in neighboring regions. In addition to noting that many of the features that are traditionally associated with Northern California are also found in the Northwest Coast and Plateau regions, Campbell comments that the retroflex/back apical stops noted by Haas extend farther south into Yokuts and that Washo resembles the Northern California languages in many ways, such as in its dual pronouns and its quinary numeral system. While these facts do not lead to further questions about the Northern California linguistic area, Campbell does question the status of the Great Basin as a linguistic area, since inclusion of Washo in the Northern California area would leave only languages from the Numic branch of Uto-Aztecan in the Great Basin area. He goes so far as to entertain the idea that the Great Basin may be an extension of the Northern California area, rather than an independent linguistic area (Campbell 1997:339).

The Northern California linguistic area was also noted in Mithun's survey of the languages of North America (Mithun 1999). Mithun's description of the features shared by the Northwest Coast and Northern California follows Haas and Campbell in highlighting the plain, aspirated, and glottalized stop series and the lateral and post-velar consonants as features found in Northern California that also occur in languages of the Northwest Coast area. She notes that while retroflex stops are frequently assumed to be a Northern California areal feature, the symbol most often used to represent these sounds (ʈ vs. a plain t) is frequently also used to represent a simple distinction between "back" and "front" (i.e. alveolar and dental) stops, particularly in Northern California. Thus it is probably a contrast between dental and alveolar stops, rather than retroflexion, that occurs across much of Northern California. Mithun goes into greater detail discussing the Clear Lake micro-area, a region about 100 miles north of San Francisco around the lake of the same name. In this region Miwok, Wintun, and Pomoan languages, as well as the Yukian language Wappo, come into contact. In addition to a large number of lexical loans between the languages of this sub-region, Mithun also notes the remarkable phonological differences between Lake Miwok and the rest of its linguistic relatives, having acquired the aspirated, ejective, and voiced stop series through contact with other Clear Lake languages, as well as several affricates, a rhotic consonant, and the lateral fricative and glottalized lateral affricate.

Areal feature spread in Northern California was the specific focus of Lisa Conathan's 2004 dissertation. Her study focused on the northwest corner of California, including Tolowa, Hupa, Chimariko, Wiyot, Yurok, and Karuk. Conathan finds a very different type of linguistic area in Northwestern California than Mithun identified in Clear Lake. While lexical borrowings and phonological adaptations define the Clear Lake contact area, Conathan finds Northwestern California to be typified by "functional convergence", or adaptation of semantic categories and pragmatic functions in response to contact without actually borrowing or calquing the formal expression of these features (2004:180). For example, the repetitive reduplication in Yurok was adopted through contact with Karuk and is used to express the same meanings (Conathan and Wood 2003). This proposed mechanism of Northwestern California linguistic change is very similar to Ross's notion of metatypy, or the restructuring of morphosyntactic characteristics of a community's primary language in response to contact

and bilingualism with an intergroup language (1996:182).

The most recent and most comprehensive study of areal traits of California is Golla's chapter on areal features of the region in his 2011 research guide to California Indian languages. Golla notes that several localized features are found within California but states somewhat vaguely that the character of the area's languages is also defined by more widespread features that co-occur in particular combinations within California (Golla 2011:203). Golla's list of California areal features includes phonological and morphological traits as well as elements he classifies as "linguistic culture", including numeral systems, names, and other culturally important linguistic domains. Golla describes California phonology in terms of many of the same features as Haas, Sherzer, Campbell, and Mithun have, but he states the distributions of these features in great detail. As had been noted in earlier studies, laryngeal feature contrasts in consonant series are common in Northern California, particularly a three-way contrast between plain, aspirated, and glottalized consonants. Golla also notes a sub-area where voiced consonants occur in contrast with voiceless and either glottalized or both glottalized and aspirated consonants. Contrasts in consonant articulation are also noted, including apical vs. palatal affricates (*c* and *č*), front and back apical stops (*t* and *ʈ*), velars versus back velars (*k* and *q*, *x* and *χ*). Golla notes that the front/back apical stop contrast is found primarily in the Coast Range and Central Valley (2011:205), which places this feature somewhat south of the area that Campbell defined as the Northern California Linguistic Area (Campbell 1997). Backed velars, he notes, are found in the northwestern corner of California and the northeastern corner, but are distributed in these two Northern California clusters rather than a single large area, despite the probability that at some great historical depth these two clusters are likely linked.

Golla also discusses many individual sounds found in California languages as well as their distributions. Labiovelars, lateral fricatives and glottalized affricates, and a contrast between front and back sibilants are found sporadically throughout California. Also distributed in the region but without a clear source or pattern of spread are velar nasals, rhotic approximants, and a rhotic/lateral contrast. The most common vowel system in Northern California is a five-vowel system, though Golla details all of the attested systems including those with rhotic or nasalized vowels. Six- and seven-vowel systems are found adjacent to the Great Basin in Miwok and Maiduan as well as Washo and some varieties of Yokuts. Systems with four vowels are found in the northernmost part of California, in Athabaskan languages, Shasta, and Klamath-Modoc. Rhotic vowels are found in Northern California only in Yurok, though they do appear in the Serrano (Takic, Uto-Aztecan) language of Southern California as well. Nasalized vowels are found in some Oregon Athabaskan languages, like Tolowa, but are otherwise rare in California.

There is some overlap between the morphological areal traits listed by Golla and those on Haas' list of Northern California features, but Golla goes even further in describing the morphological character of California languages and the possible areal traits of the region. He notes that reduplication is common, particularly for marking habitual, repetitive, or distributive categories, but its distribution in California is uneven. Other morphological traits he associates with genealogical groups. Ablaut, dependent marking, and a nominative-accusative case marking system are commonly found in Penutian. Suppletion and head marking are common in Hokan. Some languages and small families are noted for their departure from

these patterns. Golla claims that Eastern Miwok and Tübatulabal, both of which express grammatical relations through double-marking, are in transition from dependent marking to head marking and that Pomoan has completed the transition from head marking to dependent marking. Yukian and Pomoan also have case marking, although their systems are quite different from the typical Penutian system.

Features that Golla associates with the entire California region include infrequent marking of plurality on subject nominals, sound symbolic diminutives, etymologically quinary counting systems³, and systems of stem derivation such as instrumental affixes and bipartite stems. Bipartite stems, which have been discussed in depth by DeLancey (1996; 1999), are most common in the southern Cascades and northern Sierra Nevada – DeLancey’s (1996) “bipartite stem belt”, and Golla proposes that the Penutian languages that use this stem derivation strategy have acquired it through contact or substrate effects from Hokan languages. Instrumental affixes describing the shape or characteristics of an instrument or absolutive argument of the verb are less common in Northern California, according to Golla, although they are found in the Numic languages to the east. Other features that Golla finds restricted to an area closer to Northern California are the switch reference system found in Pomoan, Washo, Maiduan, and Yuki as well as Yuman farther south, and vigesimal counting systems, which proceed by twenties and are found in the Sacramento Valley and surrounding areas. Golla associates vigesimal systems with the use of clamshell beads in trading, which spread from the Cost Miwok and Pomoan areas north of the San Francisco Bay up the Sacramento Valley and into far northern parts of California.

From these discussions of areal linguistics in Northern California a core set of features commonly associated with Northern California can be compiled. These features are:

(4.1) Phonological

- plain, glottalized, and aspirated stop series
- five-vowel systems
- “back” apical stops
- “back velar” consonants
- fricated and affricated laterals
- *l/r* contrast
- velar nasal (only in a sub-area of Northern California)

Morphological

- sound symbolic diminutivization
- quinary numeral systems
- reduplication to express distributive or repetitive meanings
- case marking
- prefixed verb subject markers
- pre-verbal tense and aspect markers
- prefixed possessive markers
- *n-* and *m-* initial first and second person pronouns
- numeral classifier systems (only in a sub-area of Northern California)

³Golla states that only Takic languages like Cupeño have actual quinary systems, where the counting proceeds by fives throughout the system. The remainder of California’s “quinary” systems refer to the etymology of the numerals 1-10 in an otherwise decimal system.

These traits feature prominently in the discourse on areal diffusion of linguistic material in Northern California. Due to the relatively concrete and basic nature of these features, the data needed to analyze their distributions in Northern California and the surrounding regions is available in grammatical descriptions and other published sources. Thus, these traits serve as the focus for a new, geographically-oriented study of Northern California as a linguistic area.

4.1.2 Language Areas in Linguistic Theory

Defining the concept of a language area is a difficult and contentious issue within linguistics and there is a significant body of literature that represents the ongoing debate. The concept of a linguistic area originated as Trubetskoy's *Sprachbund* in the early twentieth century (1928). From that time, the definition of the term 'linguistic area' has remained fuzzy, since no general agreement has been reached regarding what constitutes a linguistic area and how these areas form (Thomason 2001). In several of the well-accepted linguistic areas – e.g. the Balkans (Sandfeld 1930), India (Emeneau 1956) – there is good evidence that grammatical traits have spread across genealogical boundaries within some geographically well-defined area and that the patterns in spatial feature distribution are unlikely to have arisen due to chance or genealogical inheritance. No standard criteria exist for classifying a region as a linguistic area (for discussion, see Campbell 2006), although recent work has proposed that statistical measures can evaluate areality and identify areas of feature diffusion (Donohue and Whiting 2011). For assessing existing proposals that specific regions function as linguistic areas, however, the aforementioned well-established linguistic areas share characteristics that make a good starting point. Specifically, these areas share the attestation of some linguistic traits in unrelated or very distantly related languages (i.e. features for which mutual inheritance from some proto-language is an unlikely or impossible source), and they are geographically and culturally cohesive regions. These two characteristics seem like important components of areal diffusion and will serve as a relatively theory-neutral set of criteria for evaluating the Northern California linguistic area.

Still, these two points for assessing a linguistic area would not satisfy everyone. Campbell finds little reason to worry about boundaries and geographical patterns as criteria for linguistic areas, arguing that understanding the geography of linguistic areas is at best less important than understanding the history of diffusion events in the area and in more egregious misapplications tends to overshadow the linguistic traits themselves (2006:14-16). While the assertion that geographic determinism should not shape our approach to areal linguistics is a very reasonable one, the historically-detailed approach that Campbell promotes instead is problematic in situations where the histories of some traits, and even the linguistic relationships that would make it possible to disentangle genealogical and areal signals, are unknown and possibly unknowable. Here the use of geographic patterns to gain some insight into linguistic history – even if it yields only a crude or incomplete picture – is a way to augment approaches that focus on more detailed historical information.

Campbell's criticism of the geographic component of areal linguistics is in fact based on some of the same concerns that have directed this study toward a more geographically oriented approach. His comment that "it is necessary to combat the notion that the geography

is prime and the borrowings are in some way secondary to and determined by the geography” perhaps discounts the influence of geographic phenomena on human behavior and culture more than is necessary, but seems to identify a very real problem that occurs when research is undertaken with the expectation that linguistic areas will conform to the boundaries of culturally or environmentally defined areas (Campbell 2006:16). If assumptions are made about the scale and extent of linguistic areas instead of letting these geographic parameters fall out from the distributions of areal features, the characterization of regions as linguistic areas could be misleading. For example, is Northern California considered a linguistic area because it happens to be the geographic area across which some set of features diffused? Or was its classification as a linguistic area influenced by the conception of Northern California as a cultural and environmental province as well as an expectation that the areal linguistic diffusion pattern would align with this familiar region? Reassessing the status of Northern California as a linguistic area may seem like a trivial pursuit, but if Northern California is better characterized as a collection of small linguistic areas of the sort discussed by Conathan (2004) or as an extension of the Northwest Coast area, then it would seem that geography is an important consideration. Using areal trait distributions to better understand the scale and extent of diffusion in Northern California may or may not lead to new support for the traditional view that it is in and of itself a linguistic area, but getting a better sense for the patterns and scale of Northern California areal diffusion will certainly lead to a richer understanding of the region’s linguistic history.

4.2 Methods

The methods used to analyze these trait distributions build on the ideas that a linguistic area should show evidence that features co-occur in unrelated languages due to diffusion rather than common inheritance or chance, and that the area should be geographically cohesive. These criteria are assessed in two steps. In the first step a feature index is calculated for each language. This index expresses numerically the presence or absence of the feature and whether its occurrence is expected based on the language’s family affiliation. Spatial autocorrelation analysis is performed on the mapped feature index values. This step identifies features that are likely to satisfy the first criterion: evidence of diffusion. The second step is an analysis of the overall geographic patterns that occur in the region to assess whether the region of focus is geographically cohesive. This includes a qualitative assessment of individual and cumulative feature distribution patterns, the use of spatial autocorrelation to look for a core-periphery type pattern, and a simple *t*-test to determine whether the cumulative presence of areal features in the region of focus is significantly different from its surroundings.

The dataset used for this study is a survey of proposed areal features of Northern California (see the list in 4.1) in 64 languages that include the Northern California area as well as languages that are found within 200 km of this area. As the first step in the development of the feature index, these features are coded as binary (0/1) data based on their presence or absence in a particular language. This coded data will be referred to as the feature occurrence data (*f*). Data were collected primarily from published resources⁴, and features that received

⁴Data for this study comes from: de Angulo and Freeland (1930, 1931); Aoki (1970); Banks (2007);

no mention in the literature are coded as absent. This zero coding of unmentioned data may lead to the underreporting of some features, but because traits that a language does not have are often left unmentioned in grammatical descriptions this practice was unavoidable and deemed to have only a minimal impact on the results.

Other coding issues arise when linguistic entities are found within languages as the result of synchronically transparent processes. For example, Karuk has only three basic vowels, but appears to have five vowel qualities represented in its long vowels. This happens to be the result of the coalescence of *ai* and *au* into *ee* and *oo* respectively, and does not reflect the underlying vowel system. As much as possible, the phonological features reported here reflect the language's phoneme inventory rather than simply the phonetic presence of a sound. In most cases, however, coding relied on the analysis of the linguist responsible for a language's grammatical description rather than an independent confirmation of the phonemic status of sounds. Similarly, the coding of morphological data relied on the analyses of published grammars. Case marking was coded as present for languages that mark syntactic cases on nominal phrases. Prefixed grammatical markers were diagnosed by the use of morphological affixes, whereas the category of pre-verbal tense and aspect markers was construed to include clitic-like items. The categories of personal pronouns with initial nasal consonants only took into consideration the basic, independent personal pronouns of a language; possessive pronouns and person markers were not investigated. Numeral classifier systems were coded as present for languages that express categories of object characteristics (e.g. shape, consistency) on numerals. Several languages in the study area express minimal distinctions such as human/non-human on numerals, but basic systems of gender and animacy were not considered as examples of the classifier systems noted by Haas (1976), Sherzer (1976), and others. Quinary numeral systems were coded as present in languages where existing analysis classified a numeral system as quinary or where the numeral system showed evidence of deriving numerals between five and ten through mathematical operations similar to those outlined by Haas (1976:355). Finally, the coding of reduplication and sound symbolism relied entirely on existing analyses.

The second stage in the development of the feature index involves coding each language for the familial association of the feature. Known genealogical affiliations of each language are used to develop this code. The study sample includes data from ten small California families, four Oregon language families, one Great Basin language family, and 14 language isolates or languages whose known relatives were documented only minimally or not at all before extinction. For the features analyzed in this study, every language was also assigned a binary (0/1) code representing whether that feature is associated with that language's family.

Barrett (1908b); Boas (1911); Boas and Goddard (1924); Bright (1957); Broadbent (1964); Buckley (1994); Callaghan (1963, 1970, 1984, 1987); Dayley (1989); Dixon (1910); Dixon and Kroeber (1907); Eatough (1999); Edel (1939); Frachtenberg (1914, 1922a,b); Freeland (1951); Gatschet (1890); Goddard (1912, 1929); Golla (1970, 1976, 2011); Heizer (1952, 1955); Hoijer (1966); Hymes (1966); Jansen (2010); Jany (2009); Kroeber (1907b,c, 1910, 1929, 1963); Lamb (1957); Li (1930); Mason (1916, 1918); Mithun (1999); Moshinsky (1974); Newman (1944); Okrand (1977); Olmsted (1984); Oswalt (1961, n.d., 1967); Pharris (2006); Pierce (1966); Pitkin (1984); Radin (1929); Reichard (1925); Rigsby (1966, 1969); Robins (1958); Sapir (1907, 1922b,a); Sapir and Swadesh (1960); Schlichter (1985); Seaburg (1977); Shaul (1995); Shimkin (1949); Shipley (1965); Silver (1966); Snapp et al. (1982); Teeter (1964); Thompson et al. (2006); Thornes (2003); Ultan (1967); Voegelin (1935)

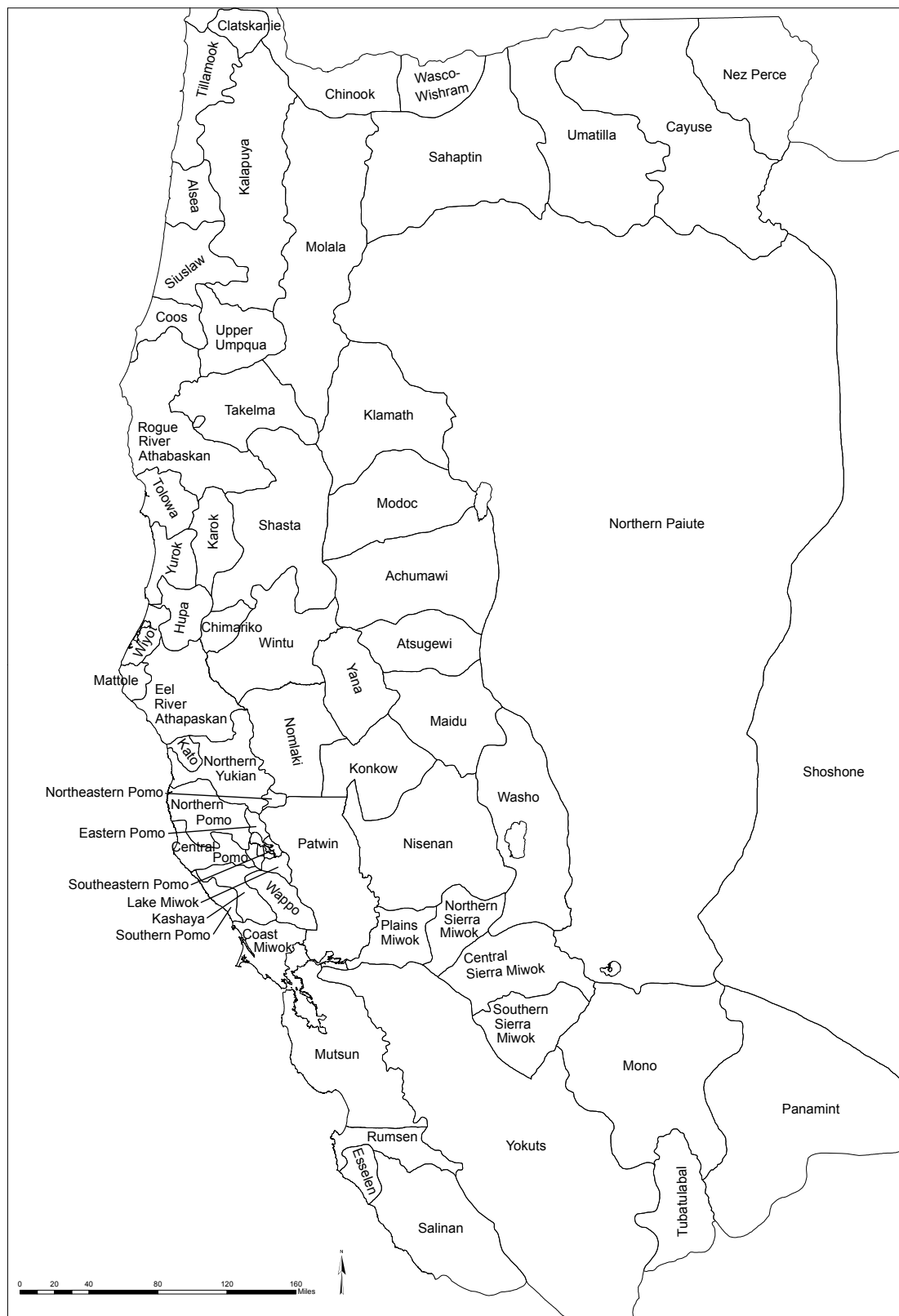


Figure 4.2: Northern California areal linguistic study area

This index (which I will refer to as g) is set to 1 for all languages that belong to families where all daughter languages have the feature; it is set to 0 for isolates and languages whose families have any daughter languages without the feature. This index (g) of genealogical group/feature association is a very simple way of summarizing the likelihood that a feature was inherited from an ancestor language. Because the amount of historical reconstruction that has been done varies from family to family and because the number of languages per family (between 2 and 7) is too small to conduct any statistical tests for association, the simple criterion of exceptionless occurrence in a family's languages was selected to minimize bias in identifying probable genealogical sources of features.

From these binary feature occurrence data (f) and genealogical group/feature association values (g), another metric is derived that weights the occurrence of features based on whether or not they are associated with a language's genealogical group. This index (which I will refer to as the feature index h) is calculated using simple arithmetic:

$$(4.2) \quad h = 2f - g$$

This procedure diminishes the weight of features in languages where all members of the family also contain the feature, and we might reasonably expect a high likelihood of that feature being inherited from the group's ancestor language. The values assigned for the feature index (h) will be highest if a feature occurs in a language but is not associated with the language's genealogical group ($h = 2$), lower if the feature occurs but is associated with the genealogical group ($h = 1$) and lowest if the feature does not occur ($h = 0$). This scheme results in values that represent the full geographic distributions of features, but also takes into account the likelihood of a genealogical source for a trait in any given language.

This feature index, which characterizes both the presence/absence of a feature and the likelihood that the feature was genealogically inherited, is better suited to the identification of areal feature spread than alternative coding methods. For example, analyzing the spatial distribution of traits using only the binary feature occurrence information (f) would not make it possible to infer whether the patterns detected represent a truly areal signal or whether they result primarily from genealogical feature inheritance. Analyzing only the occurrences of features where they are not associated with genealogical groups would divorce the suspected areas of spread from the areas that may have served as sources of those spreads, thus disrupting the geographic patterns we are trying to identify.

This feature index (h) is calculated for every feature in every language, and the resulting dataset is used to analyze geographic patterns in feature distributions. This analysis is done by calculating Moran's I , a measure of spatial autocorrelation, and the associated probability value, which measures the statistical significance of the identified patterns. Spatial autocorrelation characterizes the geographic organization of data. In highly spatially autocorrelated data, similar values will be found geographically near to one another, and more different values will be more distant. Highly negatively spatially autocorrelated data will show an evenly distributed pattern, with similar values spaced out evenly across the area of study. Data with very low spatial autocorrelation will be distributed randomly – neither clustered nor spaced evenly throughout the study area.

Spatial autocorrelation measures are frequently used to determine whether geographic parameters are missing from models relating independent and dependent variables, or whether

geographic patterns in variables exist in violation of the assumptions of certain statistical models. Frequently, ordinary least squares (OLS) regression is used to examine the relationship between variables, and the residuals of this correlation are analyzed for spatial correlation using a measure like Moran's I . If statistically significant spatial autocorrelation is found in the residuals, some geographic factor is influencing the relationship modeled by the regression, and the assumptions of the statistic are violated. In most cases the researcher will then need to identify the geographic influence and adapt the model. For this analysis of linguistic traits, the feature index (h) is used much like the residuals from OLS regression. Because the available data do not lend themselves to statistical procedures such as OLS regression, the subtraction procedure shown in (4.2) accounts for genealogical associations with traits, albeit in an abstract way, yielding something that represents unexplained variation much like an OLS residual does. If a positive, statistically significant Moran's I value is returned, the relevant linguistic variable shows evidence of greater geographic organization than we would expect by chance. In other words, where we find a positive, statistically significant Moran's I value, a feature's distribution is more likely to represent diffusion than chance occurrence. Having instituted some control for genealogical inheritance in the calculation of the feature index (h), we can conclude that features with positive, significant ($p < 0.05$) spatial autocorrelation values are likely to be areal features.

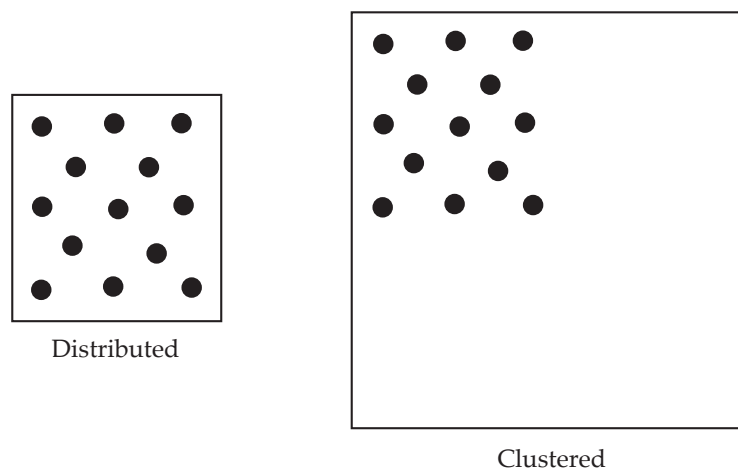


Figure 4.3: Effect of study area size on spatial autocorrelation

Measures of global spatial autocorrelation, like Moran's I , require the consideration of several other parameters. Because spatial autocorrelation metrics compare values between a point or shape and its spatial neighbors, the spatial neighborhood must be defined. Invoking Occam's Razor, a simple model of language contact might predict feature diffusion between contiguous languages, so for this study the spatial weights used to calculate Moran's I are assigned based on the "queen's case" pattern (i.e. based on the classification of any language that shares a border or touches a corner as a neighbor). This connectivity-based weighting represents a simple model of language contact, but weighting could also be based on geographic distance for studies where that is a better representation of the geographic linkages.

Another consideration in the analysis of spatial autocorrelation is the sensitivity of global spatial autocorrelation measures to the scale of the study area. In Figure 4.3, for example, the same pattern of dots appears to be evenly distributed when it occurs in a small area, but appears to be clustered when it occurs in a large area. To accurately identify patterns associated with Northern California, a study area was chosen that includes all of Northern California as well as all languages that occur within 200 km of the region's edges. This study area size allows us to identify Northern California-specific patterns as well as identifying patterns which are likely associated with a larger area. It further minimizes the risk of erroneously describing Northern California clusters as distributed patterns or the opposite, erroneously describing a pattern as clustered.

In addition to identifying likely areal traits, this study analyzes the spatial patterns in feature distributions in Northern California to assess this region's status as a linguistic area. One component of this is a qualitative assessment of the distribution of each linguistic feature. This assessment involves detecting how similar a feature's spatial distribution is to the extent of the Northern California area, determining whether features are restricted to Northern California or extend beyond the area, and examining how even or patchy the distribution is within and outside of Northern California. The quantitative component of the assessment of Northern California's linguistic area status involves counting the number of probable areal features that occur in each language and mapping these counts. Moran's I is again used to assess the spatial autocorrelation within these count data, since a core/periphery-type pattern would manifest as significant positive spatial autocorrelation. Finally, a simple two-tailed t -test is used to determine whether the counts of areal features in Northern California languages are significantly different than the counts of areal features found in the study area's non-Northern California languages. Finding any features whose distributions are similar to the extent of the Northern California area, any significant spatial autocorrelation in feature counts that corresponds to a Northern California cluster, or a significant difference in number of areal traits between Northern California and its surroundings demonstrated by a t -test would constitute evidence for Northern California being a linguistic area.

4.3 Results

Several of the phonological traits commonly considered to be areal traits of Northern California are distributed in spatial patterns that are significantly different than the patterns that would be expected due to chance. These features are likely to have arisen through the geographic diffusion of linguistic material.

Fewer of the proposed morphological traits of Northern California show significant geographic patterning based on the feature index (h), which takes into account the associations between traits and genealogical groups. Despite the small number of morphological features that show evidence of areal diffusion, some of the features most commonly associated with this area do show significant geographic organization in their distributions.

4.3.1 Phonological Traits

Contrast between plain, glottalized, and aspirated stops

The three-way contrast between plain, glottalized, and aspirated stops occurs in all of the languages of the Pomoan, Maiduan, Wintuan, Klamath-Modoc, California Athabaskan, Oregon Athabaskan, and Chinookan families included in the study sample.

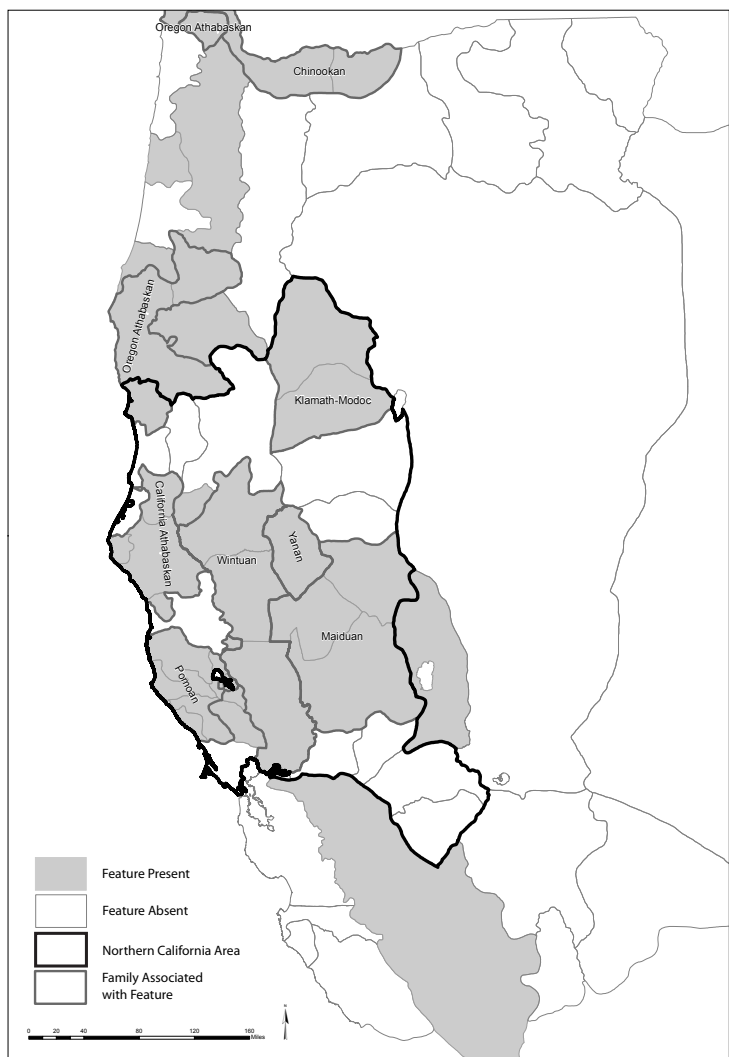


Figure 4.4: Distribution of plain, glottalized, and aspirated stop series contrast

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for this contrast is small and not statistically significant. This contrast is spatially distributed in a pattern that is similar to distributions generated by chance.

Table 4.1: Spatial autocorrelation of plain/glottalized/aspirated stop series contrast

Moran's I	p value
0.077475	0.244974

Glottalized stops are found in a wider distribution, occurring in all of the languages of the families mentioned above, as well as all of the Yukian, Shasta-Palaihnihan, Sahaptian, and Salishan languages included in the sample.

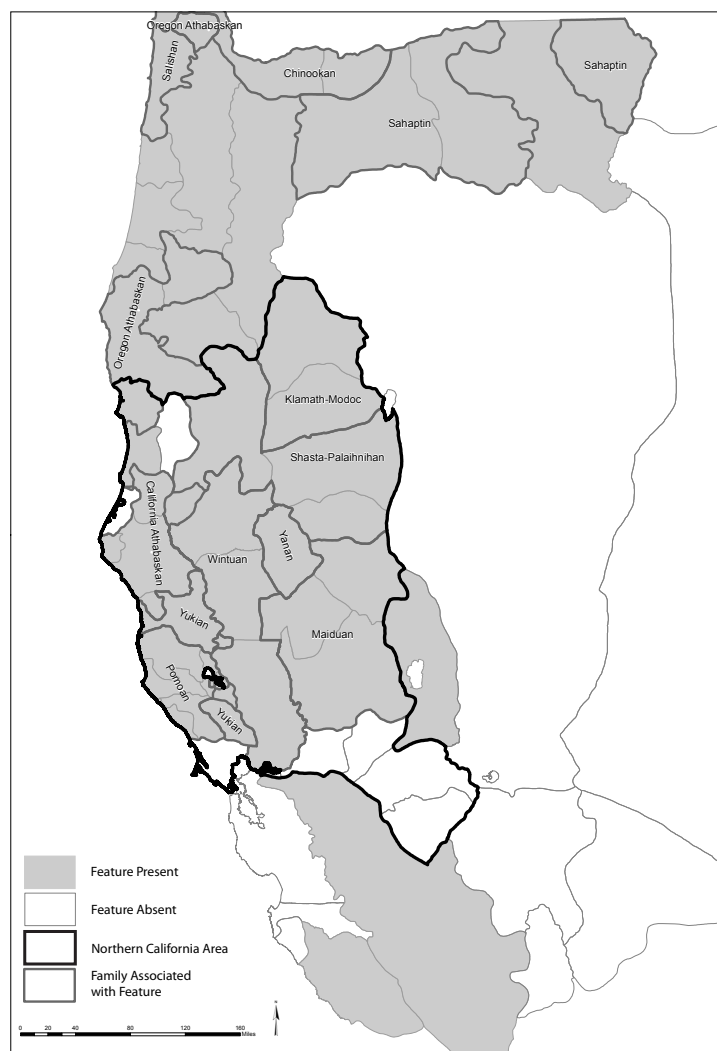


Figure 4.5: Distribution of glottalized stops

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for glottalized stops is not statistically significant. Like the 3-way contrast, glottalized stops are not distributed in a pattern that is significantly different than the patterns we would expect due to chance. However, if the scale of the study area were broadened there may be evidence of a larger West Coast cluster.

Table 4.2: Spatial autocorrelation of glottalized stops

Moran's I	p value
0.124552	0.077849

Five-vowel system

A five-vowel system occurs in all of the languages of the Ohlone, Yukian, Pomoan, and Wintuan families included in the study sample.

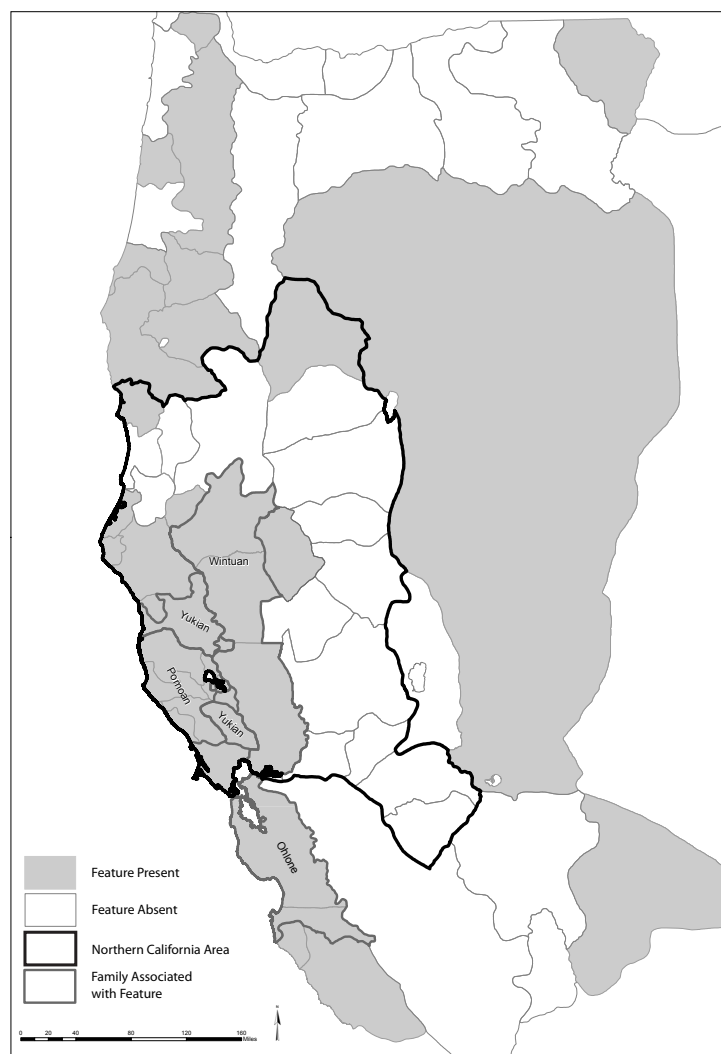


Figure 4.6: Distribution of five-vowel systems

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for the five-vowel system is not statistically significant. This vowel system is not distributed in a pattern that is significantly different than the patterns we would expect due to chance.

Table 4.3: Spatial autocorrelation of five-vowel system

Moran's I	p value
0.111131	0.115665

“Back” apical stops

“Back” apical stops are found in all of the languages of the Ohlone, Yukian, Pomoan, and Miwok families included in the study sample.

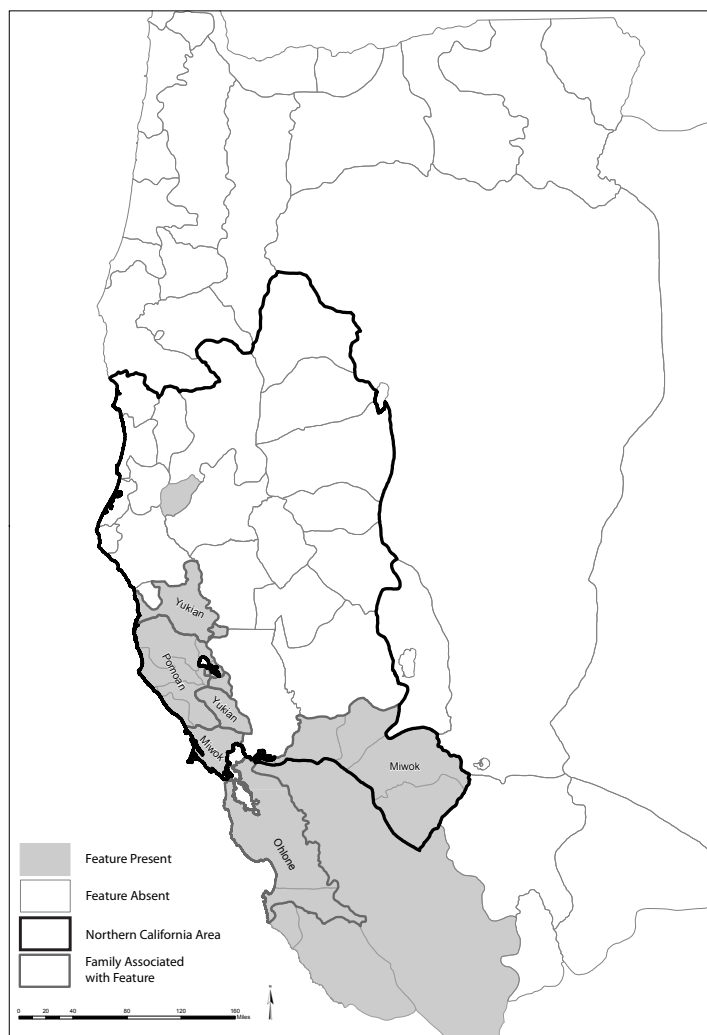


Figure 4.7: Distribution of “back” apical stops

The spatial autocorrelation value (Moran’s I) describing the geographic organization of the feature index (h) for “back” apical stops is very high and statistically significant. This feature is strongly clustered and occurs in a pattern that is very unlikely to have arisen due to chance, however this pattern is centered in Central California and extends only into Sierra Miwok and the Clear Lake area in Northern California.

Table 4.4: Spatial autocorrelation of “back” apical stops

Moran’s I	p value
0.556108	< 0.000001

“Back velar” consonants

“Back velar” consonants, including post-velar or uvular stops and fricatives, are found in all of the sampled languages of the California Athabaskan, Klamath-Modoc, Sahaptian, Chinookan, and Salish families.

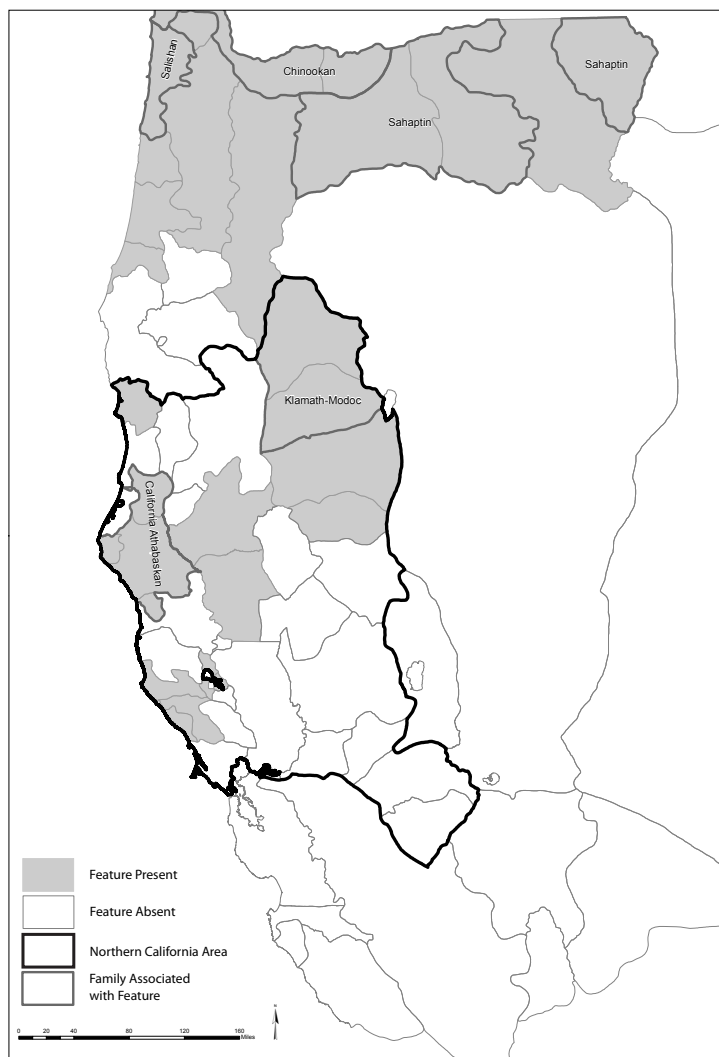


Figure 4.8: Distribution of “back velar” consonants

The spatial autocorrelation value (Moran’s I) describing the geographic organization of “back velar” consonants is moderately high and is statistically significant. This trait is distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.5: Spatial autocorrelation of “back velar” consonants

Moran’s I	p value
0.241682	0.001373

Looking specifically at the “back velar” stop (q), the pattern is again found that it occurs in all sampled languages from the California Athabaskan, Klamath-Modoc, Sahaptian, Chinookan, and Salish families.

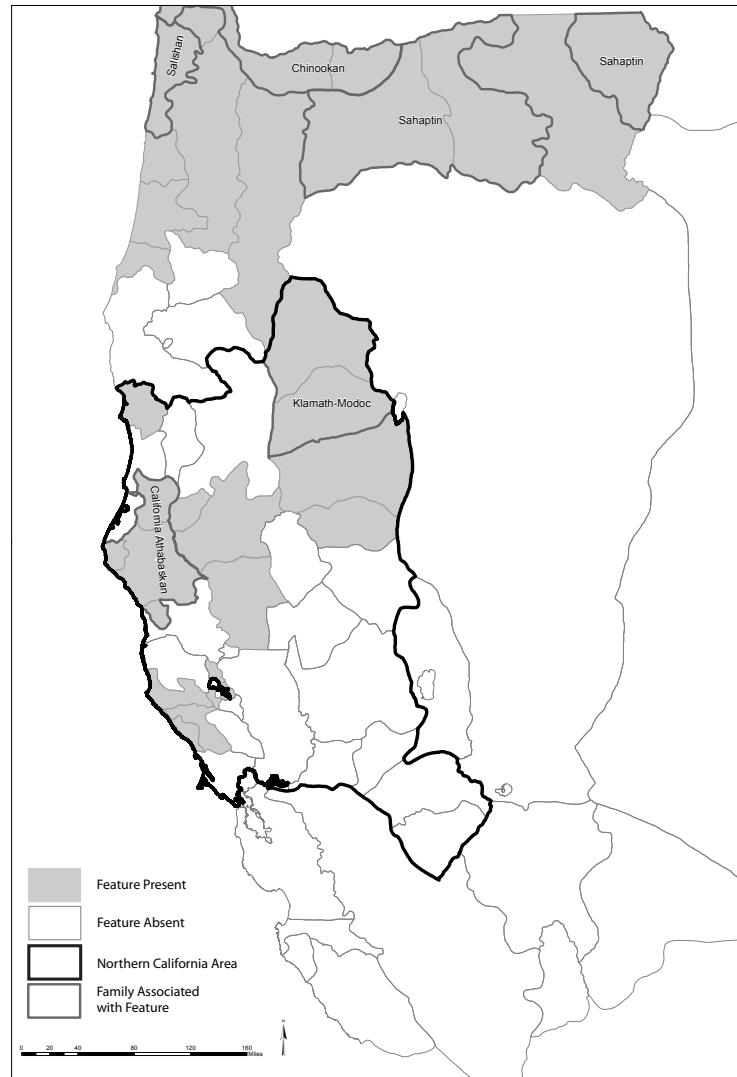


Figure 4.9: Distribution of “back velar” stops

The spatial autocorrelation value (Moran’s I) describing the geographic organization of the feature index (h) for “back velar” stops is again moderately high and is statistically significant. This type of consonant is spatially distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.6: Spatial autocorrelation of “back velar” stops

Moran’s I	p value
0.258784	0.000648

The “back velar” fricative occurs in a more restricted area, including all sampled languages from the Chinookan and Salish families.

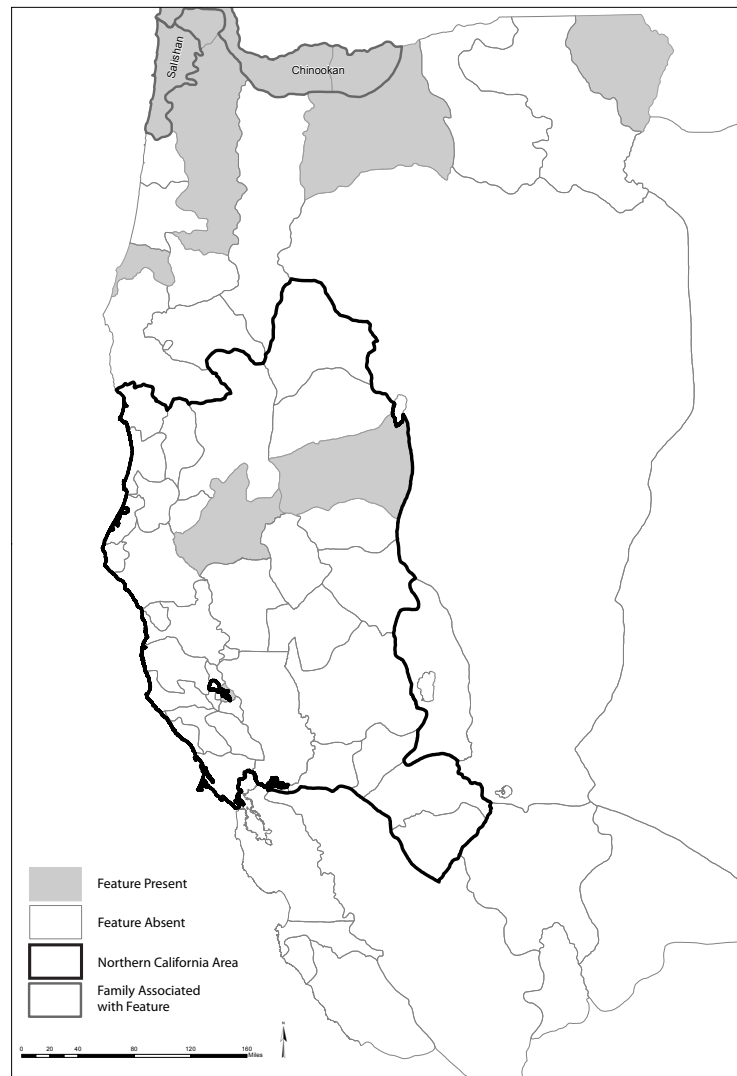


Figure 4.10: Distribution of “back velar” fricatives

The spatial autocorrelation value (Moran’s I) describing the geographic organization of “back velar” fricatives is somewhat lower but is statistically significant.

Table 4.7: Spatial autocorrelation of “back velar” fricatives

Moran’s I	p value
0.258784	0.000648

Fricated or affricated lateral consonants

At least one lateral fricative or affricate, including the glottalized lateral affricate, is found in all of the sampled languages of the Wintuan, Algic, California Athabaskan, Oregon Athabaskan, Klamath-Modoc, Sahaptian, Chinookan, and Salish families.

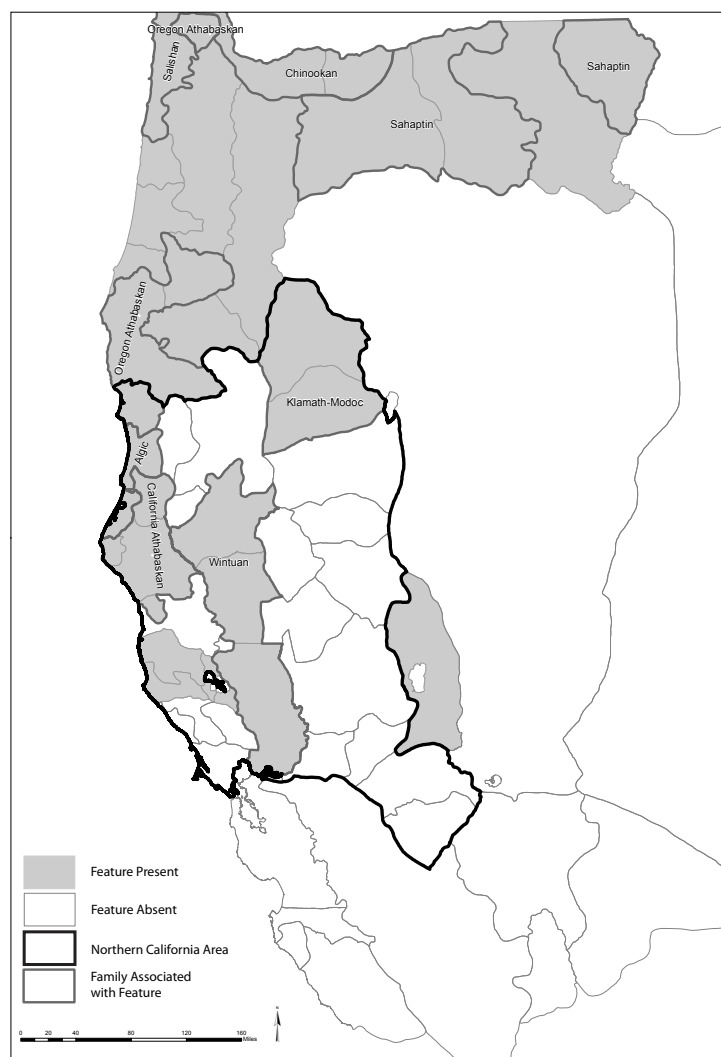


Figure 4.11: Distribution of fricated or affricated lateral consonants

The spatial autocorrelation value (Moran's I) describing the geographic organization of lateral fricatives and affricates is quite high and is statistically significant. This trait is distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.8: Spatial autocorrelation of lateral fricatives and affricates

Moran's I	p value
0.359817	0.000003

The lateral fricative, specifically, is also found in all of the sampled languages of the Wintuan, Algic, California Athabaskan, Oregon Athabaskan, Klamath-Modoc, Sahaptian, Chinookan, and Salish families.

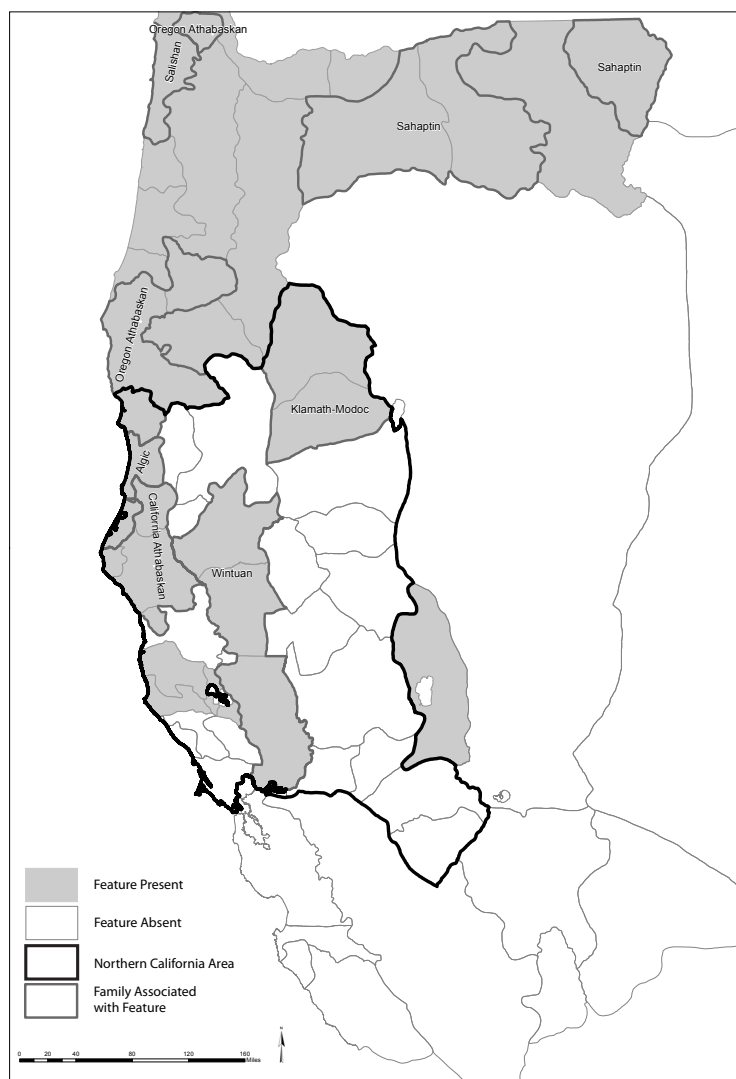


Figure 4.12: Distribution of lateral fricatives

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for the lateral fricative is quite high and is statistically significant. This consonant is distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.9: Spatial autocorrelation of lateral fricatives

Moran's I	p value
0.30007	0.000084

The plain lateral affricate is found in a more restricted area, including all of the sampled languages of only the Chinookan and Salish families. This sound does not occur phonemically in the California languages in this study area.

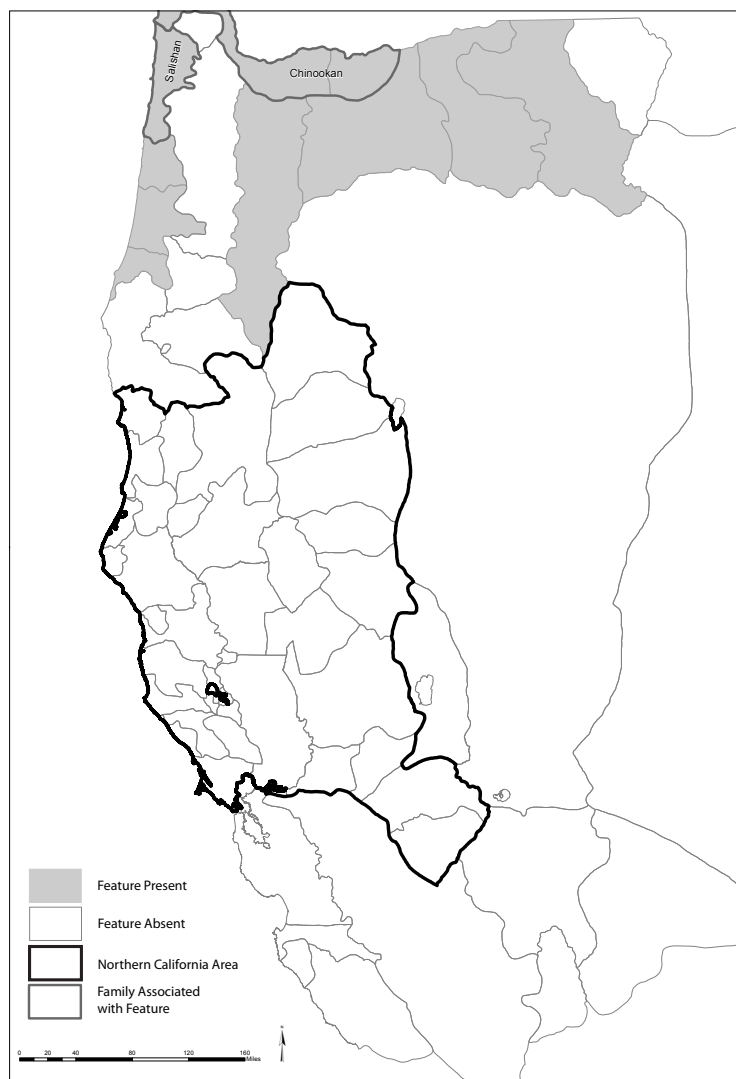


Figure 4.13: Distribution of plain lateral affricates

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for the plain lateral affricate is quite high and is statistically significant. This consonant is distributed in a pattern that is unlikely to have arisen due to chance, but that pattern falls entirely outside of Northern California.

Table 4.10: Spatial autocorrelation of plain lateral affricates

Moran's I	p value
0.289471	0.000102

The glottalized lateral affricate is found in a similar area, but also extends into Northern California. It occurs in all of the sampled languages of the Wintuan and Sahaptian families as well as all of the sampled languages of the Chinookan and Salish families.

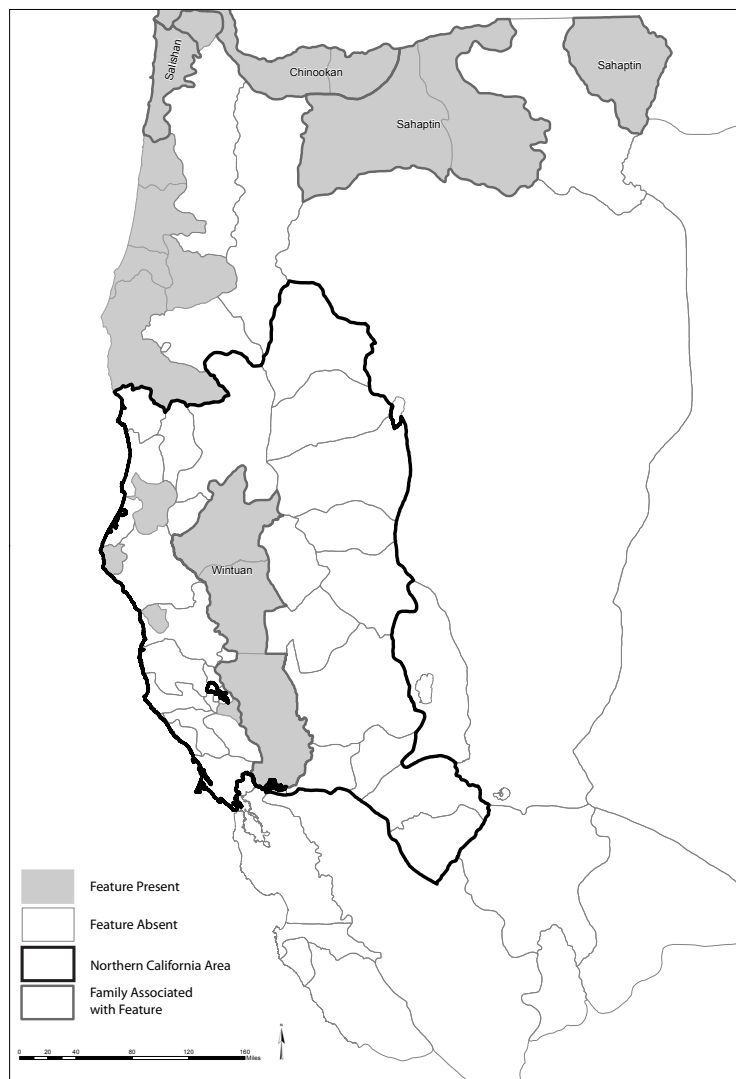


Figure 4.14: Distribution of glottalized lateral affricates

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for the glottalized lateral affricate is somewhat lower but is still statistically significant. This consonant is distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.11: Spatial autocorrelation of glottalized lateral affricates

Moran's I	p value
0.192313	0.008888

Contrast between r and l

A contrast between the rhotic and lateral approximants (r and l) is found in all of the sampled languages of the Ohlone, Wintuan, and Algic families.

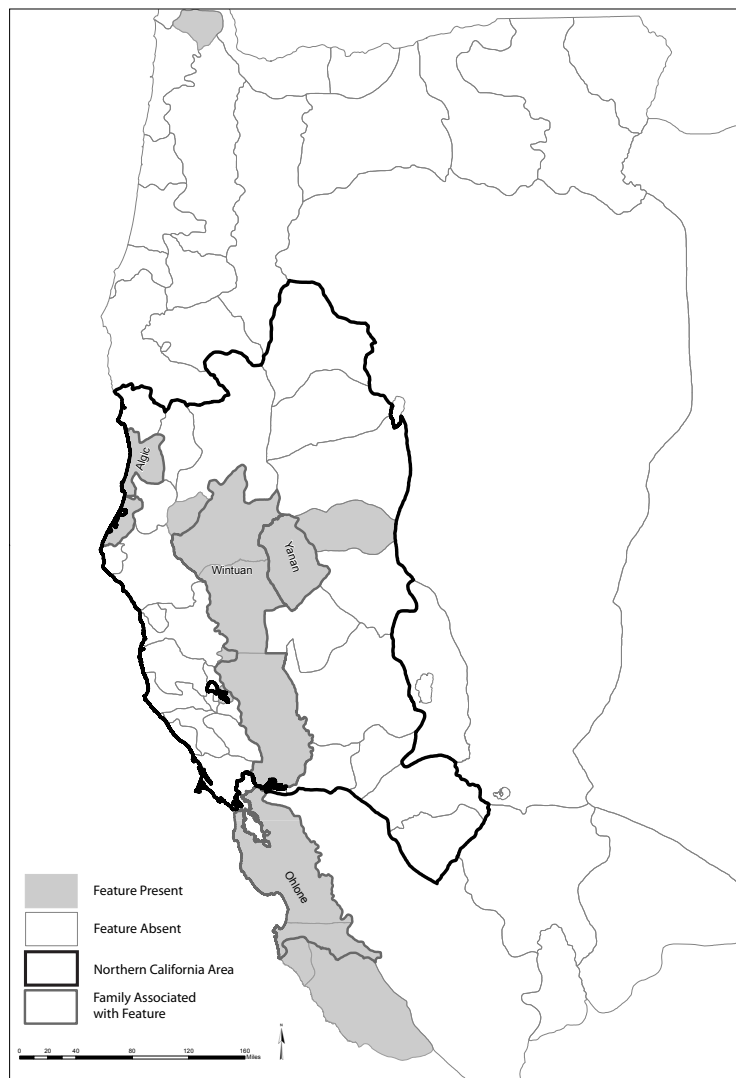


Figure 4.15: Distribution of contrast between r and l

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for the r/l contrast is moderately high and is statistically significant. This trait is distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.12: Spatial autocorrelation of the r/l contrast

Moran's I	p value
0.226348	0.001632

Velar nasal consonant

The velar nasal consonant is found in all of the sampled languages of the Uto-Aztecan and California Athabaskan families.

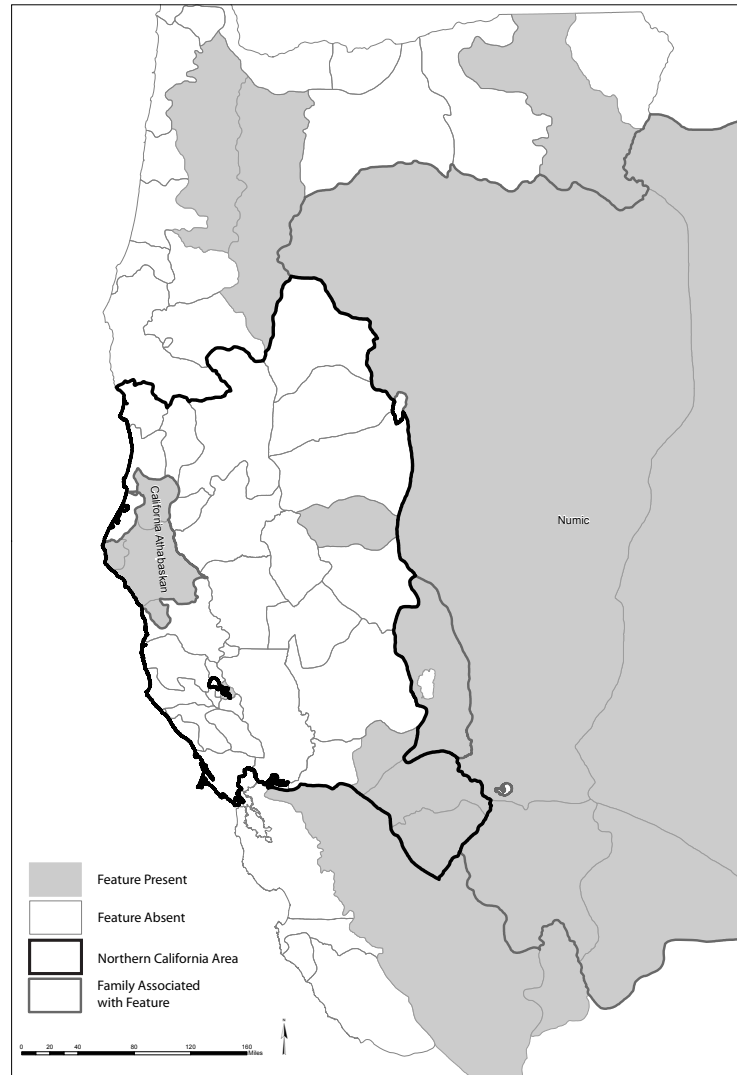


Figure 4.16: Distribution of velar nasal consonants

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for velar nasals is moderately high and is statistically significant. This trait is distributed in a pattern that is unlikely to have arisen due to chance.

Table 4.13: Spatial autocorrelation of velar nasal consonants

Moran's I	p value
0.234739	0.001632

4.3.2 Morphological Traits

Sound symbolic diminutivization

Sound symbolic diminutivization processes are found in all of the sampled languages of the Algonic, Klamath-Modoc, Chinookan, and Salish families. The spatial autocorrelation value

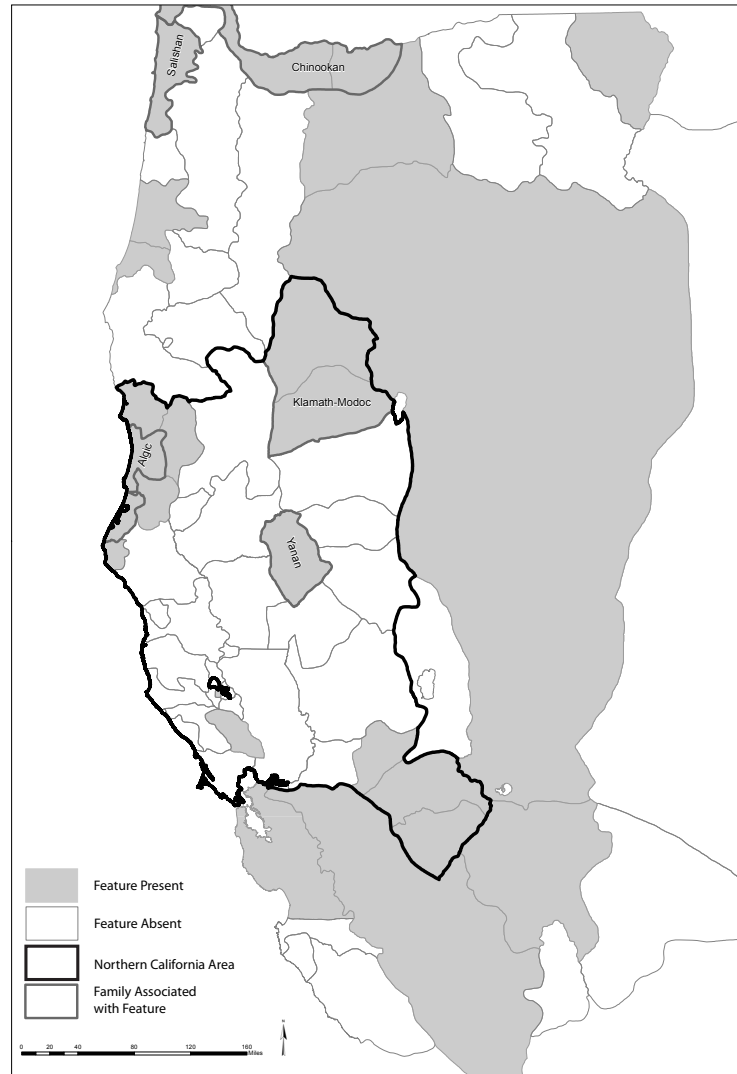


Figure 4.17: Distribution of sound symbolic diminutivization processes

(Moran's I) for sound symbolic diminutivization strategies is statistically significant and thus the distribution is more spatially organized than would be expected by chance.

Table 4.14: Spatial autocorrelation of sound symbolic diminutivization processes

Moran's I	p value
0.149348	0.039783

Quinary numeral systems

Quinary numeral systems are found in all of the sampled languages of the Wintuan, Shasta-Palainhnan, Klamath-Modoc, and Chinookan families.

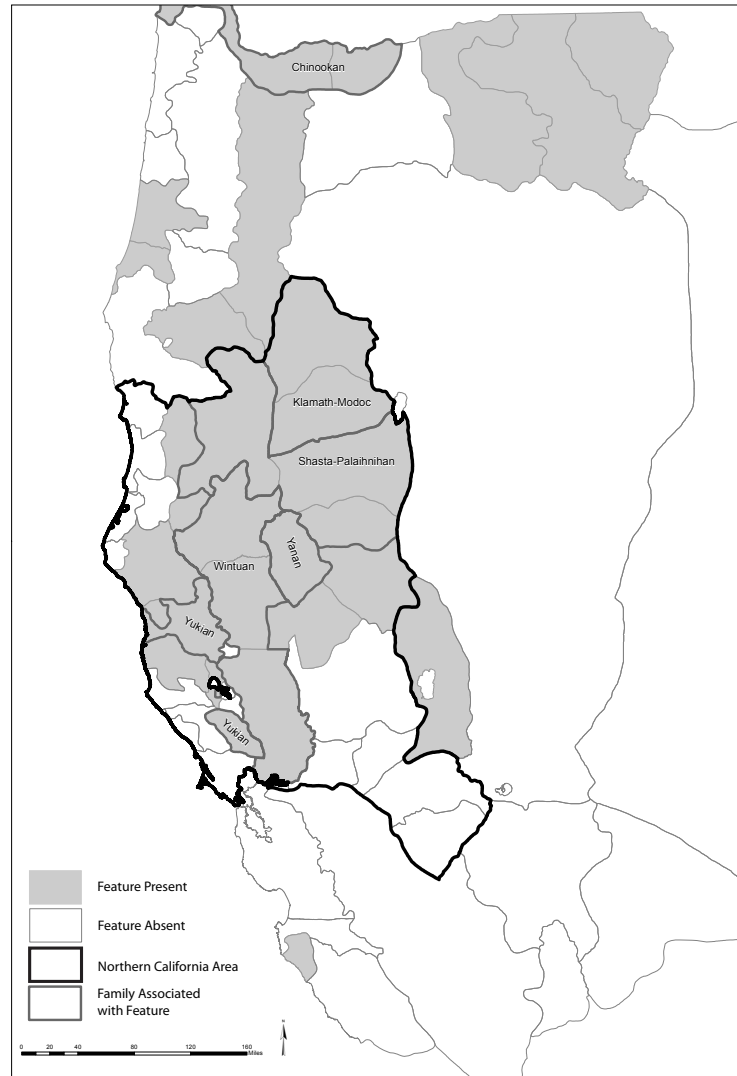


Figure 4.18: Distribution of quinary numeral systems

The spatial autocorrelation value (Moran's I) for quinary numeral systems is not statistically significant. While this feature is present in many Northern California languages, its distribution is not significantly different from the patterns that might occur due to chance.

Table 4.15: Spatial autocorrelation of quinary numeral systems

Moran's I	p value
0.104603	0.134877

Reduplication with distributive, iterative, and plural meanings

There are two kinds of reduplication that can be associated with these meanings: nominal and verbal. Nominal reduplication with a plural or distributive meaning occurs in all of the sampled languages of the Maiduan, Klamath-Modoc, Uto-Aztecan, Sahaptian, and Salish families.

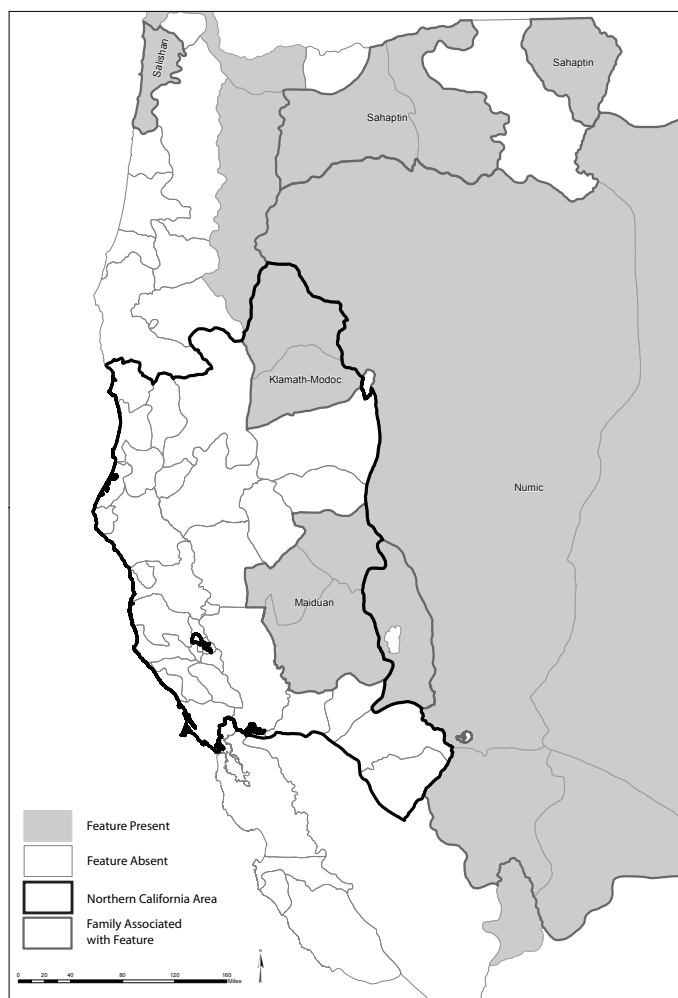


Figure 4.19: Distribution of nominal reduplication associated with plural meanings

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for nominal reduplication is moderately high and statistically significant. This feature is distributed in a more clustered pattern than would be expected by chance.

Table 4.16: Spatial autocorrelation of nominal reduplication with plural semantics

Moran's I	p value
0.272819	0.000245

Verbal reduplication with an iterative or distributive meaning occurs in all of the sampled languages of the Pomoan, Miwok, Maiduan, Wintuan, Yukian, Klamath-Modoc, Uto-Aztecan, Sahaptian, and Salish families.

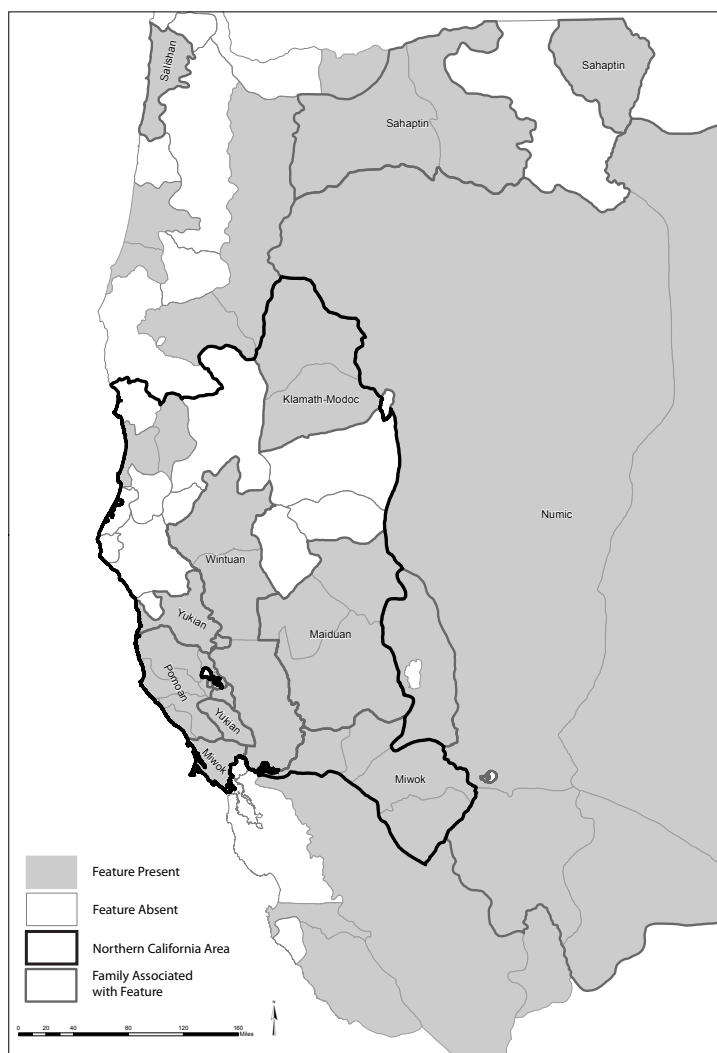


Figure 4.20: Distribution of verbal reduplication with iterative or distributive meanings

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for verbal reduplication is much lower and is not statistically significant. The negative value indicates that this feature is closer to being evenly distributed than significantly clustered, though the small I value indicates an essentially random pattern.

Table 4.17: Spatial autocorrelation of verbal reduplication with iterative or distributive semantics

Moran's I	p value
-0.004157	0.886133

Case marking on nominal phrases

Case marking occurs in all of the sampled languages of the Ohlone, Miwok, Pomoan, Wintuan, Yukian, Maiduan, Klamath-Modoc, Shasta-Palainhiyan, Uto-Aztecan, Sahaptian, and Chinookan families.

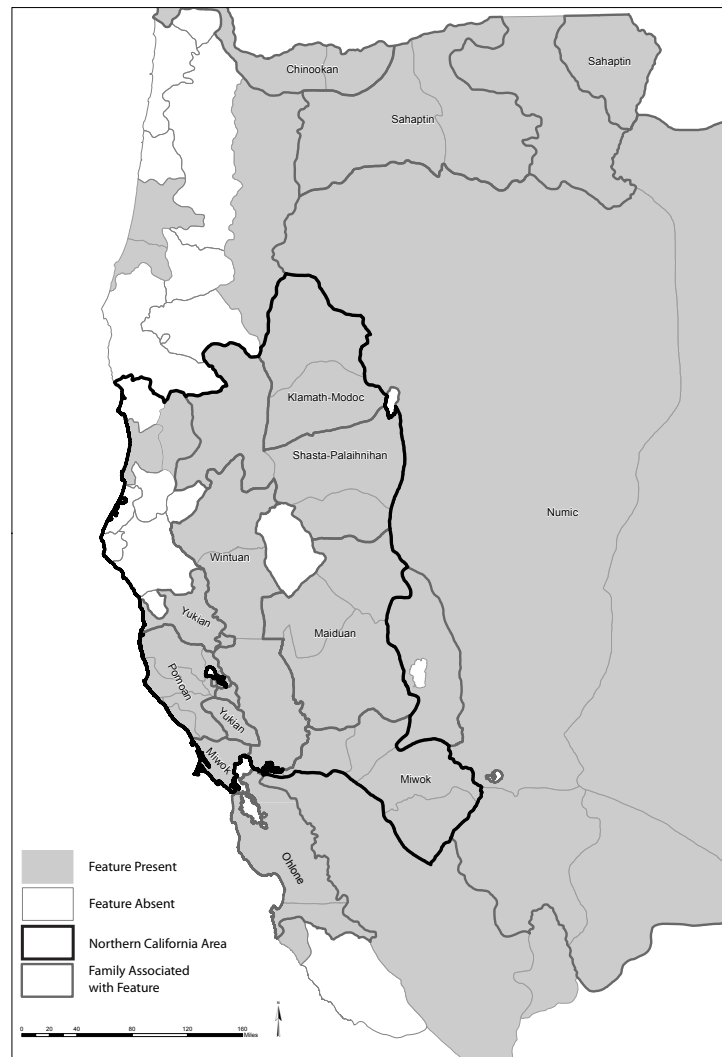


Figure 4.21: Distribution of case marking systems

The spatial autocorrelation value (Moran's I) for case marking systems is statistically significant. The distribution of this feature is more clustered than would be expected due to chance.

Table 4.18: Spatial autocorrelation of case marking systems

Moran's I	p value
0.15866	0.028749

***n*-initial first person pronoun and *m*-initial second person pronoun**

First person pronouns with the initial consonant *n* are found in all of the sampled languages of the Maiduan, Klamath-Modoc, Uto-Aztecan, and Chinookan families.

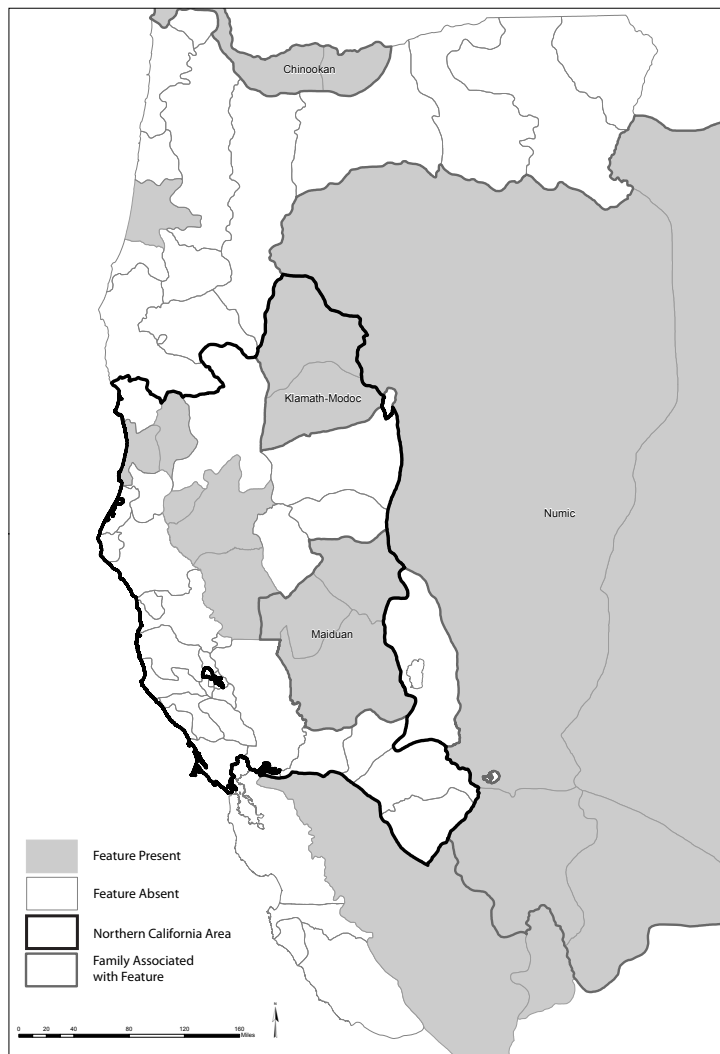


Figure 4.22: Distribution of *n*-initial first person pronoun

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for *n*-initial first person pronouns is low and is not statistically significant. The distribution of this feature is not different than might be expected due to chance.

Table 4.19: Spatial autocorrelation of *n*-initial first person pronouns

Moran's I	p value
0.032826	0.540233

Second person pronouns with the initial consonant *m* are found in all of the sampled languages of the Ohlone, Miwok, Pomoan, Wintuan, Yukian, Maiduan, Shasta-Palainhian, Klamath-Modoc, and Chinookan families.

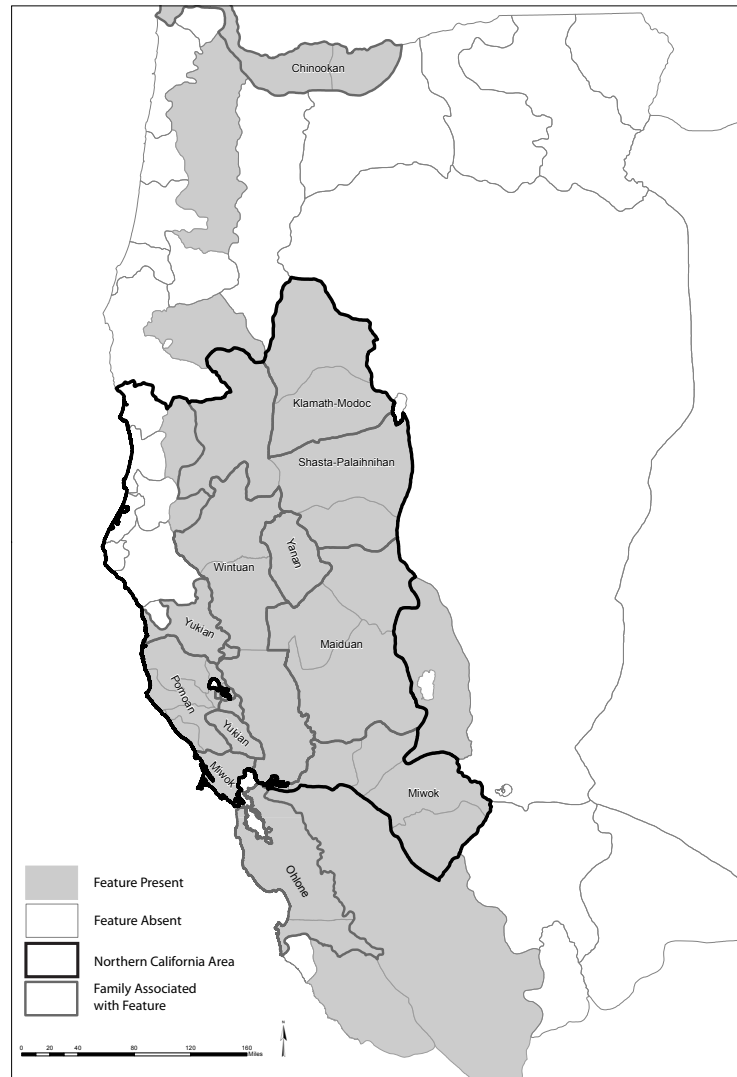


Figure 4.23: Distribution of *m*-initial second person pronoun

The spatial autocorrelation value (Moran's *I*) describing the geographic organization of the feature index (*h*) for *m*-initial second person pronouns is positive and statistically significant. The distribution of this feature is different than would be expected due to chance.

Table 4.20: Spatial autocorrelation of *m*-initial second person pronouns

Moran's <i>I</i>	<i>p</i> value
0.181926	0.013519

Prefixed possessive markers

Possessives are marked with prefixes in all of the sampled languages of the Ohlone, Pomoan, Wintuan, Maiduan, Yukian, Algic, California Athabaskan, Oregon Athabaskan, Chinookan, and Salishan families.



Figure 4.24: Distribution of prefixed possessive markers

The spatial autocorrelation value for prefixed possessive markers is small, negative, and not statistically significant. This feature's distribution is more random than geographically organized and resembles a chance-based distribution.

Table 4.21: Spatial autocorrelation of prefixed possessive markers

Moran's I	p value
-0.077843	0.435028

Prefixed subject markers

Prefixed subject markers are found in all of the sampled languages of the Shasta-Palainhihan, Algonic, California Athabaskan, Oregon Athabaskan, Sahaptian, and Chinookan families.

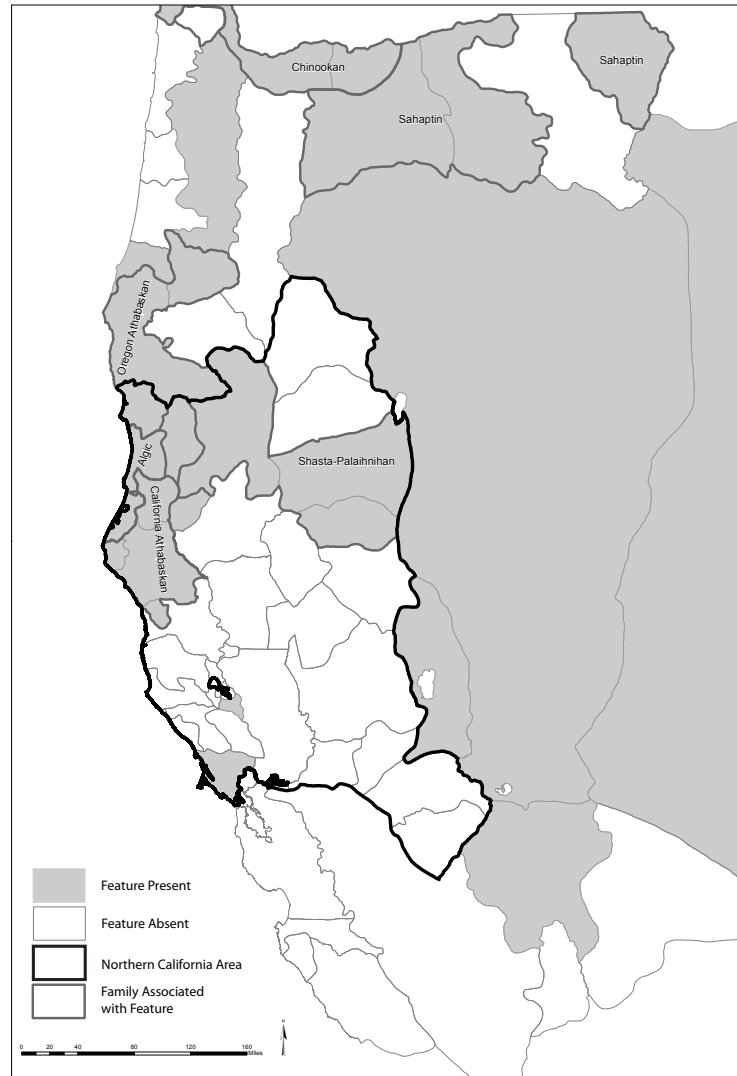


Figure 4.25: Distribution of prefixed subject markers

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for prefixed subject markers is not statistically significant. The distribution of this feature is not significantly different than might be expected due to chance.

Table 4.22: Spatial autocorrelation of prefixed subject markers

Moran's I	p value
0.08407	0.212255

Pre-verbal tense and aspect markers

Pre-verbal tense and aspect markers occur in all of the sampled languages of the Yukian, Algic, California Athabaskan, Oregon Athabaskan, Chinookan, and Salish families.

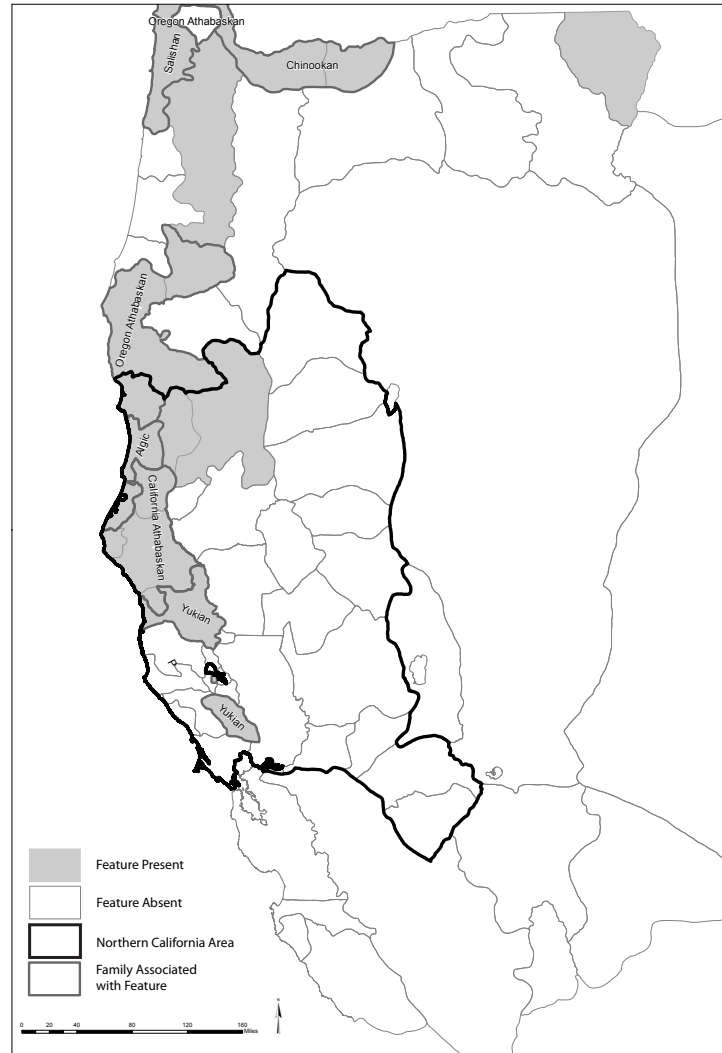


Figure 4.26: Distribution of pre-verbal tense and aspect markers

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for pre-verbal tense and aspect markers is not statistically significant. The distribution of this feature is not significantly different than might be expected due to chance.

Table 4.23: Spatial autocorrelation of pre-verbal tense and aspect markers

Moran's I	p value
0.16979	0.18663

Numeral classifier systems

Systems of numeral classifiers that express categories of shape or other characteristics of the quantified item are found in all of the sampled languages of the Algic and Salish families.

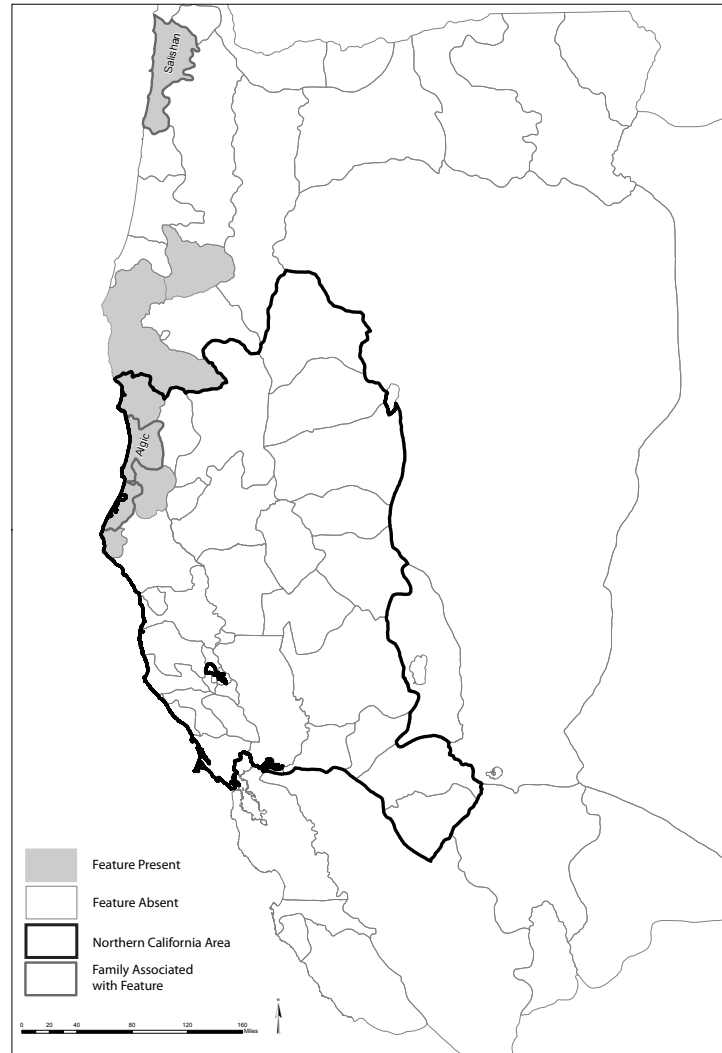


Figure 4.27: Distribution of numeral classifier systems

The spatial autocorrelation value (Moran's I) describing the geographic organization of the feature index (h) for numeral classifiers is positive and statistically significant. The very restricted distribution of this feature is significantly clustered and unlikely to have resulted from chance.

Table 4.24: Spatial autocorrelation of numeral classifiers

Moran's I	p value
0.261027	0.000266

4.4 Discussion

The patterns identified in Section 4.3 show that many of the features commonly assumed to have spread areally in Northern California have distributions that are unlikely to have arisen from chance and show significant geographic organization, even when possible genealogical sources are taken into consideration. These features are likely to have spread through geographic contact between languages. Other features commonly associated with Northern California are distributed in patterns in which no statistically significant geographic signal can be detected. These features may have spread through contact, but the likelihood that their current distributions represent areal spread rather than chance occurrence or genealogical inheritance is too low for the null hypothesis (i.e. no areality) to be rejected based on distributional facts.

The spatial patterns exhibited by these features help to determine the location and scale of areal feature diffusion in the region. Because this study examines a limited geographic area whose boundaries do not include the full extent of the spatial distributions of some features, further quantitative analysis is not appropriate for determining the scale of areal patterns in and around Northern California. Instead, the somewhat narrow focus on Northern California will be maintained and the geographic extents of feature distributions will be discussed in qualitative terms.

4.4.1 Evidence for Geographic Feature Diffusion in Northern California

Of the features analyzed in this study, more of the phonological features than morphological features show significant positive spatial autocorrelation. It is possible that this is simply a fact about this region, but it may also reflect something about the diffusion events through which different types of features have spread. For instance, phonological features may have spread through true geographic contact – that is the use of multiple languages in the same place at the same time (Thomason 2001:1) – or they may represent more recent geographic diffusion events whose geography has not been obscured by history. Morphological features may be more likely to have spread through ancient language shifts or even as a result of ancient relationships that do not leave as tidy a geographic trace in their wake. Further research is necessary to probe whether the balance of phonological and morphological feature diffusion in this region represents a meaningful historical difference.

The phonological features that are likely to have diffused geographically in this study area are:

- (4.3)
- “back” apical stops
 - “back velar” consonants
 - fricated and affricated laterals
 - *l/r* contrast
 - velar nasal (only in a sub-area of Northern California)

The morphological features that are likely to have diffused geographically in this study area are:

- (4.4)
- sound symbolic diminutivization
 - reduplication of nominals (but not reduplication of verbs)
 - case marking
 - *m*- initial second person pronouns (but not *n*-initial first person pronouns)
 - numeral classifier systems

The geographic patterns of all features are discussed below in Section 4.4.2, but the features listed above will be marked with an asterisk to designate them as highly likely to have diffused areally.

4.4.2 Geographic Patterns

None of the features investigated has a distribution that coincides with the boundaries of a Northern California area, and none occur entirely within the Northern California area. The vast majority of the traits have a patchy distribution in Northern California and are also found in neighboring regions. Discussion of the geographic distributions of features is organized based on six basic patterns below:

Features whose distributions encompass most of Northern California and parts of Central California

The traits whose distributions map most tidily onto the Northern California area are quinary numeral systems and *m*-initial second person pronouns*. Quinary numeral systems are found throughout much of the Northern California area as well as in parts of Oregon, particularly in the north. They are not found as frequently in the Northwestern California and Clear Lake micro-areas as in the rest of Northern California, though vestiges of a quinary system may exist in Yurok (Conathan 2004). The statistical insignificance of the spatial pattern in the distribution of quinary numeral systems may reflect the fact that this type of system is found without exception in several small Northern California language families. Golla also suggests that counting large numbers beyond the digits of one hand was probably largely unnecessary before the Late Period (800-1000 AD), and it is likely that many linguistic groups used quinary systems before the development of more complex trade economies. This would suggest that quinary systems may have been developed independently by many groups and inherited from ancestor languages of these small families rather than through functional convergence at a broad geographic scale.

Second person pronouns with an initial *m* are distributed in an area that includes most of Northern California, with the exception of some languages in Northwest California. This trait is also found in a few Oregon languages as well as Washo and the languages of California's Central Valley and Central Coast. Although the *n*-initial first person pronouns show weaker geographic patterning and are present in only a few Northern California languages, the *n-m* pronoun pattern has been recognized not only by Haas (1976) in Northern

California but also by Nichols and Peterson (1996), among others, as a pattern found widely in the western parts of North and South America. Nichols and Peterson conclude that this pattern is not due to chance or some typological universal but rather arises from some distant shared history among these languages. The second person m-initial pronouns found in California must be linked to this broader pattern and the deep historical connection it implies. Both quinary numeral systems and m-initial pronouns appear to have spread throughout Northern-Central California, but external evidence suggests that these features are in fact part of much larger patterns or tendencies.

Features found primarily within a subpart of the Northern California area

“Back” apical stops*, a contrast between rhotic and lateral approximants*, and numeral classifiers* are all found in small sub-areas within Northern California. “Back” apical stops are found in a highly clustered area around Clear Lake and extending through the rest of Miwok and into Yokuts and Ohlone. This distribution encompasses the southern part of the Northern California area as well as the better part of Central California. The *r/l* contrast is found in several Northern California families, notably the Wintuan and Yanan languages in the central part of Northern California, but also in the Algic languages of Northwest California. Numeral classifiers are found primarily in Northwestern California, extending into Oregon Athabaskan and also occurring, presumably independently, in Tillamook, the Salishan language in the northwest corner of the study area. We might expect these features to overlap in some core region of Northern California, but instead their distributions are mostly complementary and generally cover separate parts of the Northern California area.

Features found in much of the Oregon Coast in addition to Northern California

The contrast between plain, glottalized, and aspirated stops is found in the central area of Northern California, but it also extends up the Oregon coast into the far northwest corner of the study area, and is found in Central California’s Yokuts languages as well. As the map in Figure 4.4 shows, the feature is arrayed in a patchy manner even in the regions where it is common. Though there is good evidence for contact-based spread of this consonant distinction into Lake Miwok, it is unclear whether its occurrence outside of the Clear Lake area is the result of similar areal diffusions and just how widespread this pattern is on the Pacific coast. This feature is certainly associated with contact in the Clear Lake micro-area, but may also be a California or Pacific Coast macro-area feature. Glottalized stops, one of the three series that contrast in the aforementioned pattern, are found more consistently throughout Northern California and also throughout Central California, the Oregon coast, and the Plateau region of this study area. Neither the stop series contrast nor the glottalized stops themselves was found to have a significant geographic pattern, but it is possible that if studied at a wider geographic scale a significant pattern might be found in a large region of the Pacific Coast.

The five-vowel system is found along the Pacific Coast from the southern end of the study area to the northern end of the study area, with some exceptions such as Yurok⁵ and several

⁵Yurok’s rhotic vowel is derived through sound symbolism, and so it is probable that the historical system

languages of Northern Oregon. Five-vowel systems also occur in Northern Paiute, Panamint, and Nez Perce on the eastern side of the study area. Like the stop contrast mentioned above, this feature is not found to have a significant geographic organization to its distribution. The classification of the vowel system by number of contrastive vowels may be misleading, since the five-vowel system typical of California languages is [*i, e, a, o, u*], yet the Uto-Aztecan languages in this study area with five vowels have a high central vowel (*ɨ*) but not a mid front vowel (*e*). Separating these out, the California pattern would seem to occur along the Pacific Coast generally, rather than specifically in Northern California.

Prefixation of tense and aspect markers is found in the northwest portion of the Northern California area, as well as in Oregon Athabaskan and the Chinookan and Salish languages in the northwest corner of the study area. This pattern for inflecting verbs is associated by Sherzer (1976) with Northern California as well as the northern part of the Northwest Coast area, and the distribution shown in Figure 4.26 suggests that further analysis would be useful for determining how far up and down the Pacific Coast this feature truly extends and how continuous or patchy its distribution is. The clustering of this feature's distribution was not found to be significant.

Features found in Northern Oregon in addition to Northern California

The “back velar” consonants* occur in an assortment of Northern California's Hokan, Penutian, and Athabaskan languages that make up a rather spotty geographic coverage. The significance of this feature's spatial pattern is probably due mostly to its occurrence in a continuous band of languages across Northern Oregon. Sherzer (1976) associates back velars with both the Plateau and Northwest Coast areas, and these consonants have a strong presence in Northern Oregon. Their occurrence in Northern California is less straightforward. As mentioned in Section 4.1.1, Golla suggests that the clusters of “back velar” occurrence in Northern California – Athabaskan, Pomo, and Northern Wintuan as well as Klamath-Modoc and Palaihnihan – are probably historically linked (2011).

Features found in the Great Basin in addition to Northern California

Several features occur primarily in the Great Basin, but also extend into Northern California. The velar nasal consonant* is found in the languages of the Great Basin as well as in several Oregon and California languages which are adjacent to the Great Basin. In California it occurs in Sierra Miwok, Atsugewi, Yokuts, and California Athabaskan. While it must be a family trait of California Athabaskan, its occurrence elsewhere only in languages that share a border with a Numic language suggests that it is essentially a Great Basin feature that has spread beyond the Great Basin's borders in certain places.

Sound symbolic diminutivization processes* are patchier in their distribution both within Northern California and in other regions, including the Great Basin. Haas (1976) notes that this feature may be underreported since it is not a straightforward structural feature of language and is less likely to be discussed in grammars than traits like phonological inventories. This sort of sound symbolism is found in Western Numic in the Great Basin

included just five vowels.

as well as in a band of Central California languages including Sierra Miwok, Yokuts, and Ohlone. Its distribution within the Northern California area includes a cluster in and around the Northwest California sub-area, as well as a handful of other languages such as Wappo, Yana, and Klamath-Modoc. In Oregon it is also found in a somewhat random sample of languages, including the Chinookan and Salishan languages of the northwest and Nez Perce, Sahaptin, Coos, and Siuslaw. The primary geographic patterns in this feature's distribution is a cluster in Central California and the western Great Basin and another in the Northwestern California sub-area. It is notable that this feature is found in and around the corners of the Northern California area but is not, in fact, attested in the majority of the Northern California languages.

Nominal reduplication* is found in the languages of the Great Basin as well as in Klamath-Modoc and the Maiduan languages, two linguistic groupings which border the Great Basin. In Oregon this feature is found in the Salishan language Tillamook, as well as the Sahaptian languages, Chinook, and Molala. Like diminutive sound symbolism, this feature appears to be a Great Basin feature which has spread into some neighboring languages beyond the Great Basin border. Its occurrence in Tillamook is probably independent of the Great Basin diffusion pattern. Verbal reduplication also occurs in similar areas, but with greater geographic coverage. In addition to all of the Great Basin languages sampled, verb reduplication occurs in all of the languages of the southern end of the Northern California area (Pomoan, Wintuan, Yukian, Maiduan, Miwok) as well as the Central California languages. Farther north it occurs in Yurok, Karuk, Klamath-Modoc, and several Oregon languages. The finding that this feature's distribution is essentially random (see Table 4.17) indicates that while this is a common trait in languages throughout the entire study area, its distribution does not suggest any geographically organized pattern of diffusion.

Widely distributed features

Perhaps the most common geographical pattern is a broad spatial coverage that extends into multiple regions traditionally classified as linguistic areas. Lateral fricatives* are found in a much broader area than lateral affricates, but their distribution covers all of Oregon, except for Oregon Northern Paiute and Shasta. In California these consonants occur in the Northwest sub-area, in the Wintuan languages and in some Pomoan languages. Klamath-Modoc and Washo also have a voiceless lateral approximant (*L*) that is sometimes pronounced as a lateral fricative (Golla 2011:206). This feature's distribution is strongly linked to the Northwest Coast and Plateau areas (Sherzer 1976), and it is possible that its presence in California should be considered an extension of this larger pattern.

Case marking* is another very common feature in this study area. It is found in the Great basin, most of Northern California with the exception of the Northwest sub-area and surroundings, in Ohlone and Yokuts farther south, and in Sahaptian, Chinookan, and other Plateau languages. Whereas verb reduplication occurs with great frequency but little geographic order throughout the study area, case marking forms a fairly strong cluster including the Great Basin and the languages to the north and as far west as the San Francisco area. It is notably infrequent along the northern Pacific Coast in Northwestern California and Oregon. While this would seem to indicate a broad region throughout which a case marking system

has spread, the case marking systems represented by this feature differ from language family to language family, and so this pattern should not be assumed to correspond to a single broad feature diffusion. This pattern more likely represents a number of smaller diffusions of specific case marking systems.

Prefixed subject markers and prefixed possessive markers both occur in spotty fashion throughout the study area. Neither of these features occurs with great regularity in any of the conventional linguistic areas included in the study area. The low magnitude and statistical insignificance of the spatial autocorrelation measures for these two features represents the lack of organization in their distributions. Some very local diffusion events may have spread these features in small pockets of the study area, but they do not appear to be characteristic features of a Northern California linguistic area or any other proposed linguistic area within this study site.

Summary

The features whose distributions match the spatial extent of Northern California most closely – quinary numeral systems and *m*-initial second person pronouns – are known to be much more widespread phenomena. Furthermore, spatial analysis has not provided any evidence that the distribution of quinary numeral systems in Northern California is likely to be a result of geographic diffusion. One or more features whose spatial extent closely matched the extent of this region would support the classification of Northern California as a linguistic area, but neither quinary numeral systems nor *m*-initial second person pronouns coincides closely enough with the extent of the Northern California area to be the defining feature of this proposed linguistic area.

4.4.3 Overlap of Features

Two other patterns could suggest that Northern California is indeed a salient area of feature diffusion. The first is a core/periphery pattern, much like the isogloss patterns frequently found in dialect areas, where some core set of features is found at the heart of the region, with non-identical but overlapping feature distributions leading to smaller numbers of features shared in the area's periphery (Thomason 2001:101). The second pattern is one in which a set of traits with patchy distributions are found in consistently greater numbers among languages within the area than languages outside the area, though each language may contain a different combination of traits. This pattern's use in defining linguistic areas whose features are not evenly distributed is suggested by Conathan's mention of similar, high numbers of Northwest California traits found in the languages of that area (2004:167).

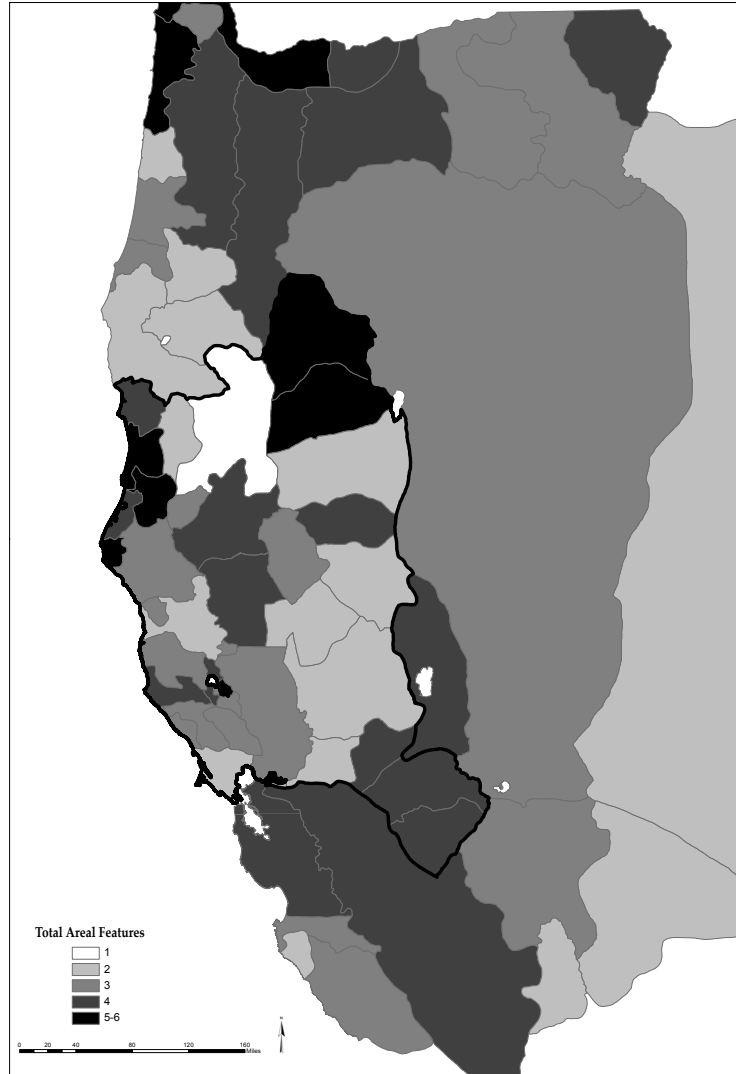


Figure 4.28: Total number of significant areal features found in Northern California and surrounding areas

Looking at a map of the total number of significantly geographically patterned features associated with each language (Figure 4.28), no impressionistic core/periphery pattern is found. The same spatial autocorrelation measure used in Section 4.3 can be used to test this impression, since what has been described as a core/periphery pattern would be manifested in an areal feature count map as a highly spatially autocorrelated cluster.

Table 4.25: Spatial autocorrelation of significant areal feature counts

Moran's I	p value
0.164489	0.054059

The low spatial autocorrelation value for this feature count pattern and its statistical insignificance at the $p < 0.05$ level indicate that no core/periphery pattern is found in this

data. Again an impressionistic appraisal of the map yields no evidence that languages within the Northern California area have more of the areal traits than languages outside of the area. A simple t -test confirms that the numbers of features associated with the Northern California languages are not statistically different than the numbers of traits associated with non-Northern California languages ($p = 0.401$).

4.5 Conclusions

Although Northern California seems to be defined by the presence of certain linguistic features, such as “back” apical stops and sound symbolic diminutivization marking, the majority of these features are unevenly distributed in the Northern California area and/or extend across far broader regions. The actual scales and locations of these features rarely coincide with the traditional boundaries of Northern California. Furthermore, in a region like Northern California where knowledge about linguistic history is also uneven, it is difficult to know whether features that occur in a particular spatial pattern have arrived at that distribution through areal diffusion, common inheritance from a genealogical ancestor, or by chance. This study addresses these questions about the Northern California linguistic area through mapping and spatial analysis. The feature index (h) introduced here is a simple way of accounting for the likelihood of a genealogical source for a particular trait in a particular language when analyzing spatial patterns in feature distributions. The use of spatial autocorrelation statistics makes it possible to identify significant geographic patterns, which are more likely to arise from geographic feature diffusion, and random geographic patterns, which are more likely to have arisen by chance. This procedure helps to determine which of the features in Northern California may have spread areally in the region without knowing the details of individual diffusion events. Patterns identified in the distributions of individual features and in the cumulative feature set are important for evaluating Northern California as a linguistic area.

The findings of this analysis suggest that Northern California is not a linguistic area. No feature distribution is matched well enough to the boundaries of Northern California to serve as a basis for classifying this region as a linguistic area. No core/periphery pattern is found in the cumulative distribution of the region’s features. And no significant difference is found between the number of “Northern California features” found in Northern California and in the surrounding areas. The conclusion that must be drawn from these findings is that Northern California is not, in fact, a true linguistic area.

Well-documented linguistic micro-areas are found in the northwest corner of Northern California (the Northwestern California area noted by Bright and Bright 1965; Haas 1967, 1970; Conathan 2004; O’Neill 2008) and the southwest corner (Mithun’s Clear Lake area). Other areal features, illustrated by many of the maps in Section 4.3, are found in larger geographic areas. The impression that Northern California is a single linguistic area may be influenced by the presence of smaller micro-areas and extensions of Northwest Coast, Basin, and even Plateau features into Northern California. These patterns suggest that Northern California might be more accurately described as a collection of micro-areas or a peripheral subregion of the Northwest Coast.

The methodology introduced in this chapter is a first step toward a geographic approach to the study of areal feature diffusion that can help to distinguish areal, genealogical, and chance patterns in linguistic feature distributions. Like any new method, it will benefit from future refinements. However, it has provided answers to the questions asked by this study and has opened the door to more quantitative rigor in the assessment of linguistic areas. Promising avenues for improving this approach are numerous. Conducting similar studies in larger study areas would add some robustness to these findings and allow for the identification of larger-scale patterns. Further geographic analysis could also be done to provide a more concrete measure of the geographic scales at which we find evidence of feature diffusion.

Chapter 5

Conclusion

This dissertation has focused on two primary goals:

1. To make progress on several seemingly intractable problems in Californianist historical linguistics.
2. To explore methodologies that can identify patterns in linguistic data, test the significance of these patterns, and be meaningfully related to processes of historical change.

The three case studies presented in Chapters 2 through 4 represent three different scales of language change, and demonstrate the use of new and underutilized quantitative methods for answering questions about California's linguistic history at each of these scales. Each chapter presents new evidence on a question in Californianist historical linguistics, summarized below in Section 5.1, and implements a methodology that was developed or chosen specifically to answer the question at hand. These methodological innovations are discussed in Section 5.2.

5.1 Summary of Findings

Chapter 2 presents evidence from isogloss patterns and linguistic distance measures to support the boundaries traditionally drawn within Eastern Miwok, but it also suggests that not all of these boundaries represent language-level divisions. In particular, the boundary between Northern Sierra Miwok and Central Sierra Miwok is limited to a few lexical isoglosses and no phonological differences, and distance measures show that this boundary is weakest at upper elevations. While the status of Plains Miwok is unchallenged by the findings of this chapter and Southern Sierra Miwok appears to have been an emerging language, Northern Sierra Miwok and Central Sierra Miwok together are probably better characterized as a single language with internal dialect diversity.

This chapter also finds evidence of uphill and downhill dialects within all of the Sierra Miwok languages. These emerging dialects may reflect patterns of contact associated with environmental adaptation and a preference for migration within similar ecological zones, rather than across distinctly different climates and vegetation regimes. The idea that elevation and vegetation are influential in the development of dialect networks is supported

by the cost distance modeling component of this chapter. Environmental influences act on linguistic contact at a spatial scale roughly similar to the territories traditionally ascribed to individual Eastern Miwok “languages”. While none of these geographic parameters appears to operate at a whole-Eastern Miwok scale, significant relationships between environmental costs and linguistic distances do exist in this dialect network. Elevation plays a significant role in contact within each of the three Sierra Miwok divisions. Southern Sierra Miwok shows the additional influence of vegetation, Central Sierra Miwok shows the additional influence of surface water, and in Northern Sierra Miwok all three of these factors are found to be significant influences on dialect network development.

Chapter 3 implements a test for statistically significant phylogenetic signals among California languages as a way to examine hypotheses that the well-supported small families of the region are genealogically linked through very deep relationships. This test accurately identifies clusters associated with the independently established Maiduan, Miwok, Pomoan, Yuman, and Uto-Aztecan families, and also finds evidence of a link between Shasta and Achumawi. However, this test fails to find any evidence of Hokan or Penutian, or any other cluster larger than the established families. It is possible that Hokan and Penutian exist and are simply too temporally distant to be within the range of this test, but a simpler conclusion to draw is that it is highly unlikely that the Hokan and Penutian classifications represent genealogical relationships.

Chapter 4 identifies several features whose distributions suggest that they have spread through geographic diffusion in Northern California. These features are “back velar” consonants, fricated and glottalized lateral consonants, an *r/l* contrast, sound symbolic diminutivization strategies, reduplication of nominals (but not verbs), case marking, *m*-initial second person pronouns (but not *n*-initial first person pronouns), and, in smaller subareas of Northern California, velar nasal consonants, numeral classifier systems, and “back” apical stops. Only one of these features, *m*-initial second person pronouns, occurs in a distribution whose north, east, and west boundaries correspond roughly with the Northern California area. This feature’s extension into Central California and its association with a broader pattern in the Americas make it insufficient as a basis for classifying Northern California as a linguistic area.

Analysis of the overall patterns of feature distribution and overlap in Northern California similarly failed to identify evidence of a Northern California linguistic area. Neither a core-periphery pattern nor an overall high occurrence of probable areal features was identified in Northern California. In sum, the lack of significant spatial patterns associated with the Northern California linguistic area suggests that this region may not, in fact, represent the scale and location of the diffusion events associated with theorized “Northern Californian” traits. The Northwestern California and Clear Lake areas contained within Northern California are better supported as linguistic areas. Northern California itself appears to be overlapped to the east by the Great Basin, a possible linguistic area from which several traits extend into Eastern California. To the north, a number of features found within California are also associated with the Northwest Coast area, and it is likely that Northern California is a peripheral part of this larger, better supported linguistic area.

5.2 Discussion of Methodological Advances

Chapter 2 employed aggregate measures of linguistic data, which have been developed and refined within the field of dialectometry, to quantify the similarity between varieties of Eastern Miwok. These aggregate measures are a particularly effective tool for analysis in a language family like Eastern Miwok, where the dialect data that exists includes missing items and researcher-related noise. The primary innovations of this chapter were the use of cost distance modeling to evaluate hypotheses about language contact and the comparison of distance matrices using Mantel correlations to characterize dialect networks and evaluate how environmental factors may have influenced their development. This cost distance modeling process makes it possible to understand travel in a world with significantly less infrastructure than currently exists, and allows us to test very specific hypotheses about contact. The procedure implemented in Chapter 2 is particularly well suited to the Eastern Miwok case, where data limitations rule out other methods and dialect diversification has occurred in a very physically and ecologically diverse setting. However, this methodology could also be implemented in other situations, both geographically broader and historically deeper, where understanding the likely paths of human contact or migration might improve our ability to evaluate linguistic hypotheses.

Chapter 3 implemented a procedure involving Kessler’s R^2 , a metric which was developed and published over ten years ago but has not yet been applied to many of the cases where it would be particularly useful. This metric quantifies the recurrence of sound correspondences, and together with a Monte Carlo test for significance and a multilateral clustering method, it is a sensitive way to test for very deep genealogical relationships. This study introduced several innovations to the method outlined in Kessler (1999, 2001). First, the methods used in Chapter 3 implement a normalization scheme that makes it possible to compare languages with phonological inventories of very different sizes. This normalization method is simple – expressing the correspondence metric as a ratio of the actual Kessler’s R^2 to the mean value generated by the Monte Carlo simulation – but it is crucial for accurate comparisons of languages with very different sound systems. This study also made explicit the procedure for conducting multilateral comparison in this type of test. Specifically, when comparing a language to a cluster, the data from the individual language are compared to the corresponding combinations of data that represent each cluster member’s values. Finally, this study selected a very low α value to serve as the acceptance level for statistical significance. This decision introduced a sort of informal Bonferroni correction to the method to control for the possibility of type I error (false positives) that is frequently associated with multiple comparison.

Chapter 4 developed a simple index for characterizing the occurrence of linguistic traits that accounts for the likelihood that they were genealogically inherited. This index is adjusted according to whether the trait occurs in all other members of a family, but in a study where larger language families are included in the sample, the associations between language families and linguistic traits could be established through statistical independence tests. The feature occurrence index developed in Chapter 4 was used to map linguistic traits and identify patterns of spatial autocorrelation in their distributions. By identifying spatial patterns that were unlikely to have occurred by chance that were also unlikely to represent

genealogical inheritance, this method was able to characterize features as likely or unlikely to have diffused areally. Spatial autocorrelation was also measured to test whether the aggregate feature patterns in a hypothesized language area showed a sort of core-periphery pattern. Spatial autocorrelation, used in the manner demonstrated in Chapter 4, is an effective tool for testing hypotheses about linguistic areas and areal traits.

5.3 Theoretical Impacts

The findings of the studies presented in Chapters 2–4 bear on several of the broad theoretical questions introduced in Section 1.2. The dialect study in Chapter 2 demonstrates links between several elements of the physical environment and the development of patterns in Eastern Miwok diversity. These findings not only contribute to our understanding of how divisions within Eastern Miwok have arisen but also exemplify the types of effects arising from environmental constraints on linguistic history that we should strive to recognize. This study also finds some specific, quantifiable patterns in the diversification of Eastern Miwok languages that help to answer questions about the outcomes of contact between closely related language varieties. The finding that the area traditionally associated with Southern Sierra Miwok is both very different from the rest of Sierra Miwok and contains within it evidence of fairly linguistically distinct and geographically orderly dialect divisions can be related specifically to characteristics of the environment and the resulting contact dynamics. The relatively high costs of travel imposed by this region’s rugged terrain and the ecological and resource diversity of the high valleys where many Southern Sierra Miwok settlements were located suggest that mobility was lower in this region than in the areas associated with the traditional Central and Northern divisions of Sierra Miwok. In contrast, Central Sierra Miwok is the least distinct and least orderly part of the study area, and is found in a much gentler terrain where we would expect greater mobility. The association of greater linguistic differences from neighboring varieties, and more distinct internal differentiation with the area where we would expect more restricted contact echos Ross’s (2003) association of tight-knit communities with the maintenance of linguistic differences in polylectal situations, but the relationship between language and geography in Central Sierra Miwok fits less neatly into Ross’s framework. This suggests that the relationships between conditions and outcomes of language/dialect-internal contact are somewhat different than those that have been proposed for polylectal contact. While it would be premature to propose the sort of generalizations between conditions and outcomes that have been put forth for cross-linguistic contact based on just one study, the Central Sierra Miwok facts suggest that social networks that are open to contact from very similar lects might be linked with shallow but complex patterns of variation across the network rather than the calquing and convergence that is found in open polylectal communities.

Chapters 3 and 4 both present negative results regarding the detection of very deep genetic and areal signals in California. At first blush this might suggest that the importance of genealogical inheritance and areal diffusion for explaining widespread similarities among languages is simply not something that can be examined at this temporal scale. But these results are perhaps better interpreted as showing that the resemblances between these lan-

guages are neither wholly genealogical in nature, nor are they derived entirely through the regional diffusion of linguistic traits. The tests discussed in Chapters 3–4 introduce ways to objectively evaluate the significance of genetic and areal signals in linguistic data and while the hypotheses they were applied to in this dissertation are not supported, the methods are validated by the support found for well-established genealogical and areal relationships. Rather than proving discouraging, the results suggest that more progress could be made in understanding this region’s linguistic prehistory by more carefully considering the interplay of genealogical and areal processes, reformulating hypotheses, and continuing work on detecting and characterizing the genetic and areal signals that may be discoverable.

5.4 Directions for Future Work

The approaches to diachronic questions introduced in this dissertation could be extended in many different ways and applied in a wide variety of regions and studies. However, a few particularly promising directions exist for future research. The quantitative but ecologically sensitive approach to dialect geography developed in Chapter 2 is highly sensitive to scale and local patterns in human-environment relations. Yet by carefully implementing it in other regions, where different social and environmental patterns exist, it may be possible to identify larger themes regarding the environmental factors that affect linguistic contact and the scale at which these impacts are incurred. The languages of the Eastern Sierra Nevada and California’s Central Valley would serve as good starting points for this work, since they involve distinct linguistic groups that bear some cultural resemblance to Eastern Miwok, but whose dialect geography has developed in the context of very different physical landscapes.

Directions for improving the phylogenetic study in Chapter 3 are more straightforward. One priority is to implement a more complex phonological coding scheme to prevent false negative results like the absence of the Palaihnihan family from the clustering results. Another improvement to this study that could be implemented is the collection of data from a larger number of California languages and a new test using this larger sample. Retesting with a larger sample would lead to a more convincing result, even if that test also finds no evidence for deep relationships.

Finally, the areal study in Chapter 4 could be extended by broadening the study area and testing patterns in a larger geographic area including the Northwest Coast and the Great Basin. A more sophisticated implementation of this methodology would also take into consideration typological and broad geographic tendencies of the sort discussed by Nichols (1992). Future work could also implement the Kulldorff and Nagarwalla’s statistic discussed by Donohue and Whiting (2011) to characterize the scale and extent of areal patterns without relying on assumptions about language and culture spread or biases about regional boundaries.

This dissertation has demonstrated that progress on even the thorniest of historical questions can be made through the creative but careful implementation of quantitative methods. In California, the availability of data may be an obstacle to investigating certain questions, but the quantitative approaches embraced in this dissertation overcome this impediment through the use of statistics. With sophisticated new technologies and powerful computers,

it is possible to answer many of the questions about California's linguistic history that have not previously found solutions. The methods explored in this dissertation are promising tools for more general use as well, even in situations where there is not insufficient data for more traditional approaches. The techniques employed in these studies to understand the histories of California languages can be used in a wide variety of contexts to detect areal and genealogical signals in linguistic data and examine the interplay of descent and diffusion in linguistic history.

Bibliography

- Allen, J. and K. Lu. 2003. Modeling and prediction of future urban growth in the Charleston region of South Carolina: a GIS-based integrated approach. *Ecology and Society* 8:2.
- de Angulo, Jaime and L. S. Freeland. 1930. The Achumawi language. *International Journal of American Linguistics* 6:77–120.
- de Angulo, Jaime and L. S. Freeland. 1931. The Lutuami language. *Journal de la Société des Américanistes* 1:1–45.
- de Angulo, Jaime and L. S. Freeland. n.d. Jaime de Angulo and L.S. Freeland papers. Survey of California and Other Indian Languages, University of California, Berkeley.
- Aoki, Haruo. 1970. *Nez Perce grammar*, volume 62 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Balstrøm, Thomas. 2002. On identifying the most time-saving walking route in a trackless mountain terrain. *Geografisk Tidsskrift, Danish Journal of Geography* 102:51–58.
- Banks, Jonathan. 2007. The verbal morphology of Santiam Kalapuya. *Northwest Journal of Linguistics* 1:1–98.
- Barrett, S. A. 1908a. *The geography and dialects of the Miwok Indians*, volume 6 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Barrett, S. A. and Edward Winslow Gifford. 1933. *Miwok material culture*, volume 2 of *Milwaukee Public Museum Bulletin*. Milwaukee: Order of the Board of Trustees.
- Barrett, Samuel A. 1893-1977. Samuel Alfred Barrett papers. BANC MSS 86/172c, The Bancroft Library, University of California, Berkeley.
- Barrett, Samuel A. 1906. Miscellaneous California Indian vocabularies and ethnographic notes. Ethnological Documents of the Department and Museum of Anthropology, University of California, Berkeley, BANC FILM 2216 , The Bancroft Library, University of California, Berkeley.
- Barrett, Samuel A. 1908b. *The ethno-geography of the Pomo and neighboring Indians*, volume 6.1 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.

- Beckman, Jill N. 1998. Positional faithfulness. Ph.D. Dissertation, University of Massachusetts Amherst. Ph.D.
- Bennyhoff, James A. 1977. *Ethnogeography of the Plains Miwok*, volume 5 of *Center for Archaeological Research at Davis Publications*. Davis: University of California Davis.
- Bergsland, Knut and Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3:115–153.
- Berman, Howard. 1983. Some California Penutian morphological elements. *International Journal of American Linguistics* 49:400–412.
- Bethel, Rosalie, Paul V. Kroskrity, Christopher Loether and Gregory A. Reinhardt. 1993. A dictionary of Western Mono, second edition.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart, and Winston.
- Boas, Franz. 1911. Chinook. In *Handbook of American Indian Languages, part 2*, ed. Franz Boas, volume 40.1 of *Bureau of American Ethnology Bulletin*, 559–677. Washington D.C.: Smithsonian Institution.
- Boas, Franz. 1929. Classification of American Indian languages. *Language* 5.
- Boas, Franz and Ives Goddard. 1924. Ts'ets'aut, an Athapaskan language from Portland Canal, British Columbia. *International Journal of American Linguistics* 3:1–35.
- Bolnick, Deborah A. (Weiss), Beth A. (Schultz) Shook, Lyle Campbell and Ives Goddard. 2004. Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation. *American Journal of Human Genetics* 75:519–523.
- Braimoh, A.K. and T. Onishi. 2007. Spatial determinants of urban land use change in Lagos, Nigeria. *Land Use Policy* 24:502–515.
- Bright, Jane O. and William Bright. 1965. Semantic structures in Northwestern California and the Sapir-Whorf hypothesis. *American Anthropologist* 67:249–258.
- Bright, William. 1957. *The Karok language*. Berkeley: University of California Press.
- Bright, William. 1968. *A Luiseño dictionary*. Berkeley: University of California Press.
- Bright, William and Susan Gehr. 2008. Karuk dictionary (online). <http://corpus.linguistics.berkeley.edu/karuk/karuk-lexicon.html>.
- Broadbent, Sylvia M. 1955-1958. Miwok field notes. Survey of California and Other Indian Languages, University of California, Berkeley.
- Broadbent, Sylvia M. 1960. A grammar of Southern Sierra Miwok. Ph.D. Dissertation, University of California Berkeley. Ph.D.

- Broadbent, Sylvia M. 1964. *The Southern Sierra Miwok language*. Berkeley: University of California Press.
- Broadbent, Sylvia M. and Harvey Pitkin. 1964. A comparison of Miwok and Wintun. In *Studies in California linguistics*, ed. William Bright. Berkeley.
- Buckley, Eugene. 1994. *Theoretical aspects of Kashaya phonology and morphology*. Dissertations in Linguistics. Stanford, California: Center for the Study of Language and Information.
- Callaghan, Catherine. 1958-1960. Miwok field notes. Survey of California and Other Indian Languages, University of California, Berkeley.
- Callaghan, Catherine. 1984. *Plains Miwok dictionary*. Berkeley: University of California Press.
- Callaghan, Catherine A. 1963. A grammar of the Lake Miwok language. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Callaghan, Catherine A. 1965. *Lake Miwok dictionary*. Berkeley: University of California Press.
- Callaghan, Catherine A. 1970. *Bodega Miwok dictionary*, volume 60 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Callaghan, Catherine A. 1979a. Miwok color terms. *International Journal of American Linguistics* 45:1–4.
- Callaghan, Catherine A. 1979b. The reconstruction of “two” and related words in Proto-Miwok. *International Journal of American Linguistics* 45:176–181.
- Callaghan, Catherine A. 1985. Proto-Miwok tense and aspect. *International Journal of American Linguistics* 51:356–358.
- Callaghan, Catherine A. 1987. *Northern Sierra Miwok dictionary*. Berkeley: University of California Press.
- Callaghan, Catherine A. 1994. Proto-Miwok numerals. *International Journal of American Linguistics* 60:161–176.
- Callaghan, Catherine A. 1997. Evidence for Yok-Utian. *International Journal of American Linguistics* 63:18–64.
- Callaghan, Catherine A. 2001. More evidence for Yok-Utian: A reanalysis of the Dixon and Kroeber sets. *International Journal of American Linguistics* 67:313–346.
- Callaghan, Catherine A. 2003. Proto-Utian (Miwok-Costanoan) case system. *International Journal of American Linguistics* 69:49–75.

- Callaghan, Catherine A. and Sylvia M. Broadbent. 1960. Comparative Miwok: A preliminary survey. *International Journal of American Linguistics* 26:301–316.
- Campbell, Lyle. 1997. *American Indian languages: The historical linguistics of Native America*. Oxford: Oxford University Press.
- Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In *Linguistic areas: Convergence in historical and typological perspective*, eds. Yaron Matras, April McMahon, and Nigel Vincent, 1–31. Basingstoke, U.K.: Palgrave MacMillan.
- Caruso, G.C., M. Rounsevell and G. Cojocaru. 2005. Exploring a spatio-dynamic neighbourhood-based model of residential behaviour in the brussels periurban area. *International Journal of Geographical Information Science* 19:103–123.
- Chambers, J.K. and Peter Trudgill. 1980. *Dialectology*. Cambridge: Cambridge University Press.
- Chartkoff, Joseph L. 1990. Cracking and grinding, chipping and swapping: Summers on Skunk Creek. *Proceedings of the Society for California Archaeology* 3:21–34.
- Chitoran, Ioana, Louis Goldstein and Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In *Laboratory phonetics*, eds. Carlos Gussenhoven and Natasha Warner. Berlin, New York.
- Conathan, Lisa. 2004. The linguistic ecology of northwestern California: Contact, functional convergence, and dialectology. Ph.D. Dissertation, University of California Berkeley. Ph.D.
- Conathan, Lisa and Esther Wood. 2003. Repetitive reduplication in Yurok and Karuk: Semantic effects of contact. In *Papers of the thirty-fourth algonquian conference*, ed. H. C. Wolfart, 19–33. Winnipeg: University of Manitoba.
- Cook, Sherburne F. 1955. The aboriginal population of the San Joaquin Valley, California. *University of California Anthropological Records* 16:31–80.
- Couro, Ted and Christina Hutcheson. 1973. *Dictionary of Mesa Grande Diegueño*. Banning, California: Malki Museum Press.
- Crawford, James M. 1989. *Cocopa dictionary*. Berkeley: University of California Press.
- Cross, Coca. 1982. Miwok field notes. Survey of California and Other Indian Languages, University of California, Berkeley.
- Daley, Jon P. 1989. *Tümpisa (Panamint) Shoshone dictionary*, volume 116 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Davis, Alva L. and Jr. Raven I. McDavid. 1950. Northwestern Ohio: A transition area. *Language* 26:264–273.

- Davis, James T. 1974. *Trade routes and economic exchange among the Indians of California*. Ramona, California: Ballena Press.
- Dayley, Jon P. 1989. *Tümpisa (Panamint) Shoshone grammar*, volume 115 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Delancey, Scott. 1996. Penutian in the bipartite stem belt: Disentangling areal and genetic correspondences. In *Proceedings of the 22nd annual meeting, special session*, eds. David Librik and Roxanne Beeler, volume 22s of *BLS*, 37–54. Berkeley Linguistics Society.
- Delancey, Scott. 1999. Lexical prefixes and the bipartite stem construction in Klamath. *International Journal of American Linguistics* 65:56–83.
- Dixon, Roland B. 1905. The Shasta-Achomawi: A new linguistic stock, with four new dialects. *American Anthropologist* 7:213–217.
- Dixon, Roland B. 1906. The pronominal dual in the languages of California. In *Boas anniversary volume*, 80–84. New York: GE Stechert & Company.
- Dixon, Roland B. 1910. *The Chimariko Indians and language*, volume 5.5 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Dixon, Roland B. and A. L. Kroeber. 1913. New linguistic families in California. *American Anthropologist* 15:647–655.
- Dixon, Roland B. and A. L. Kroeber. 1919. *Linguistic families of California*, volume 16 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Dixon, Roland B. and Alfred L. Kroeber. 1903. The native languages of California. *American Anthropologist* 5:1–26.
- Dixon, Roland B. and Alfred L. Kroeber. 1907. Numeral systems of the languages of California. *American Anthropologist* 9.4:663–690.
- Donohue, Mark and Simon Musgrave. 2007. Typology and the linguistic macrohistory of Island Melanesia. *Oceanic Linguistics* 46:348–387.
- Donohue, Mark and Bronwen Whiting. 2011. Quantifying areality: A study of prenasalization in Southeast Asia and New Guinea. *Typology* 15:101–122.
- Donohue, Mark, Søren Wichmann and Mihai Albu. 2008. Typology, areality, and diffusion. *Oceanic Linguistics* 47:223–232.
- Driver, Harold. 1961. *Indians of North America*. Chicago: University of Chicago Press.
- Dunn, Michael. 2009. Contact and phylogeny in Island Melanesia. *International Journal of American Linguistics* 119:1664–1678.

- Dunn, Michael, Robert A. Foley, Stephen C. Levinson, Ger Reesink and Angela Terrill. 2007. Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Oceanic Linguistics* 46:388–403.
- Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley and Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309:2072–2075.
- Eatough, Andrew. 1999. *Central Hill Nissenan texts with grammatical sketch*, volume 132 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Edel, May. 1939. The Tillamook language. *International Journal of American Linguistics* 10:1–57.
- Emeneau, Murray. 1956. India as a linguistic area. *Language* 32:3–16.
- Felsenstein, J. 2001. PHYLIP: Phylogeny inference package: Version 3.6. Department of Genetics, University of Washington.
- Frachtenberg, Leo J. 1914. Lower Umpqua texts and notes on the Kusan dialect. *California University Contributions to Anthropology* 4:141–150.
- Frachtenberg, Leo J. 1918. Comparative studies in Takelman, Kalapuyan and Chinookan lexicography, a preliminary paper. *International Journal of American Linguistics* 1:175–182.
- Frachtenberg, Leo J. 1922a. Coos. In *Handbook of American Indian Languages, part 2*, ed. Franz Boas, volume 40.2 of *Bureau of American Ethnology Bulletin*, 297–429. Washington D.C.: Smithsonian Institution.
- Frachtenberg, Leo J. 1922b. Siuslawan. In *Handbook of American Indian Languages, part 2*, ed. Franz Boas, volume 40.2 of *Bureau of American Ethnology Bulletin*, 431–629. Washington D.C.: Smithsonian Institution.
- Freeland, Lucy S. 1948-1972. Lucy Shepard Freeland papers, 1948-1972. University of California, Santa Cruz. University Library. Special Collections and Archives.
- Freeland, Lucy S. 1951. *Language of the Sierra Miwok*. Bloomington, Indiana: Indiana University Press.
- Freeland, Lucy S. n.d. Central Sierra Miwok field notes. Survey of California and Other Indian Languages, University of California, Berkeley.
- Freeland, Lucy S. and Sylvia M. Broadbent. 1960. *Central Sierra Miwok dictionary with texts*. Berkeley: University of California Press.
- Gatschet, Albert Samuel. 1890. *The Klamath Indians of Southwestern Oregon, Parts 1 and 2*. Washington D.C.: Government Printing Office.

- Gesh, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck and D. Tyler. 2002. The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing* 68:5–11.
- Gesh, D. B. 2007. The National Elevation Dataset. In *Digital elevation model technologies and applications: The DEM users manual, 2nd edition*, ed. D. Maune, 99–118. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing.
- Gifford, Edward Winslow. 1916. *Miwok moieties*, volume 12 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Gifford, Edward Winslow. 1926. Miwok lineages and the political unit in aboriginal California. *American Anthropologist* 28:389–401.
- Giovani, B. and R. F. Goldman. 1971. Predicting metabolic energy cost. *Journal of Applied Physiology* 30:429–433.
- Goddard, Ives. 1996. The classification of the native languages of North America. In *Handbook of North American Indians Vol. 17*, ed. W. C. Sturtevant, 290–324. Washington, D.C.: Smithsonian Institution.
- Goddard, Pliny Earle. 1912. *Elements of the Kato language*, volume 11.1 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Goddard, Pliny Earle. 1929. *The Bear River dialect of Athapaskan*, volume 24.5 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Golla, Victor. 1970. Hupa grammar. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Golla, Victor. 1976. Tututni (Oregon Athapaskan). *International Journal of American Linguistics* 42:217–227.
- Golla, Victor. 2011. *California Indian languages*. Berkeley: University of California Press.
- Gooskens, Charlotte. 2004. Norwegian dialect differences geographically explained. In *Papers from the Second International Conference on Language Variation in Europe*, eds. B. L. Gunnarson, L. Bergstrom, G. Eklund, S. Fridella, L. H. Hansen, A. Karstadt, B. Nordberg, E. Sundgren, and M. Thelander, 195–206. Uppsala, Sweden: Uppsala University.
- Gow, David W., Janis Melvold and Sharon Manuel. 1996. How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing. In *Proceedings of the International Conference on Spoken Language Processing*.
- Greenberg, Joseph. 1987. *Language in the Americas*. Palo Alto: Stanford University Press.

- Gursky, Karl-Heinz. 1974. Der Hoka-Sprachstamm. Eine bestandsaufnahme des lexikalischen beweismaterials. *Orbis* 23:170–215.
- Haas, Mary R. 1963. Shasta and Proto-Hokan. *Language* 39:40–59.
- Haas, Mary R. 1967. Language and taxonomy in Northwestern California. *American Anthropologist* 69:358–362.
- Haas, Mary R. 1970. Consonant symbolism in Northwestern California: A problem in diffusion. In *Languages and cultures of western North America: Essays in honor of Sven S. Liljeblad*, ed. Jr. Earl H. Swanson, 86–96. Pocatello: Idaho State University Press.
- Haas, Mary R. 1976. The Northern California linguistic area. In *Hokan studies*, eds. Margaret Langdon and Shirley Silver. The Hague.
- Halpern, Abraham M. 1964. A report on a survey of Pomo languages. In *Studies in Californian linguistics*, ed. William Bright, volume 34, University of California Publications in Linguistics, 88–93. Berkeley: University of California Press.
- Hare, Timothy S. 2004. Using measures of cost distance in the estimation of polity boundaries in the postclassic Yautepec Valley, Mexico. *Journal of Archaeological Science* 31:799–814.
- Haynie, Hannah J. and Loretta Kelsey. 2007. Southeastern Pomo field notes.
- Heeringa, Wilbert. 2004. Measuring dialect pronunciation differences using Levenshtein distance. Ph.D. Dissertation, Rijksuniversiteit Groningen. Ph.D.
- Heeringa, Wilbert and John Nerbonne. 2005. Dialect areas and dialect continua. ms.
- Heizer, R. F. 1952. California Indian linguistic records: the Mission Indian vocabularies of Alphonse Pinart. *University of California Anthropological Records* 15:1–84.
- Heizer, R. F. 1955. California Indian linguistic records: the Mission Indian vocabularies of H. W. Henshaw. *University of California Anthropological Records* 15:85–202.
- Hill, Jane H. 2001. Proto Uto-Aztecan: A community of cultivators in Central Mexico? *American Anthropologist* 103:317–358.
- Hill, Kenneth C., ed. 1998. *Hopi dictionary/Hopìikwa lavàytutuveni: A Hopi-English dictionary of the Third Mesa dialect, with an English-Hopi finder list and a sketch of Hopi grammar*. Tucson: University of Arizona Press.
- Hoiijer, Harry. 1966. Galice Athapaskan: A grammatical sketch. *International Journal of American Linguistics* 32:320–327.
- Hymes, Dell. 1966. Some points of Siuslaw phonology. *International Journal of American Linguistics* 32:328–342.
- Jacobsen, William H. 1958. Washo and Karok: An approach to comparative Hokan. *International Journal of American Linguistics* 24:195–212.

- Janin, A., J.P. Léna, N. Ray, C. Delacourt, P. Allemand and P. Joly. 2009. Assessing landscape connectivity with calibrated cost-distance modelling: predicting common toad distribution in a context of spreading agriculture. *Journal of Applied Ecology* 46:833–841.
- Jansen, Joana Worth. 2010. A grammar of Yakima Ichishkíin/Sahaptin. Ph.D. Dissertation, University of Oregon. Ph.D.
- Jany, Carmen. 2009. *Chimariko grammar: Areal and typological perspective*, volume 142 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Johnson, Keith. 2008. *Quantitative methods in linguistics*. Oxford: Blackwell.
- Kaufman, Terrence. 1988. A research program for reconstructing Proto-Hokan: First groupings. In *Papers from the 1988 Hokan-Penutian languages workshop*, ed. Scott DeLancy, 50–168. Eugene, Oregon: Department of Linguistics, University of Oregon.
- Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. In *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, 60–66. San Francisco: Morgan Kaufmann Publishers Inc.
- Kessler, Brett. 1999. Estimating the probability of historical connections between languages. Ph.D. Dissertation, Stanford University. Ph.D.
- Kessler, Brett. 2001. *The significance of word lists*. Stanford: CSLI Publications.
- Kessler, Brett and Annukka Lehtonen. 2006. Multilateral comparison and significance testing of the Indo-Uralic question. In *Phylogenetic methods and the prehistory of languages*, eds. Peter Forster and Colin Renfrew, 33–42. Cambridge: McDonald Institute for Archaeological Research.
- Kleiweg, Peter. 2012. RuG/L04 software for dialectometrics and cartography. URL <http://www.let.rug.nl/kleiweg/L04/>.
- Kroeber, Alfred L. 1869-1972. A. L. kroeber papers. BANC FILM 2049, The Bancroft Library, University of California, Berkeley.
- Kroeber, Alfred L. 1906. The dialectic divisions of the Moquelumnan family in relation to the internal differentiation of the other linguistic families of California. *American Anthropologist* 8:652–663.
- Kroeber, Alfred L. 1907a. *Phonetic constituents of the native languages of California*, volume 2 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Kroeber, Alfred L. 1907b. *The Washo language of East Central California and Nevada*, volume 4.5 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.

- Kroeber, Alfred L. 1907c. *The Yokuts language of south central California*, volume 2: 165–377 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Kroeber, Alfred L. 1908. *On the evidences of the occupation of certain regions by the Miwok Indians*, volume 6 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Kroeber, Alfred L. 1909. Classificatory systems of relationship. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland* 39:77–84.
- Kroeber, Alfred L. 1910. *The Chumash and Costanoan languages*, volume 9.2 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Kroeber, Alfred L. 1925. *Handbook of the Indians of California*. Bureau of American Ethnology Bulletin. Washington, D.C.: Smithsonian Institution.
- Kroeber, Alfred L. 1929. *The Valley Nisenan*, volume 24.4 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Kroeber, Alfred L. 1963. *Yokuts dialect survey*, volume 11: 177–251 of *Anthropological Records*. Berkeley: University of California Press.
- Labov, William. 2007. Transmission and diffusion. *Language* 83:344–387.
- Lamb, Sydney. 1958. Monachi dictionary manuscript. Berkeley.
- Lamb, Sydney McDonald. 1957. Mono grammar. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Lees, Robert B. 1953. The basis of glottochronology. *Language* 29:113–127.
- van Leusen, Pieter Martijn. 2002. Pattern to process: Methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes. Ph.D. Dissertation, Rijksuniversiteit Groningen. Ph.D.
- Levy, Richard. 1978. Eastern Miwok. In *Handbook of North American Indians Vol. 8: California*, ed. Robert F. Heizer, 398–415. Washington, D.C.: Smithsonian Institution.
- Li, Fanggui. 1930. *Mattole: An Athabaskan language*. University of Chicago Publications in Anthropology: Linguistic Series. Chicago: University of Chicago Press.
- Livingood, Patrick. 2009. Down the river and through the woods: Cost-distance calculations of travel time in the Mississippian. Paper for the 74th Annual Meeting of the Society for American Archaeology, Atlanta, Georgia.
- Maniery, James Gary. 1983. A chronicle of Murphys Rancheria (*Mol-Pee-So*): An historic Central Sierra Miwok village. *Journal of California and Great Basin Anthropology* 5:176–198.

- Maniery, James Gary. 1987. *Six Mile and Murphys rancherias: An ethnohistorical and archaeological study of two Central Sierra Miwok village sites*. San Diego: San Diego Museum of Man.
- Mason, J. Alden. 1916. *The Mutsun dialect of Costanoan based on the vocabulary of de la Cuesta*, volume 11.7 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Mason, J. Alden. 1918. *The language of the Salinan Indians*, volume 14.1 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- McLendon, Sally. 1964. Northern Hoka (B) and (C): A comparison of Eastern Pomo and Yana. In *Studies in Californian linguistics*, ed. William Bright, volume 34 of *University of California Publications in Linguistics*, 126–144. Berkeley: University of California Press.
- McMahon, April and Robert McMahon. 2005. *Language classification by numbers*. Oxford: Oxford University Press.
- Merriam, C. Hart. 1850-1974. Papers relating to work with California Indians, 1850-1974 (bulk 1898-1938). BANC FILM 1022, The Bancroft Library, University of California, Berkeley.
- Merriam, C. Hart. 1904. Distribution of Indian tribes in the southern Sierra and adjacent parts of the San Joaquin Valley, California. *Science* 19:912–917.
- Merriam, C. Hart. 1907. Distribution and classification of the Mewan stock of California. *American Anthropologist* 9:338–357.
- Mickelson, Andrew. 2003. A GIS-based optimal foraging model of human walking and load transport. ms.
- Milroy, James and Lesley Milroy. 1985. Linguistic change, social networks, and speaker innovation. *Journal of Linguistics* 21:339–384.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Montague, Suzanna T. 2010. High-elevation prehistoric land use in the Central Sierra Nevada, Yosemite National Park, California. Master's thesis, California State University, Sacramento.
- Moshinsky, Julius. 1965. Julius Moshinsky papers. Survey of California and Other Indian Languages, University of California, Berkeley.
- Moshinsky, Julius. 1974. *A grammar of Southeastern Pomo*, volume 72 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Nerbonne, John. 2008. Measuring the diffusion of linguistic change. ms.

- Nerbonne, John. 2009. Data-driven dialectology. *Language and Linguistic Compass* 3:175–198.
- Nerbonne, John and Wilbert Heeringa. 2007. Geographic distributions of linguistic variation reflect dynamics of differentiation. In *Roots: linguistics in search of its evidential base*, ed. Sam Featherston, 267–318. Berlin: Mouton de Gruyter.
- Nerbonne, John and Peter Kleiweg. 2006. Toward a dialectological yardstick. *Literary and Linguistic Computing* 21:477–492.
- Newman, Stanley S. 1944. *Yokuts language of California*. New York: Viking.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 1993. Ergativity and linguistic geography. *Australian Journal of Linguistics* 13:39–89.
- Nichols, Johanna. 1999. The historical geography of pharyngeals and laterals in the Caucasus. In *Proceedings of the twenty-fifth annual meeting of the Berkeley Linguistics Society: Special session on Caucasian, Dravidian, and Turkic languages*, eds. Jeff Good and Alan Yu, volume 25, 1–13. Berkeley: Berkeley Linguistics Society.
- Nichols, Johanna and David A. Peterson. 1996. The Amerind personal pronouns. *Language* 72:336–371.
- Nolan, Kevin C. and Robert A. Cook. 2012. A method for multiple cost surface evaluation of a model of Fort Ancient interaction. In *Least cost analysis of social landscapes: Archaeological case studies*. Salt Lake: University of Utah Press.
- Okrand, Marc. 1977. Mutsun grammar. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Olmsted, D. L. 1956. Palaihnihan and Shasta I: Labial stops. *Language* 32:73–77.
- Olmsted, D. L. 1957. Palaihnihan and Shasta II: Apical stops. *Language* 33:136–138.
- Olmsted, D. L. 1959. Palaihnihan and Shasta III: Dorsal stops. *Language* 35:637–644.
- Olmsted, D. L. 1961. Atsugewi morphology I: Verb inflection. *International Journal of American Linguistics* 27:91–113.
- Olmsted, D. L. 1966. *Achumawi dictionary*. Berkeley: University of California Press.
- Olmsted, D. L. 1984. *A lexicon of Atsugewi*. Berkeley: Survey of California and Other Indian Languages.
- Olmsted, D. L. and William Bright. 1959. A Shasta vocabulary. *The Kroeber Anthropological Society Papers* 20:1–55.

- O'Neill, Sean. 2008. *Cultural contact and linguistic relativity among the Indians of North-western California*. Norman, Oklahoma: University of Oklahoma Press.
- Oswalt, Robert. 1957. *Kashaya vocabulary*, volume 32 of *Kashaya Pomo Language in Culture Project Working Papers*. Department of Anthropology, California State College, Sonoma.
- Oswalt, Robert L. 1961. A Kashaya grammar (Southwestern Pomo). Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Oswalt, Robert L. 1964. The internal relationships of the Pomo family of languages. In *Actas y memorias*, volume II, 413–427. Mexico: XXXV Congreso Internacional de Americanistas.
- Oswalt, Robert L. 1970. The detection of remote linguistic relationships. *Computer Studies* 3:119–129.
- Oswalt, Robert L. n.d., 1967. Robert Louis Oswalt papers on Pomoan languages. Survey of California and Other Indian Languages, University of California, Berkeley.
- Pharris, Nicholas J. 2006. Winuunsi tm talapaas: A grammar of the Molalla language. Ph.D. Dissertation, University of Michigan. Ph.D.
- Pierce, Joe E. 1966. Genetic comparisons and Hanis, Miluk, Alsea, Siuslaw, and Takelma. *International Journal of American Linguistics* 32:379–387.
- Pingel, Thomas J. 2010. Modeling slope as a contributor to route selection in mountainous areas. *Cartography and Geographic Information Science* 37:137–148.
- Pitkin, Harvey. 1984. *Wintu grammar*, volume 94 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Pitkin, Harvey. 1985. *Wintu dictionary*. Berkeley: University of California Press.
- Pitkin, Harvey and William Shipley. 1958. A comparative survey of California Penutian. *International Journal of American Linguistics* 24:174–188.
- Powell, John Wesley. 1891. *Indian linguistic families of America, north of Mexico*. Seventh Annual Report of the Bureau of Ethnology to the Secretary of the Smithsonian Institution, 1885-1886. Washington, D.C.: Government Printing Office.
- Powers, Stephen. 1877. Tribes of California. In *Contributions to North American ethnology, volume 3, Department of the Interior, U. S. Geographical and Geological Survey of the Rocky Mountain Region, J. W. Powell in charge*. Washington: Government Printing Office.
- Radin, Paul. 1929. *A grammar of the Wappo language*, volume 27.1 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Ray, N., A. Lehmann and P. Joly. 2002. Modeling spatial distribution of amphibian populations: a GIS approach based on habitat matrix permeability. *Biodiversity and Conservation* 11:2143–2165.

- Rees, W. G. 2003. Least-cost paths in mountainous terrain. *Computers and Geosciences* 30:203–209.
- Reichard, Gladys A. 1925. *Wiyot grammar and texts*, volume 22.1 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Remote Sensing Lab, USDA Forest Service Pacific Southwest Region. 2009. Vegetation descriptions, South Sierran Ecological Province, CALVEG Zone 4. Web resource, URL <http://prdp2fs.ess.usda.gov/Internet/FSE.DOCUMENTS/fsbdev3.046270.pdf>.
- Remote Sensing Lab, USDA Forest Service Pacific Southwest Region. 2010. CALVEG VegetationTiles30. Vector digital data, URL <http://www.fs.fed.us/r5/rsl/projects/mapping/>.
- Rigsby, Bruce. 1966. On Cayuse-Molala relatibility. *International Journal of American Linguistics* 32:369–378.
- Rigsby, Bruce. 1969. The Waiilatpuan problem: More on Cayuse-Molala relatibility. *North-west Anthropological Research Notes* 3.1:68–146.
- Ringe, Donald, Tandy Warnow and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100:139–151.
- Robins, R. H. 1958. *The Yurok language: Grammar, texts, lexicon*, volume 15 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Ross, Malcolm. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In *The comparative method reviewed: Regularity and irregularity in language change*, eds. Mark Durie and Malcolm Ross, 180–217. Oxford: Oxford University Press.
- Ross, Malcolm. 2003. Diagnosing prehistoric language contact. In *Motives for language change*, ed. Raymond Hickey, 174–198. Cambridge: Cambridge University Press.
- Sandfeld, Kristian. 1930. *Linguistique balkanique: Problèmes et résultats*. Paris: Société Linguistique de Paris.
- Sankoff, David. 1972. Reconstructing the history and geography of an evolutionary tree. *American Mathematics Monthly* 79:596–603.
- Sapir, Edward. 1907. Preliminary report on the language and mythology of the Upper Chinook. *American Anthropologist* 9:533–544.
- Sapir, Edward. 1917. *The position of Yana in the Hokan stock*, volume 13 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: The University of California Press.
- Sapir, Edward. 1921a. A bird’s-eye view of American languages north of Mexico. *Science* 54:408.

- Sapir, Edward. 1921b. A characteristic Penutian form of stem. *International Journal of American Linguistics* 2:261–273.
- Sapir, Edward. 1922a. *The fundamental elements of Northern Yana*, volume 13.6 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Sapir, Edward. 1922b. The Takelma language of Southwestern Oregon. In *Handbook of American Indian Languages, part 2*, ed. Franz Boas, volume 40.2 of *Bureau of American Ethnology Bulletin*, 1–296. Washington, D.C.: Smithsonian Institution.
- Sapir, Edward and Morris Swadesh. 1960. *Yana dictionary*, volume 22 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Sawyer, Jesse. n.d. Southern Sierra Miwok field notes. Survey of California and Other Indian Languages, University of California, Berkeley.
- Schlichter, Alice. 1985. The Yukian language family. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Seaburg, William R. 1977. A Wailaki (Athapaskan) text with comparative notes. *International Journal of American Linguistics* 43:327–332.
- Séguy, Jean. 1973. La dialectometrie dans l’atlas linguistique de Gascogne. *Revue de Linguistique Roman* 37:1–24.
- Shaul, David. 1995. The Huelel (Esselen) language. *International Journal of American Linguistics* 61:191–239.
- Sherzer, Joel. 1976. *An areal-typological study of American Indian languages north of Mexico*. Amsterdam: North Holland Publishing Co.
- Shimkin, D. B. 1949. Shoshone, I: Linguistic sketch and text. *International Journal of American Linguistics* 15:175–188.
- Shipley. 1965. *Maidu grammar*, volume 41 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Shipley, William F. 1963. *Maidu texts and dictionary*. Berkeley: University of California Press.
- Shipley, William F. 1980. Penutian among the ruins: A personal assessment. In *Proceedings of the sixth annual meeting of the Berkeley Linguistics Society*, 437–441.
- Silver, Shirley. 1961. Shirley Silver papers. Survey of California and Other Indian Languages, University of California, Berkeley.
- Silver, Shirley. 1964. Shasta and Karok: A binary comparison. In *Studies in Californian linguistics*, ed. William Bright, volume 34 of *University of California Publications in Linguistics*, 170–181. Berkeley: University of California Press.

- Silver, Shirley. 1966. The Shasta language. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- Singleton, P.H., W.L. Gaines, J.F. Lehmkuhl and Or.) Pacific Northwest Research Station (Portland. 2002. *Landscape permeability for large carnivores in washington: a geographic information system weighted-distance and least-cost corridor assessment*. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- Sjoberg, Andrée and Gideon Sjoberg. 1956. Problems in glottochronology. *American Anthropologist* 58:296–308.
- Snapp, Allen, John Anderson and Joy Anderson. 1982. Northern Paiute. In *Studies in Uto-Aztecan grammar, Volume 3: Uto-Aztecan grammatical sketches*, ed. Ronald W. Langacker, volume 56.3 of *Summer Institute of Linguistics Publications in Linguistics*, 1–92. Dallas: Summer Institute of Linguistics.
- Swadesh, Morris. 1952. Lexico-statistical dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96:453–463.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21:121–137.
- Teeter, Karl V. 1964. *The Wiyot language*, volume 37 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- The Southern Ute Tribe. 1979. *Ute dictionary, preliminary edition*. Ignacio, Colorado: Ute Press.
- Thomason, Sarah Grey. 2001. *Language contact*. Washington, D.C.: Georgetown Press.
- Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thompson, Sandra A., Joseph Sung-Yul Park and Charles N. Li. 2006. *A reference grammar of Wappo*, volume 138 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Thornes, Timothy Jon. 2003. A Northern Paiute grammar: With texts. Ph.D. Dissertation, University of Oregon. Ph.D.
- Tobler, Waldo. 1993. Non-isotropic geographic modeling. In *Three presentations on geographic analysis and modeling*, volume 93-1 of *Center for Geographic Analysis Technical Report*. Santa Barbara: University of California.
- Tozzer, Alfred M. 1900a. Sierra Miwok field notes. Survey of California and Other Indian Languages, University of California, Berkeley.

- Tozzer, Alfred M. 1900b. Sierra Miwok field notes. Ethnological Documents of the Department and Museum of Anthropology, University of California, Berkeley, BANC FILM 2216, The Bancroft Library, University of California, Berkeley.
- Trubetskoy, Nikolai S. 1928. Proposition 16. In *Acts of the first international congress of linguists*, 17–18. Leiden.
- Uldall, Hans J. and William Shipley. 1966. *Nisenan texts and dictionary*. Berkeley: University of California Press.
- Ultan, Russell. 1967. Konkow. Ph.D. Dissertation, University of California, Berkeley. Ph.D.
- U.S. Geological Survey. 1999. National Hydrography Dataset. Vector digital data, URL <http://nhd.usgs.gov/index.html>.
- Valls, Esteve, John Nerbonne, Jelena Prokić, Martijn Wieling, Esteve Clua and Maria-Rosa Lloret. 2010. Applying Levenshtein distance to Catalan dialects. a brief comparison of two dialectometric approaches. ms.
- Vihman, Eero. 1970. Eero Vihman papers. Survey of California and Other Indian Languages, University of California, Berkeley.
- Voegelin, C. F. and Voegelin. 1966. Recent classifications of genetic relationships. map.
- Voegelin, C. F. and F. M. Voegelin. 1973. Map of North American Indian Languages. *Annual Review of Anthropology* 2:139–151.
- Voegelin, Charles F. 1935. *Tübatulabal grammar*, volume 34 of *University of California Publications in American Archaeology and Ethnology*. Berkeley: University of California Press.
- Whistler, Kenneth and Victor Golla. 1986. Proto-Yokuts reconsidered. *International Journal of American Linguistics* 52:913–934.
- Wood, Brian M. and Zoe J. Wood. 2006. Energetically optimal travel across terrain: visualizations and a new metric of geographic distance with archaeological applications. In *SPIE Proceedings: San Jose, CA*, volume 60.
- Zigmond, Maurice L., Curtis G. Booth and Pamela Munro. 1990. *Kawaiisu: A grammar and dictionary, with texts*. Berkeley: University of California Press.

Appendix A

Lexical Isogloss Maps of Eastern Miwok

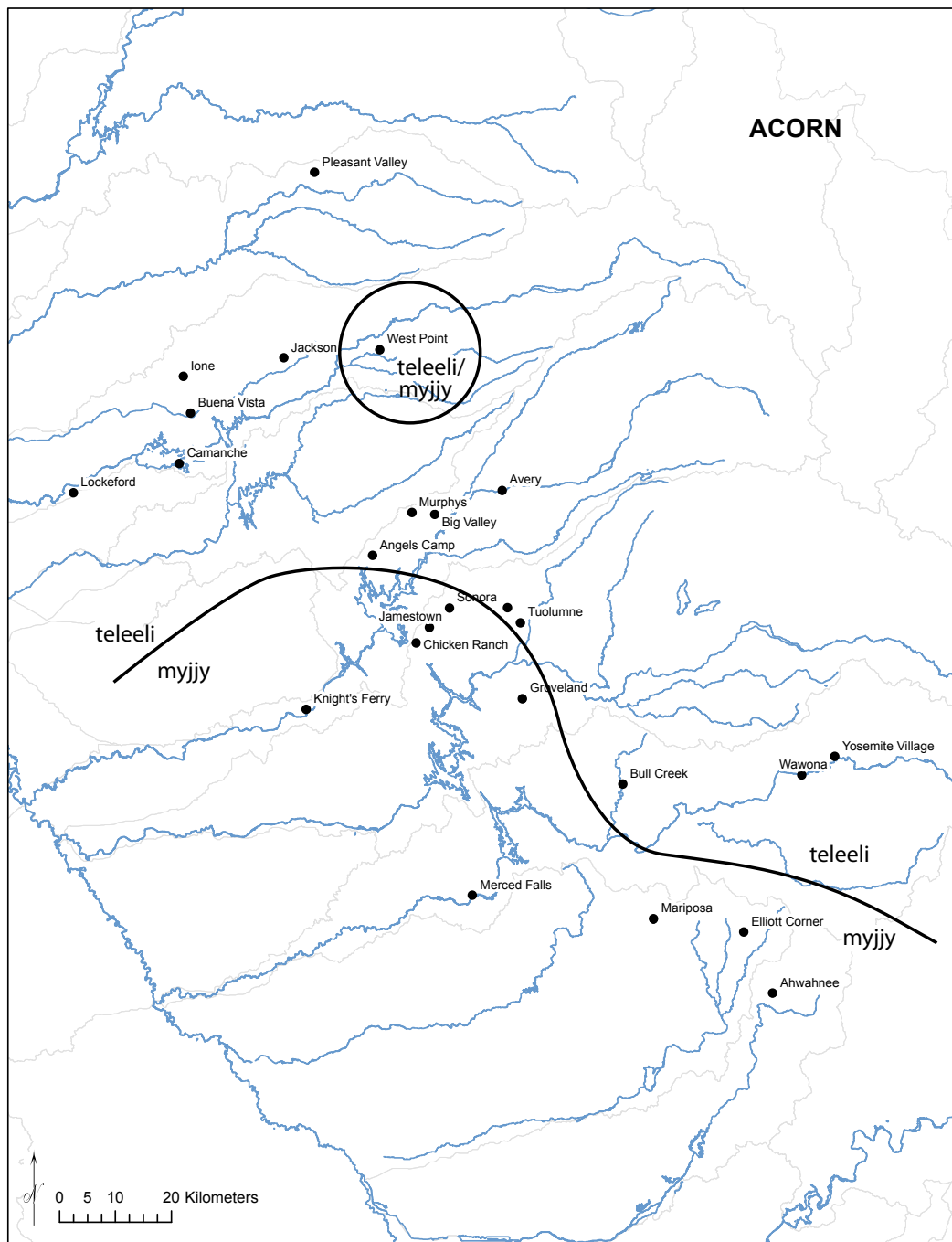


Figure 5.1: Eastern Miwok isogloss map: Acorn

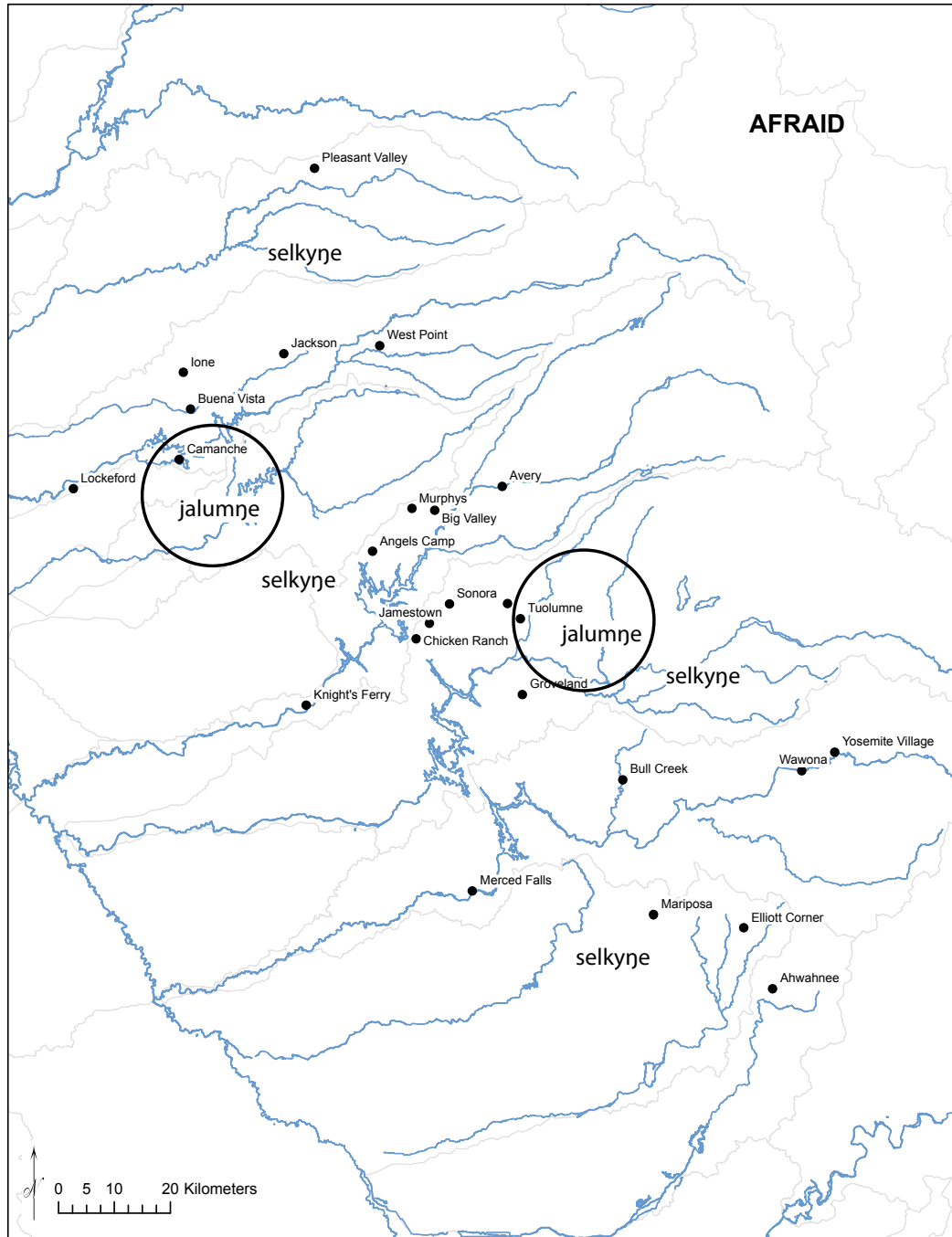


Figure 5.2: Eastern Miwok isogloss map: Afraid

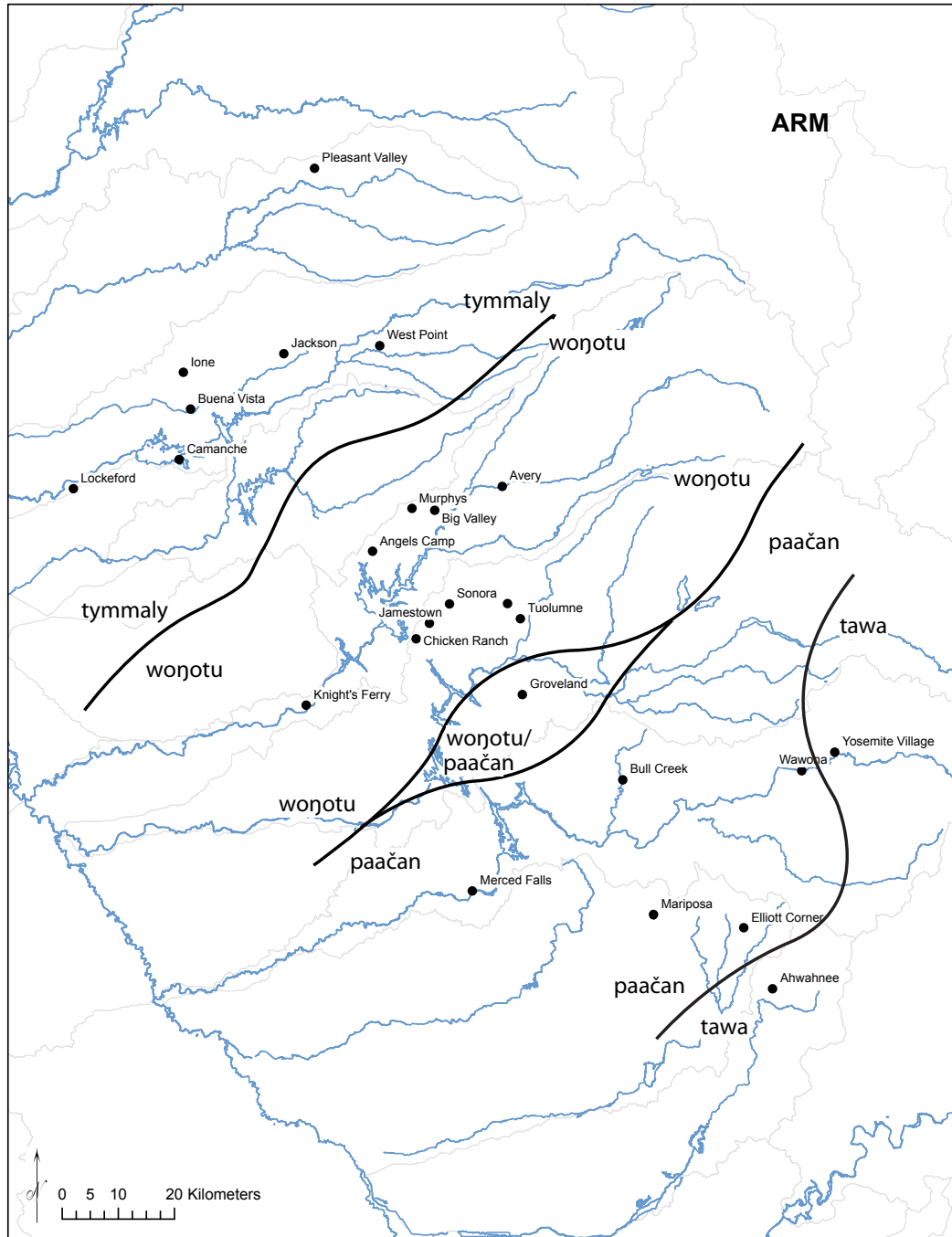


Figure 5.3: Eastern Miwok isogloss map: Arm

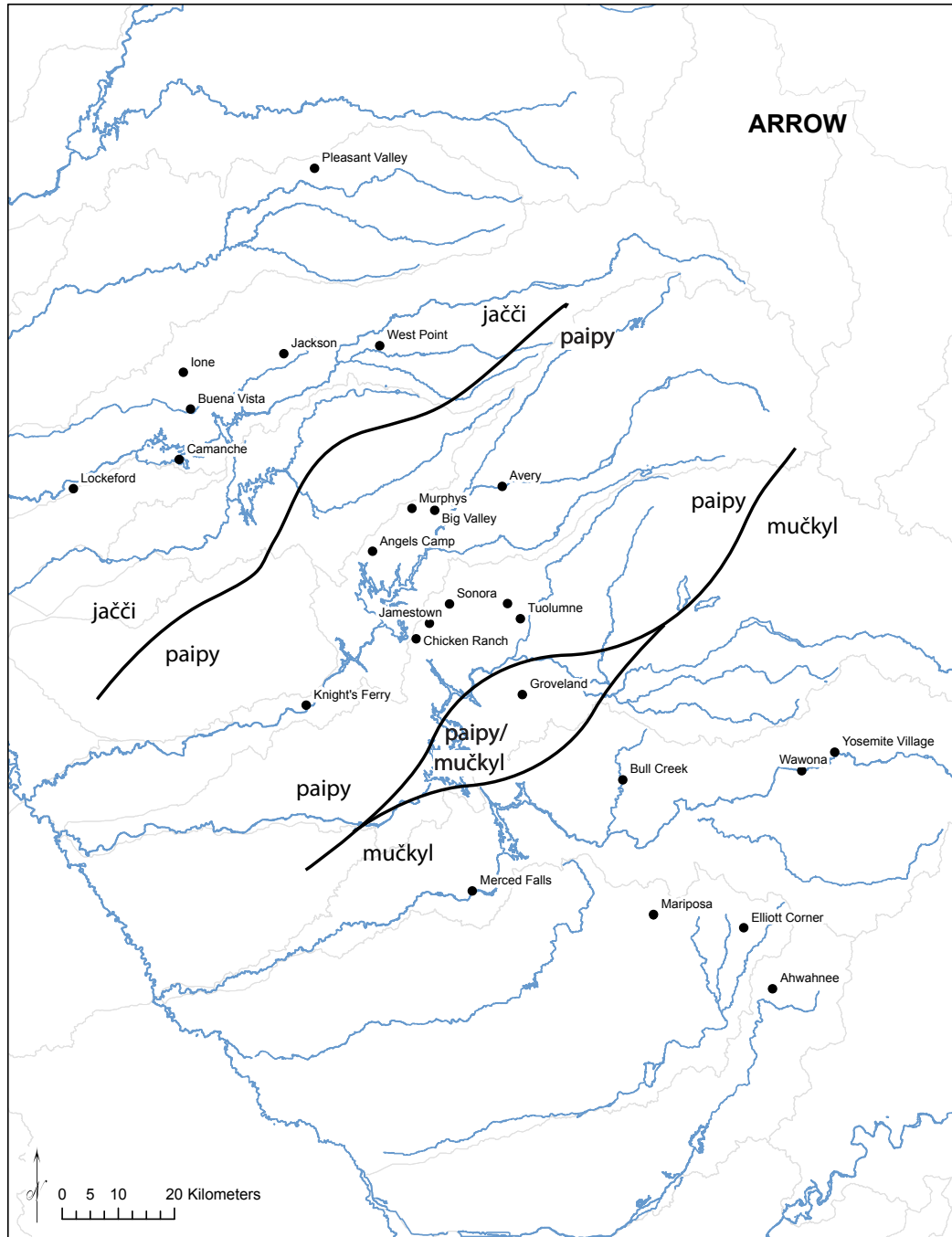


Figure 5.4: Eastern Miwok isogloss map: Arrow

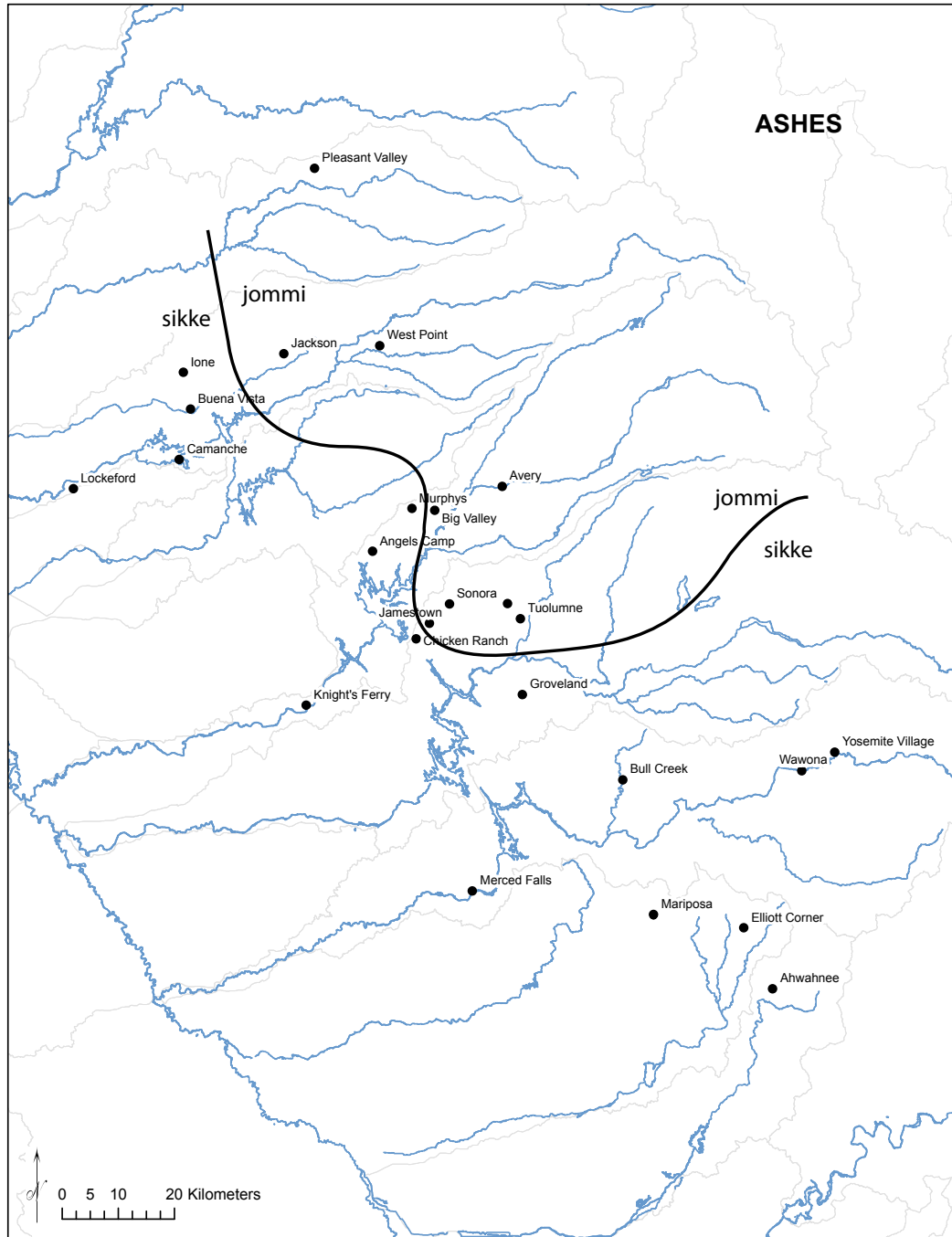


Figure 5.5: Eastern Miwok isogloss map: Ashes

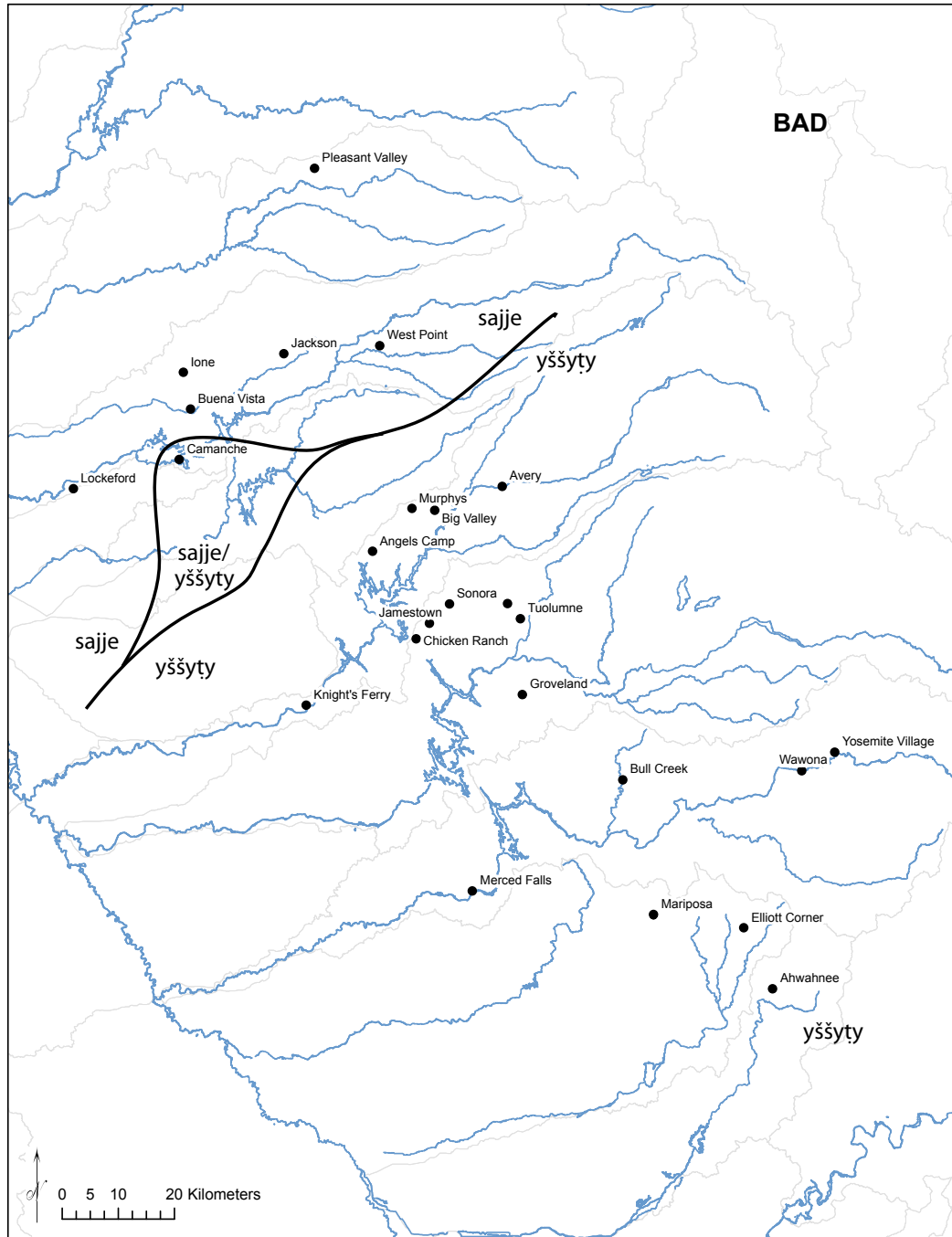


Figure 5.6: Eastern Miwok isogloss map: Bad

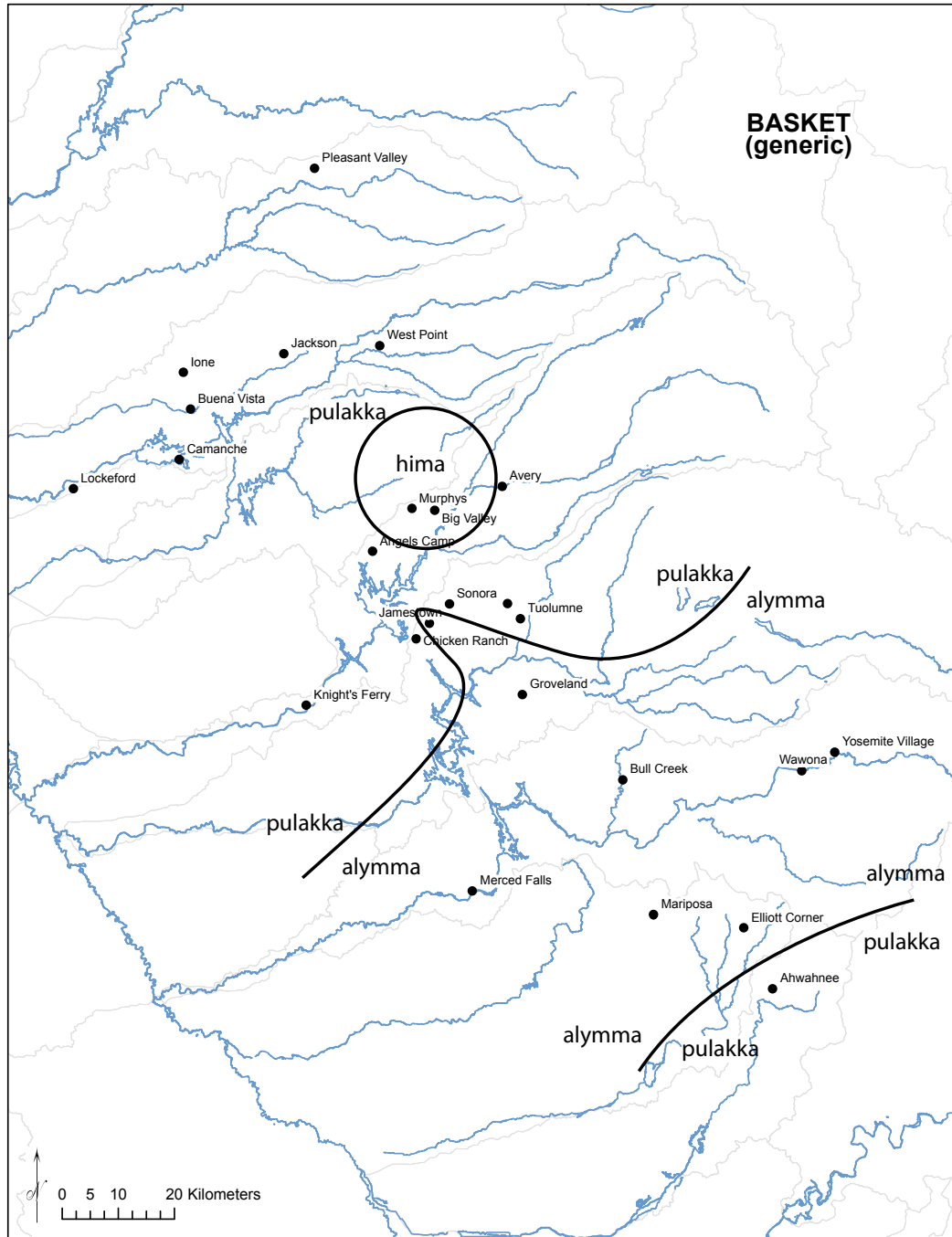


Figure 5.7: Eastern Miwok isogloss map: Basket (generic)

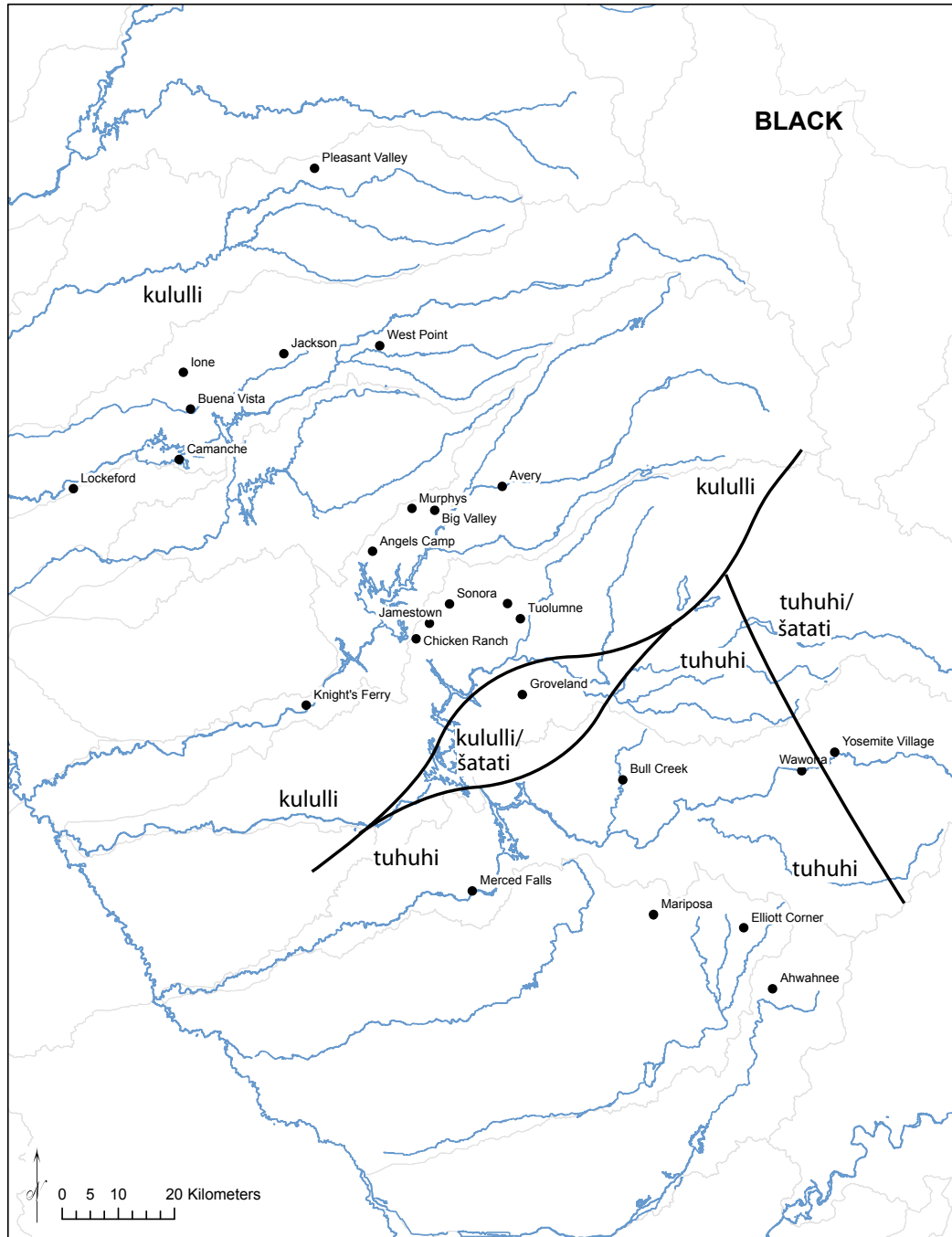


Figure 5.8: Eastern Miwok isogloss map: Black

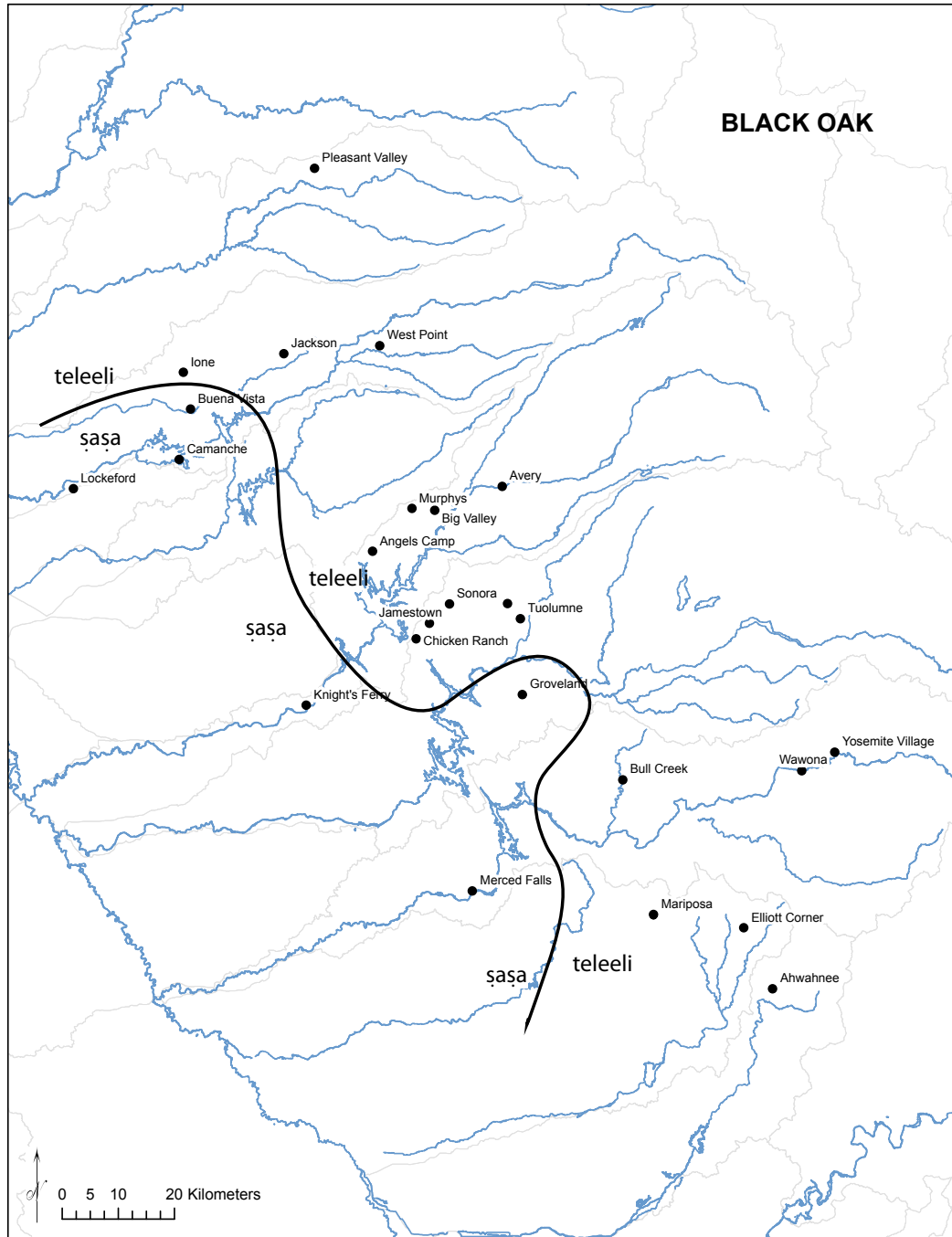


Figure 5.9: Eastern Miwok isogloss map: Black Oak

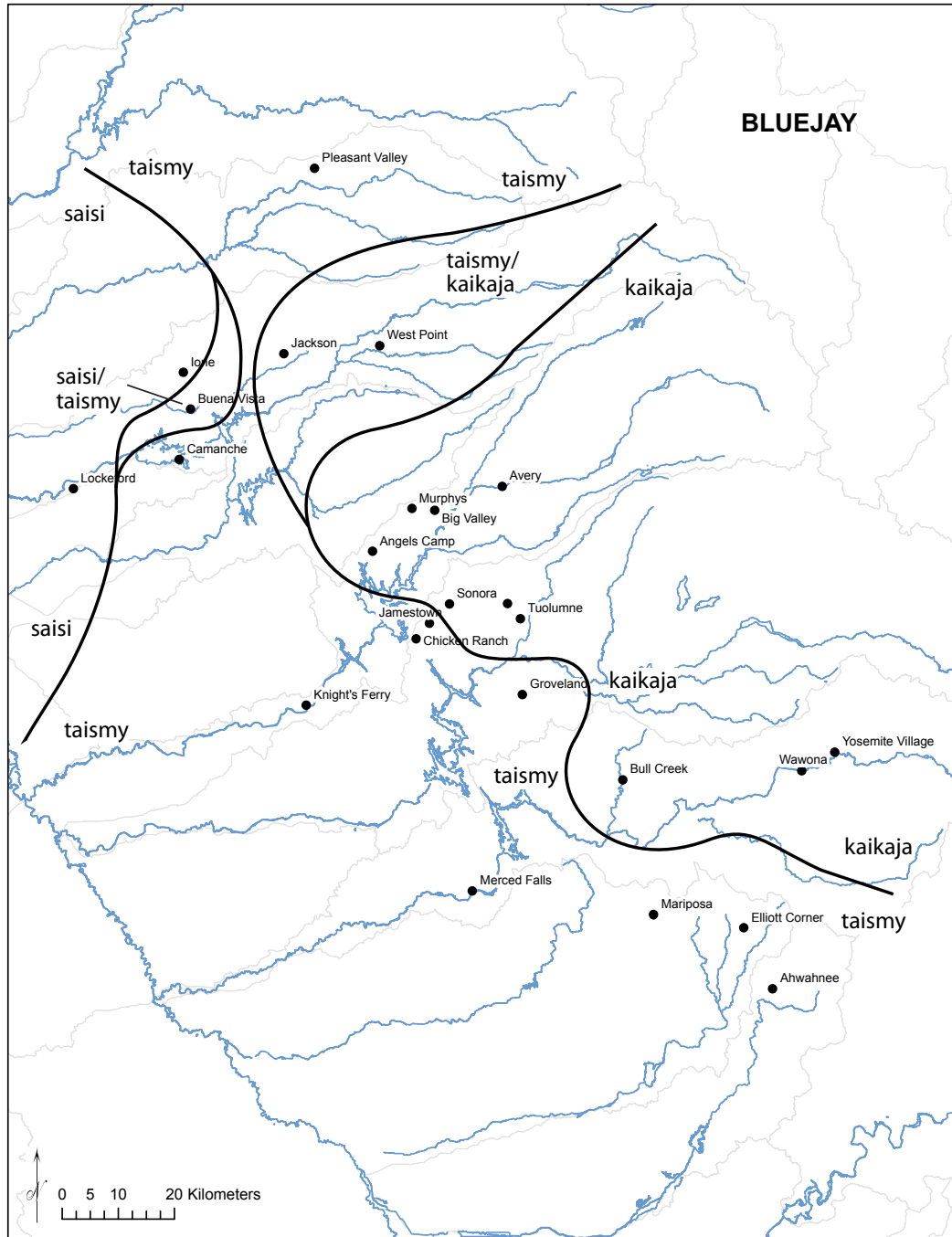


Figure 5.10: Eastern Miwok isogloss map: Bluejay

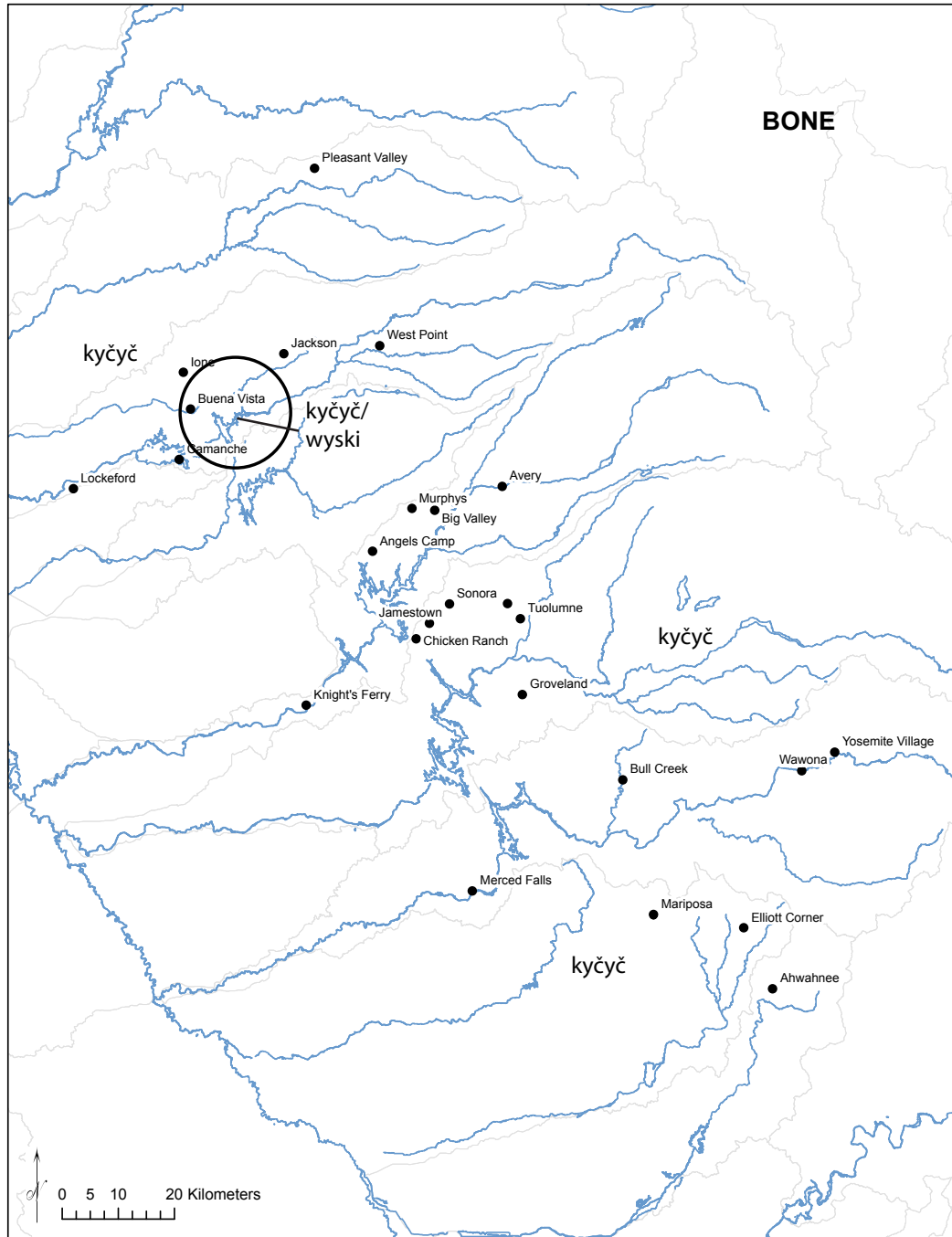


Figure 5.11: Eastern Miwok isogloss map: Bone

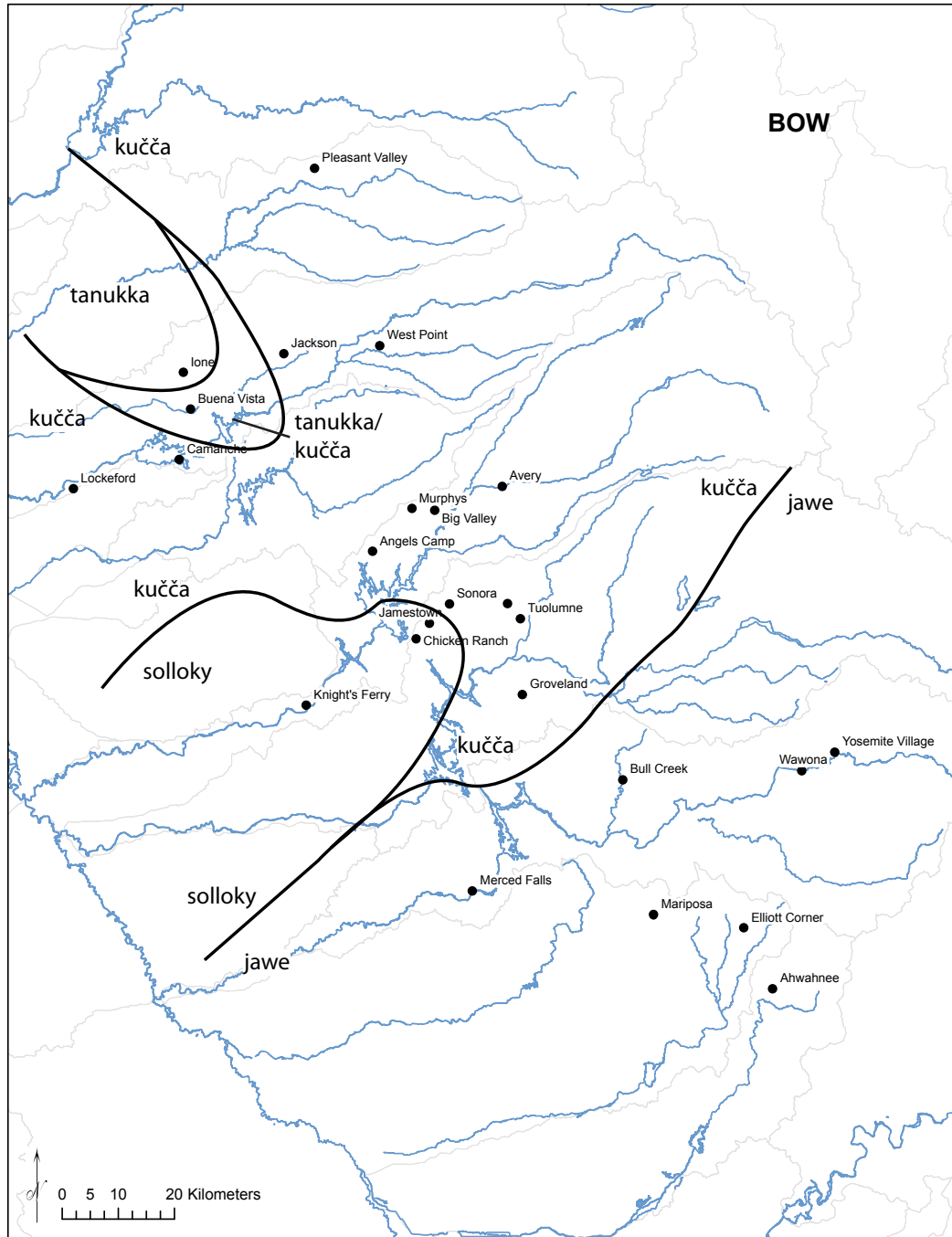


Figure 5.12: Eastern Miwok isogloss map: Bow

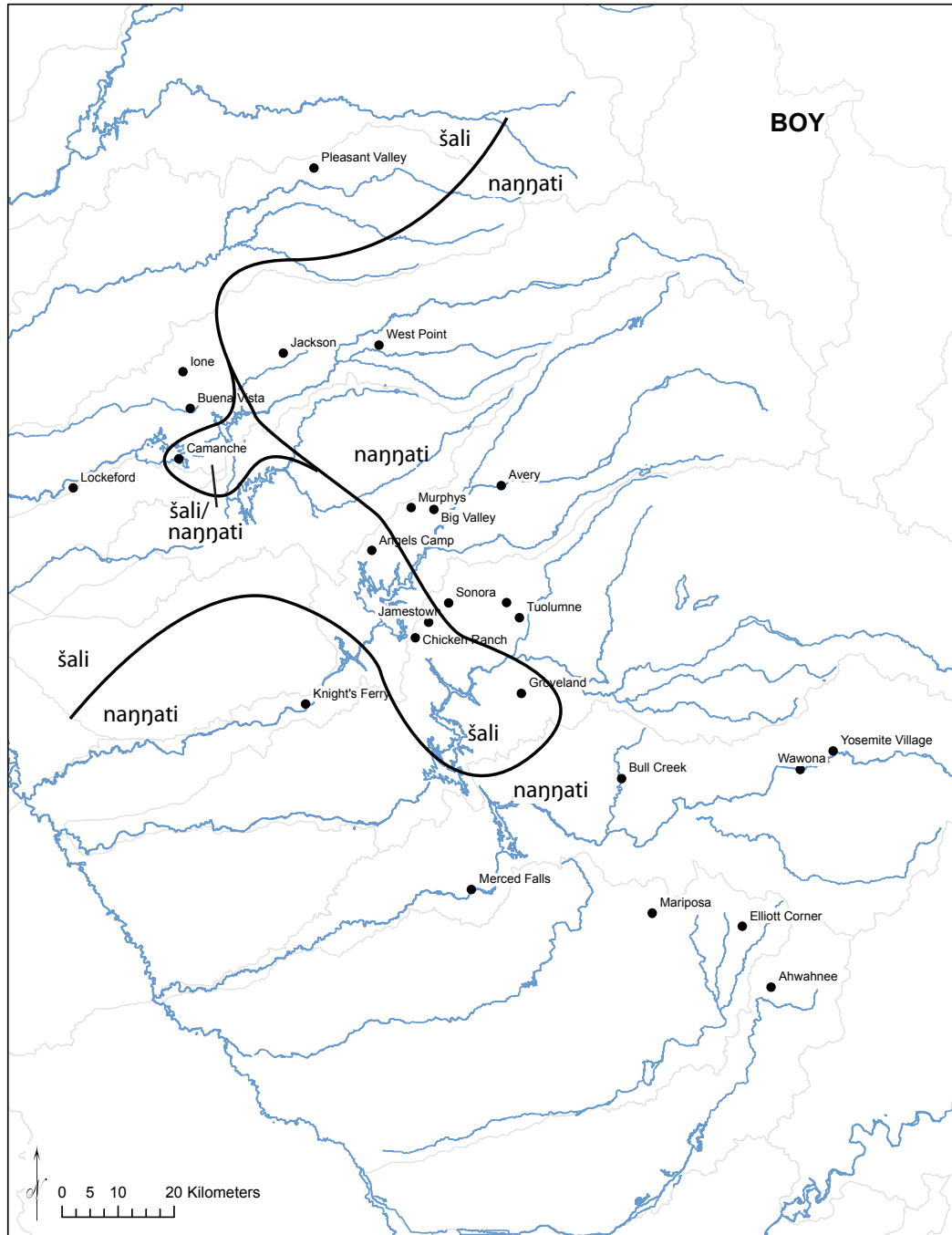


Figure 5.13: Eastern Miwok isogloss map: Boy

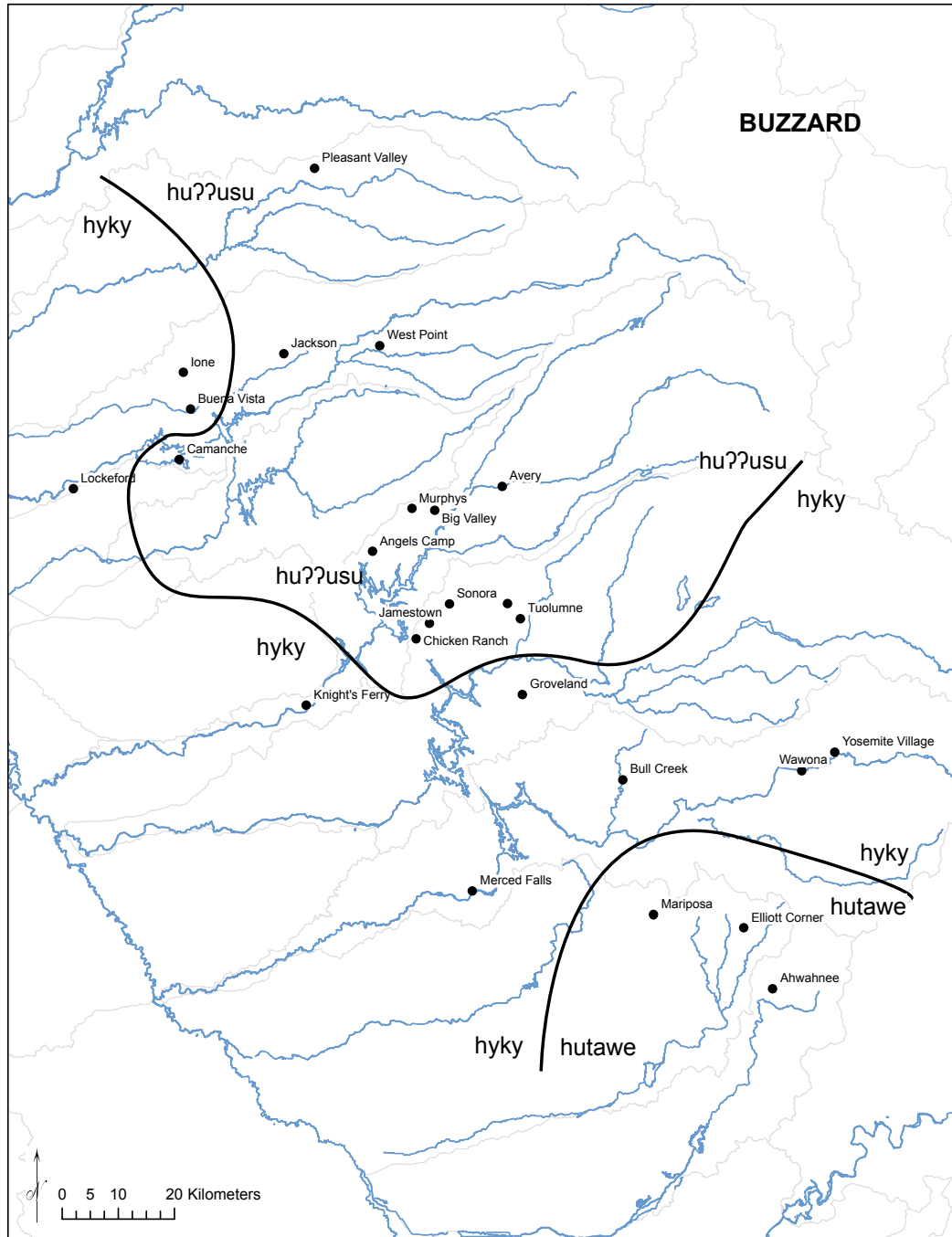


Figure 5.14: Eastern Miwok isogloss map: Buzzard

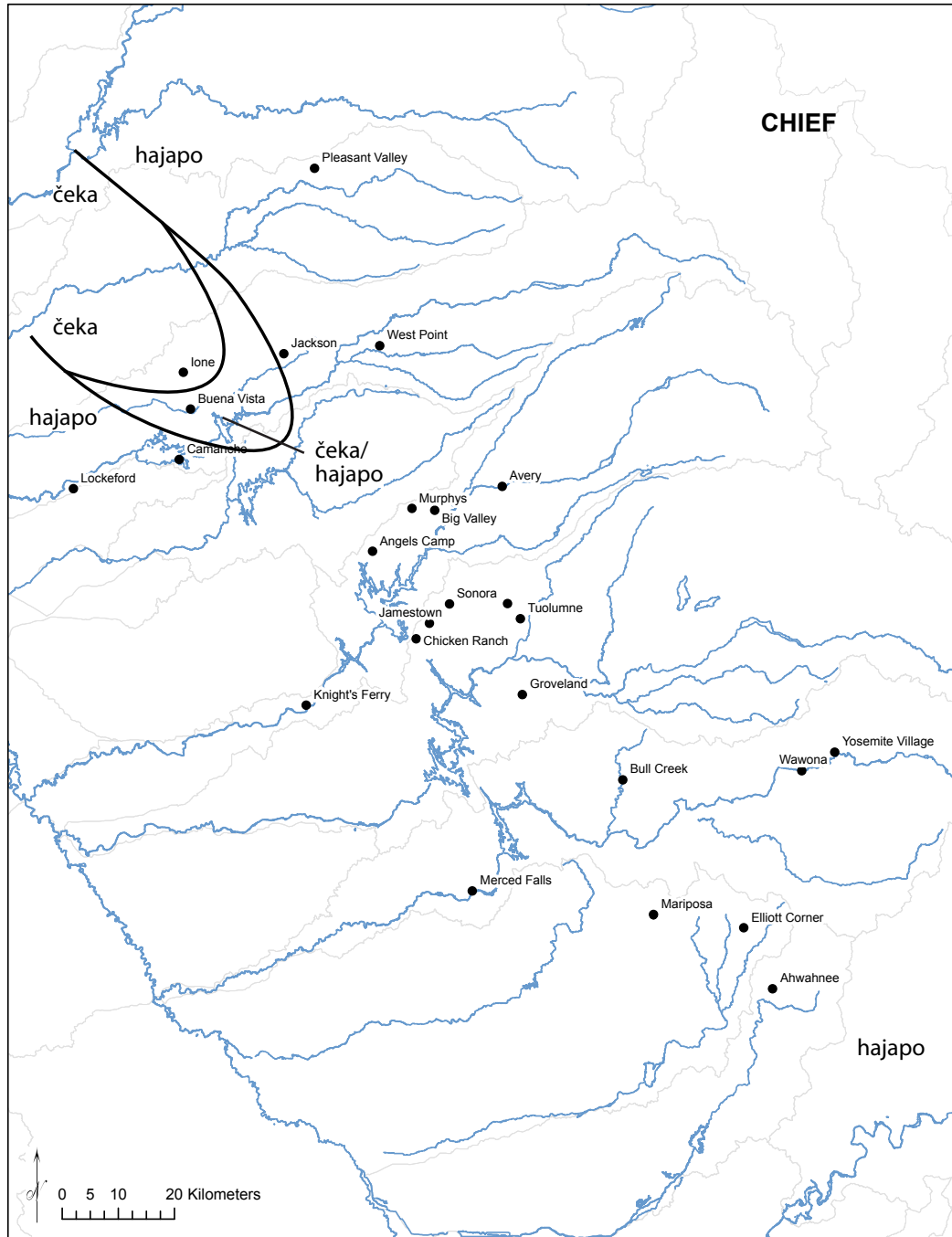


Figure 5.15: Eastern Miwok isogloss map: Chief

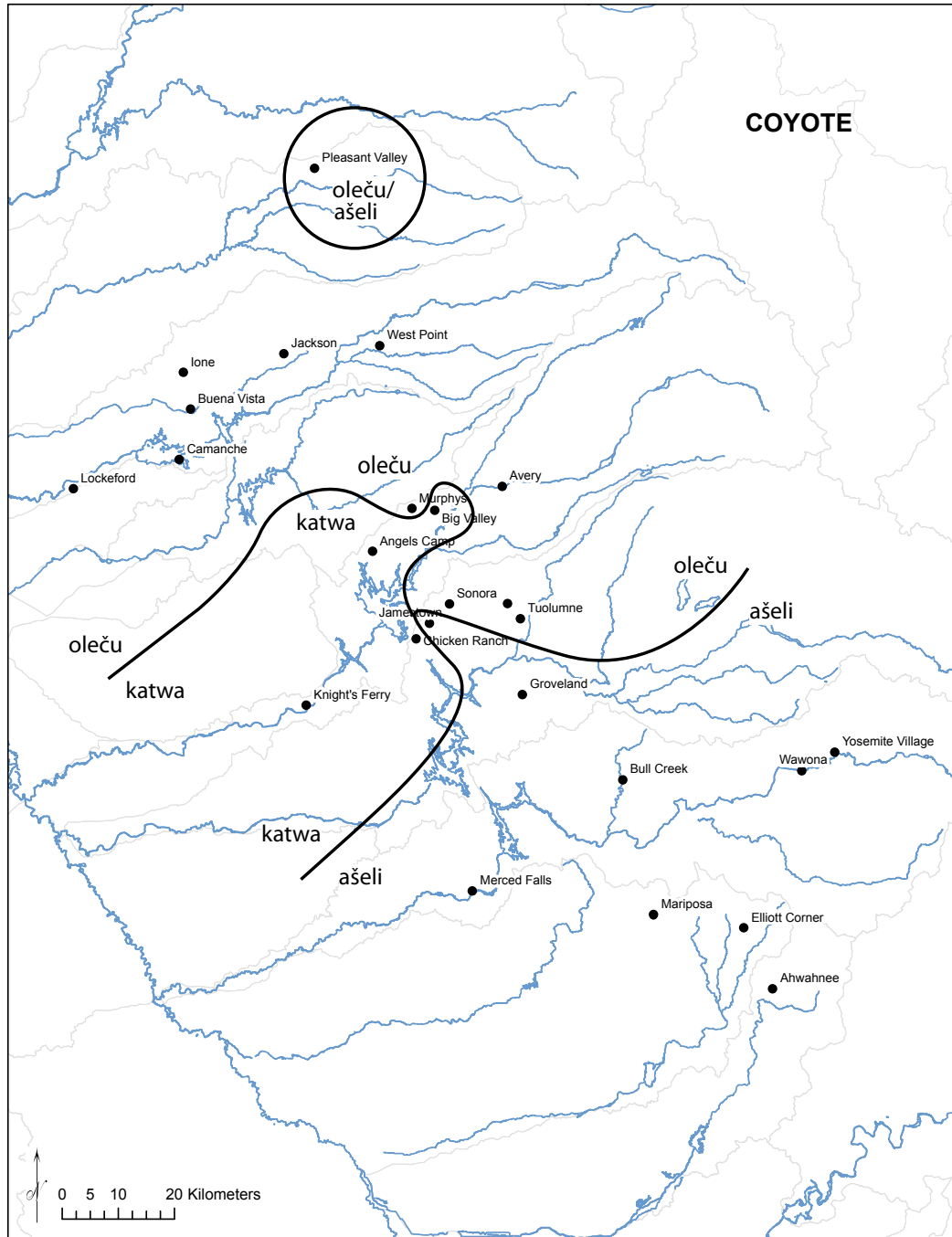


Figure 5.16: Eastern Miwok isogloss map: Coyote

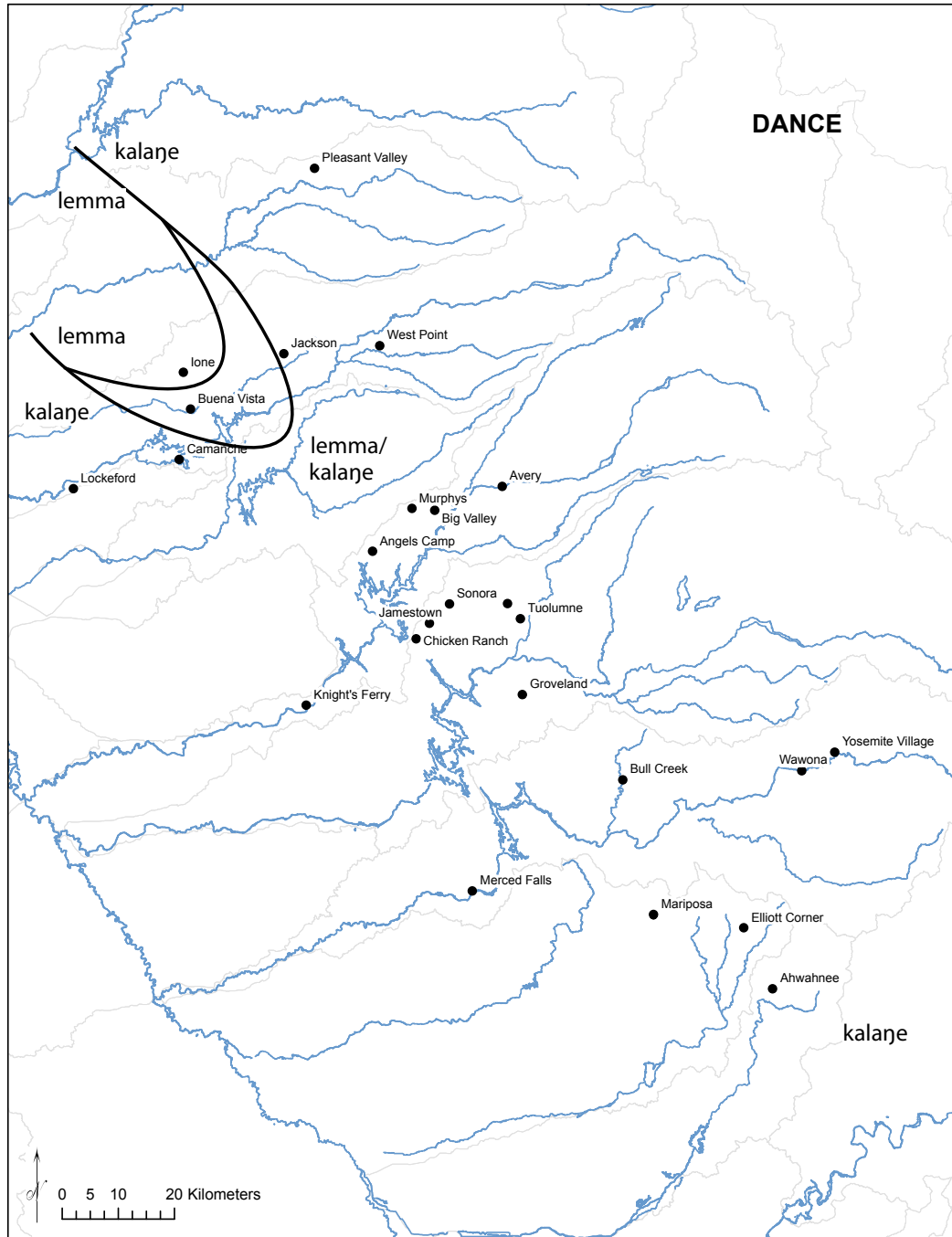


Figure 5.17: Eastern Miwok isogloss map: Dance

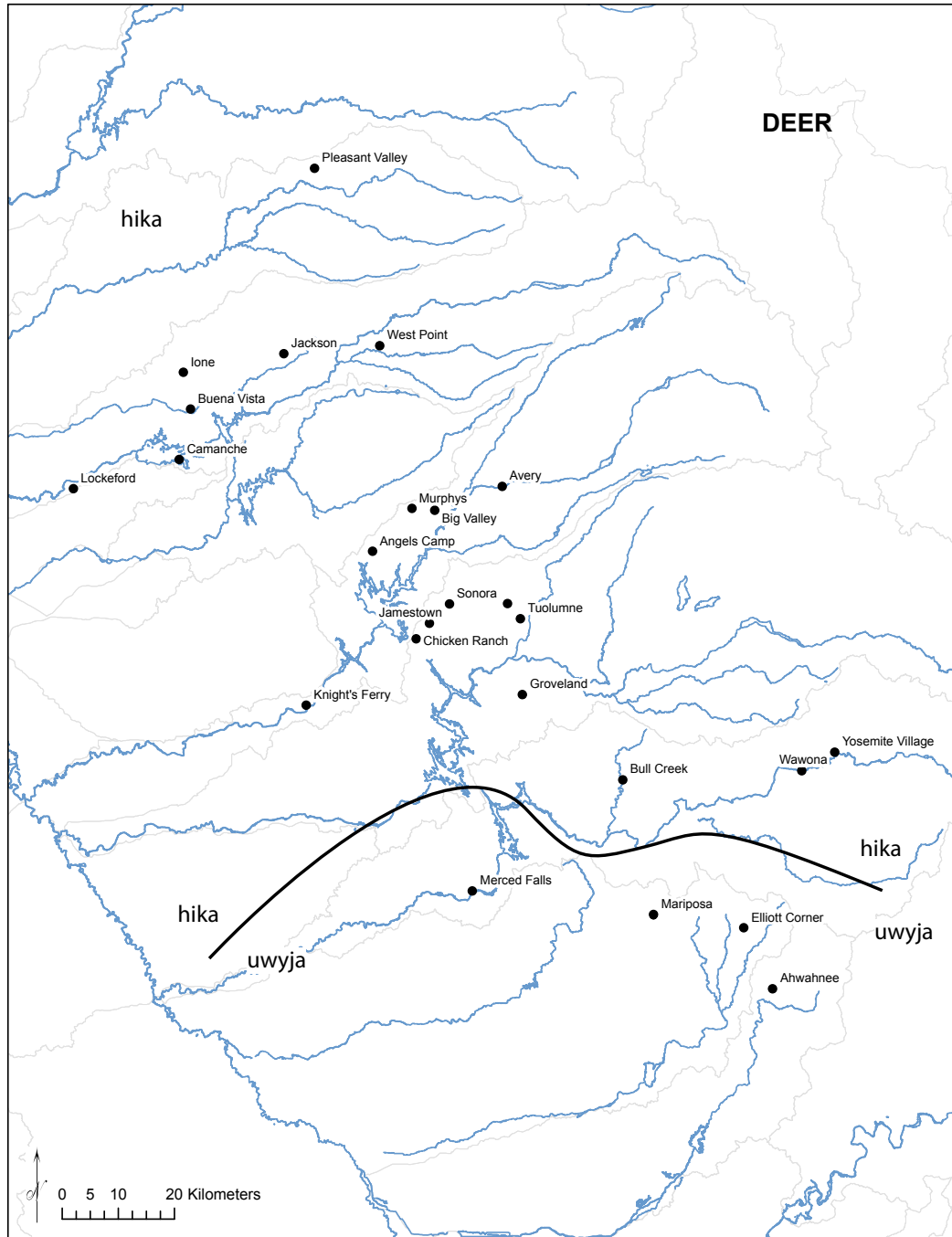


Figure 5.18: Eastern Miwok isogloss map: Deer

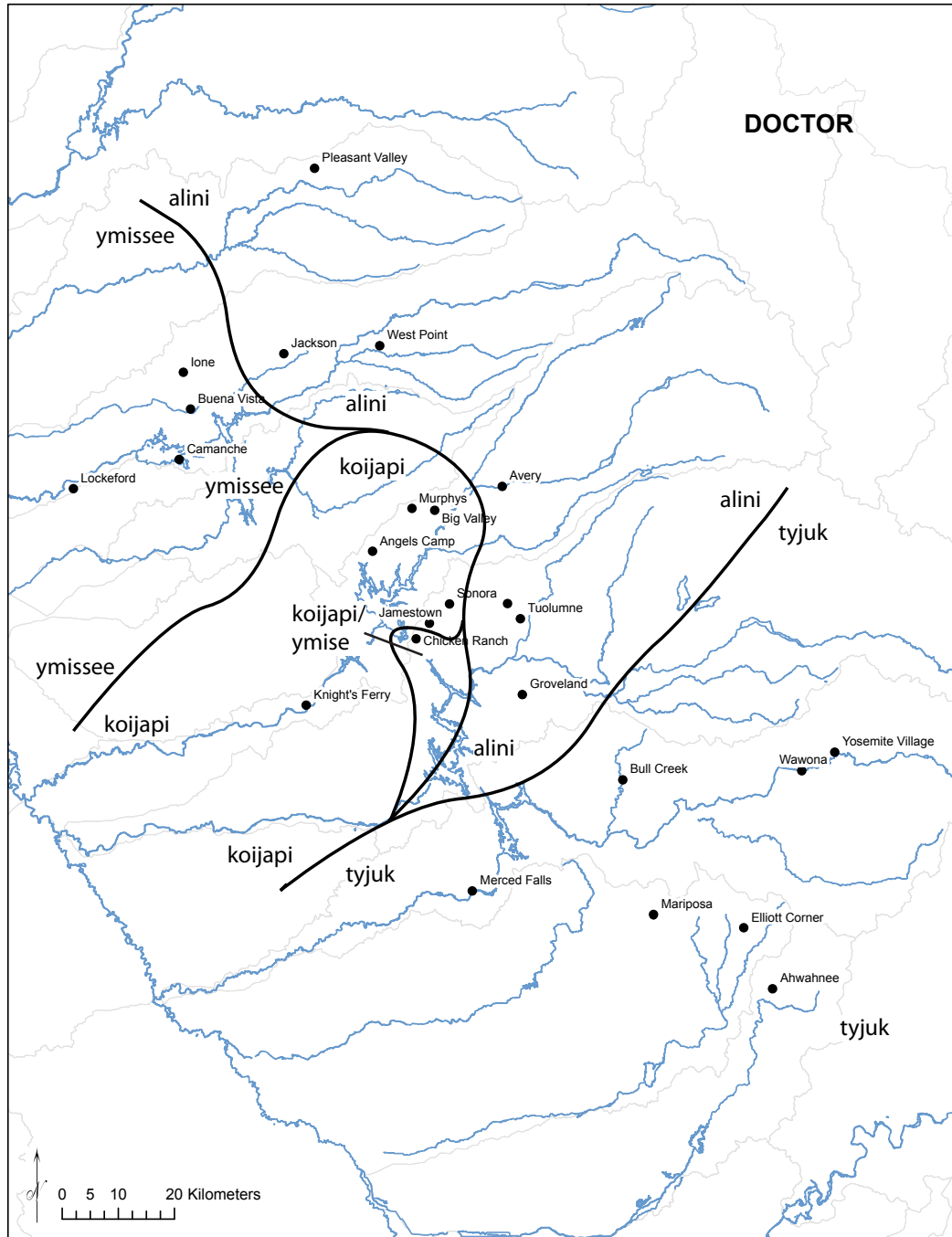


Figure 5.19: Eastern Miwok isogloss map: Doctor

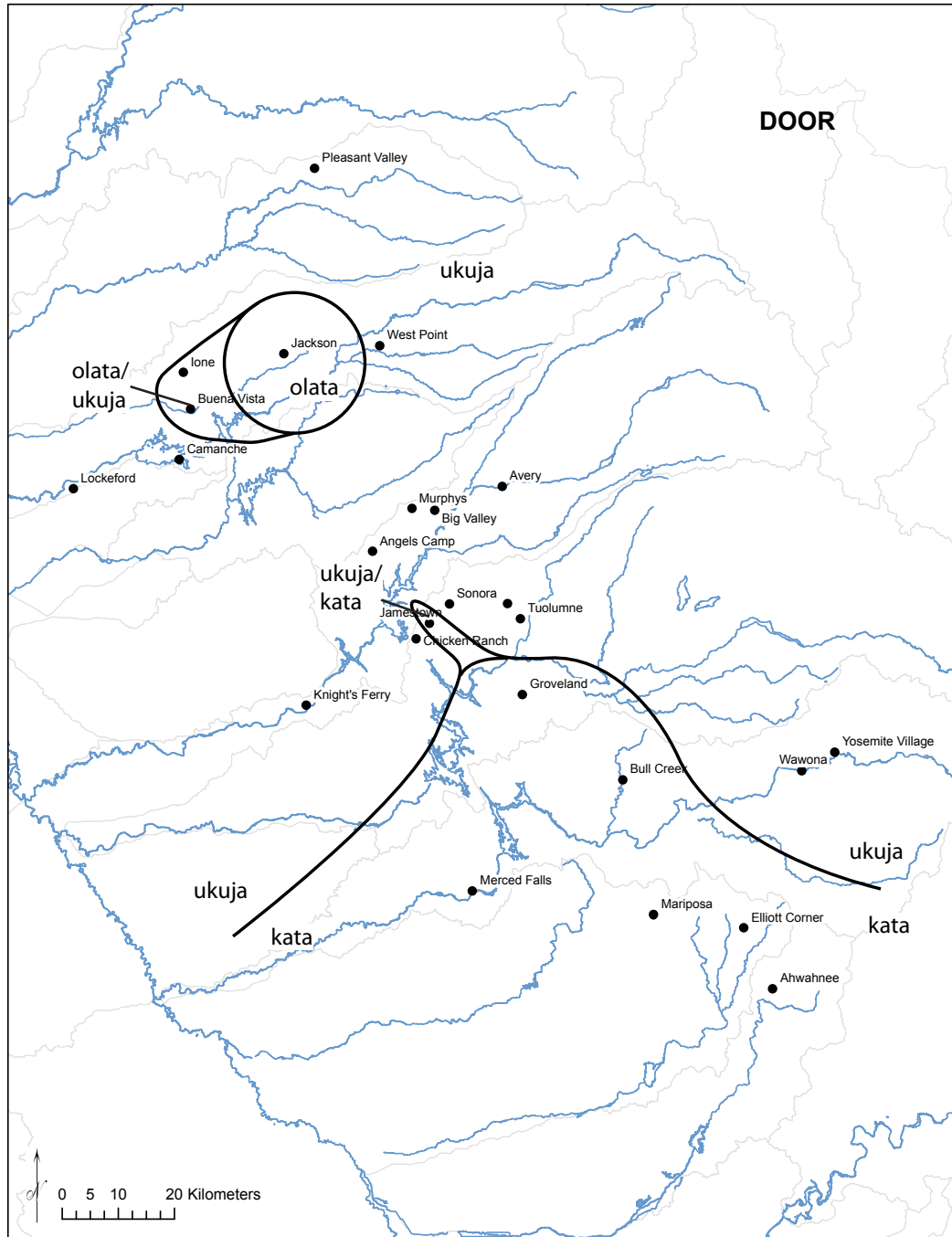


Figure 5.20: Eastern Miwok isogloss map: Door

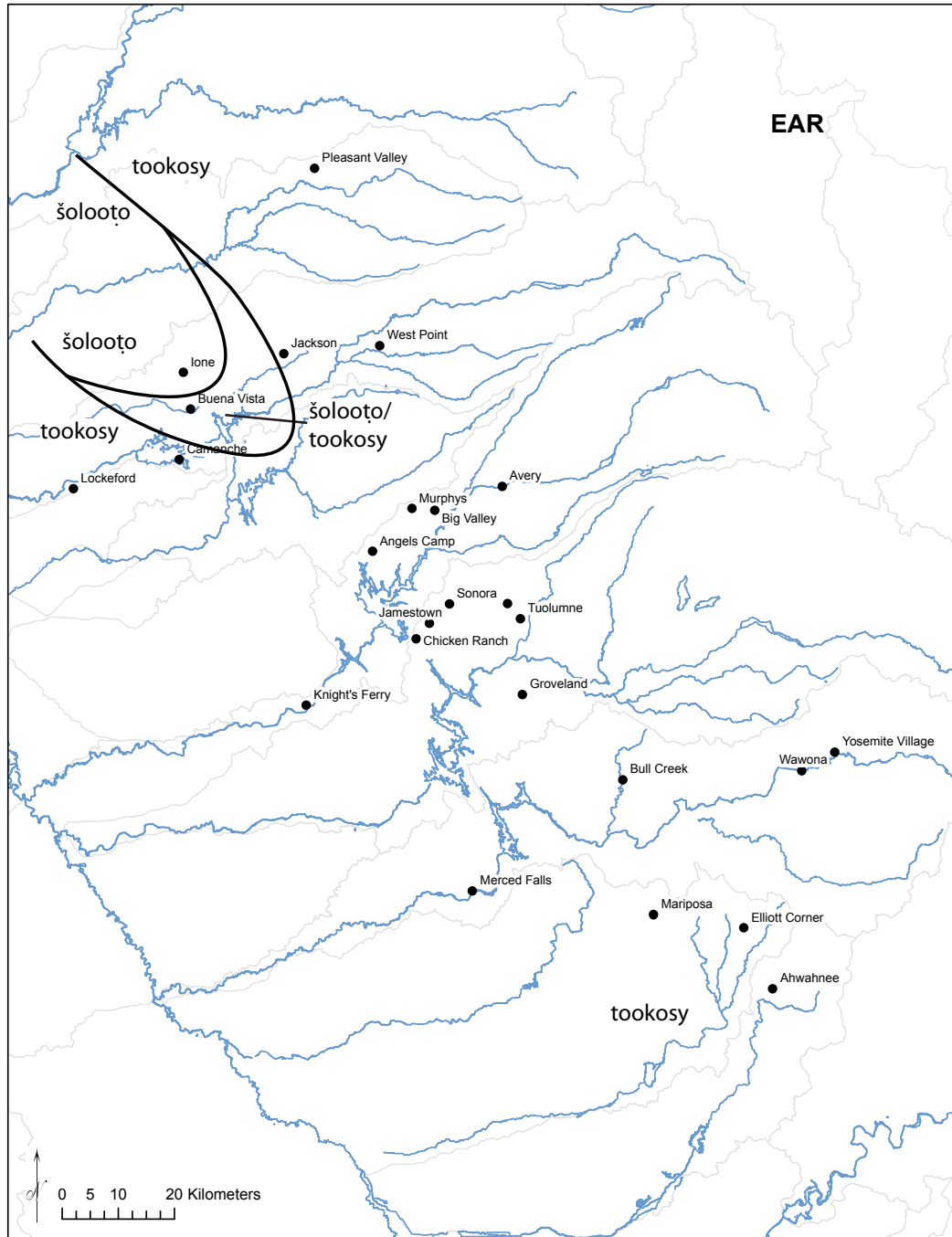


Figure 5.21: Eastern Miwok isogloss map: Ear

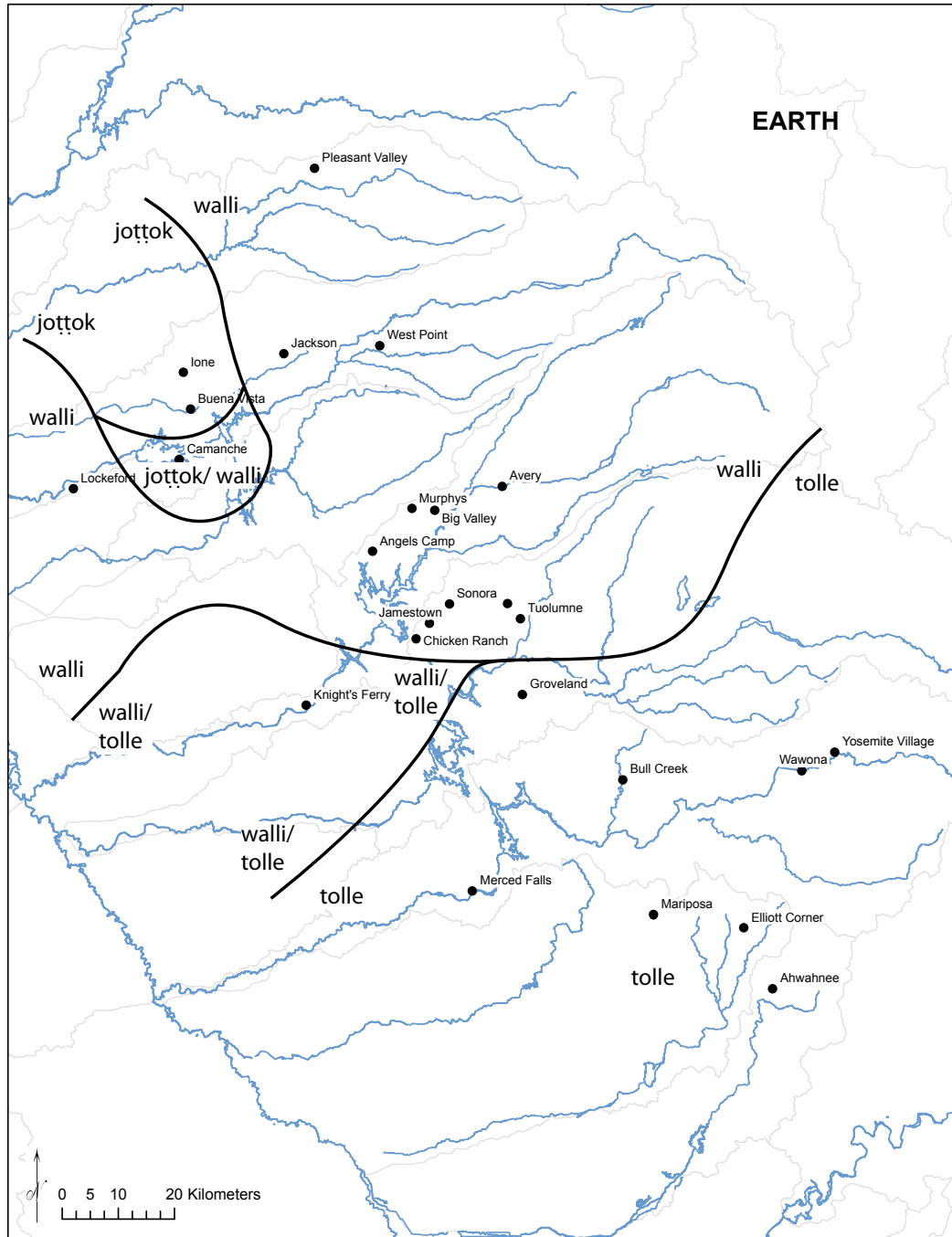


Figure 5.22: Eastern Miwok isogloss map: Earth

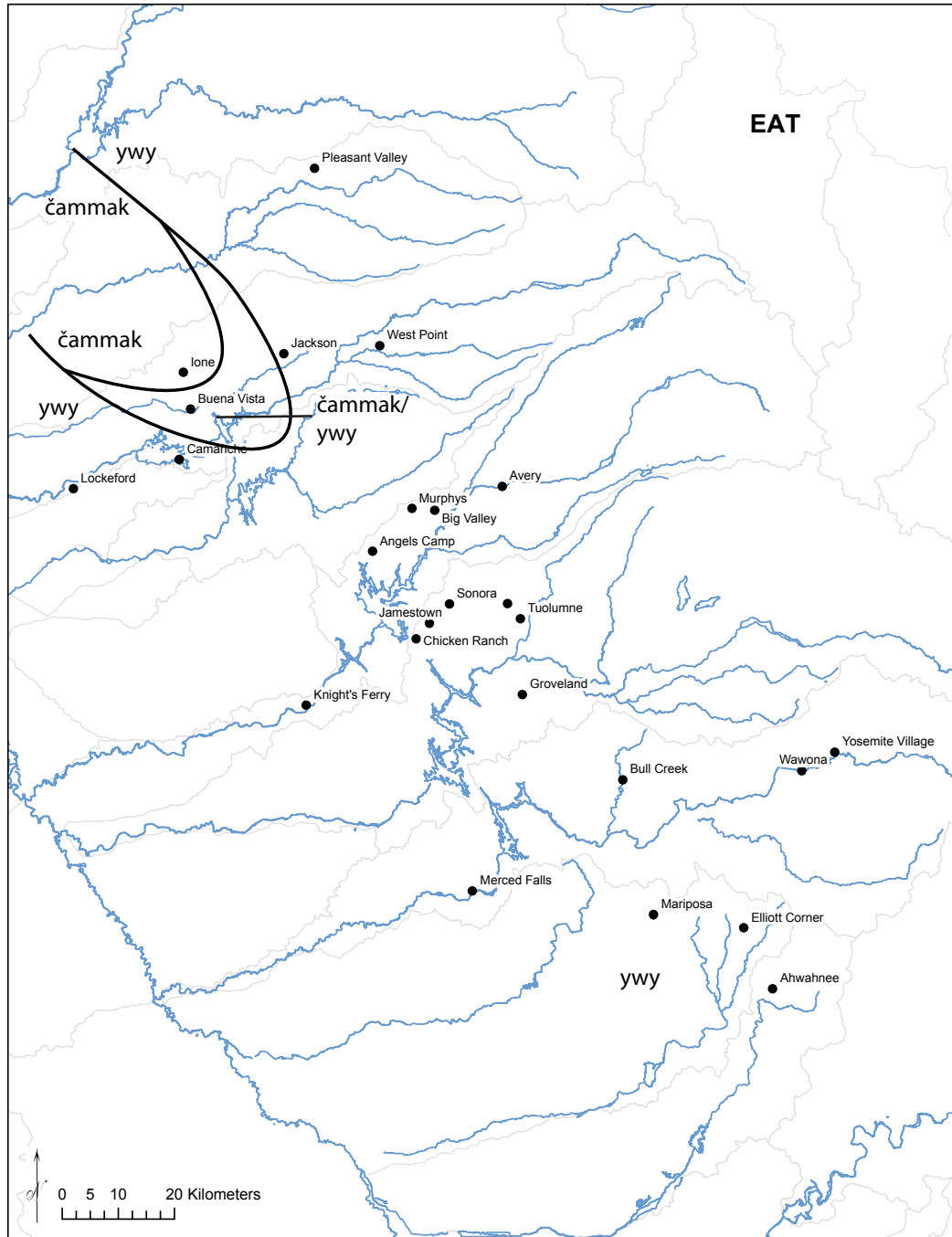


Figure 5.23: Eastern Miwok isogloss map: Eat

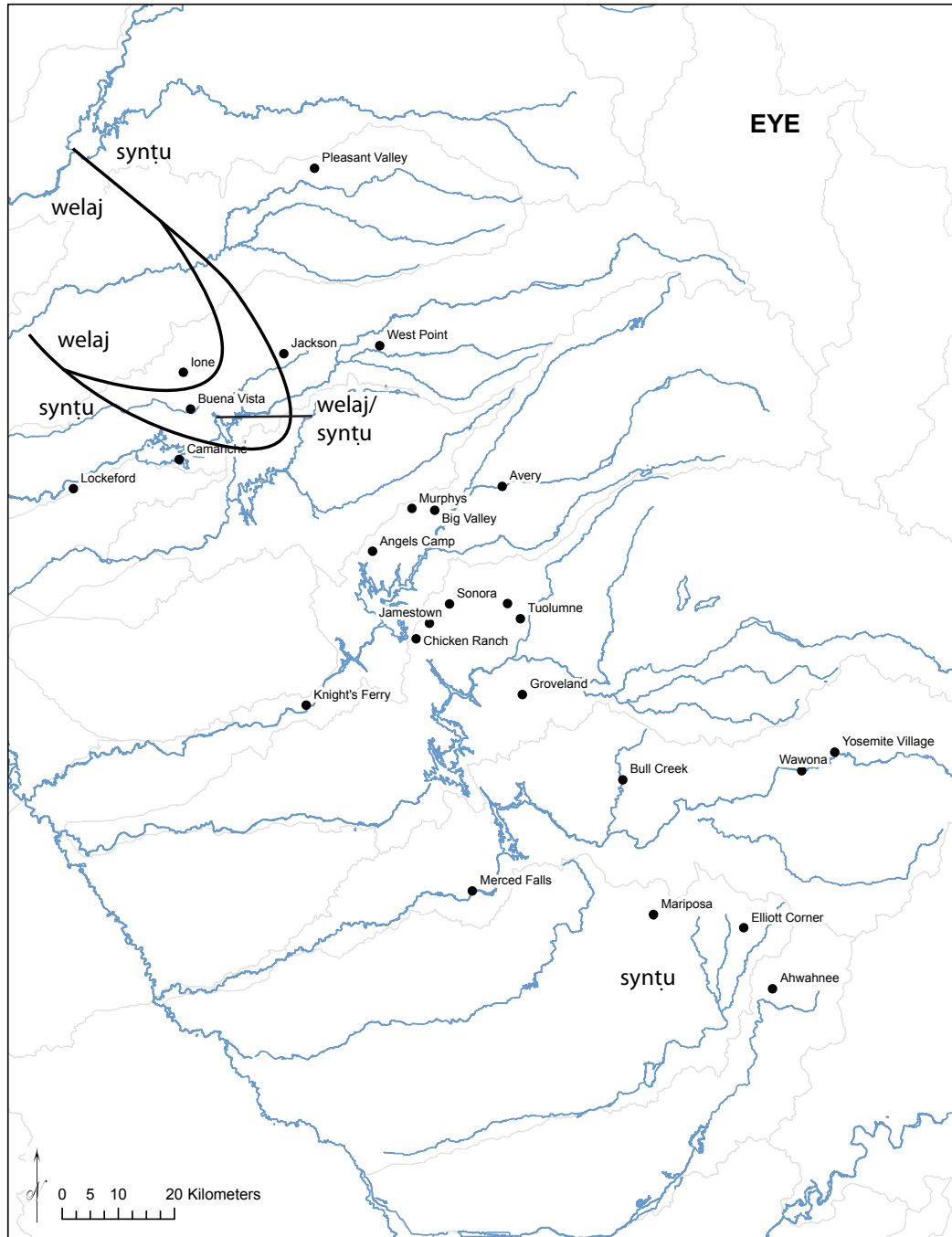


Figure 5.24: Eastern Miwok isogloss map: Eye

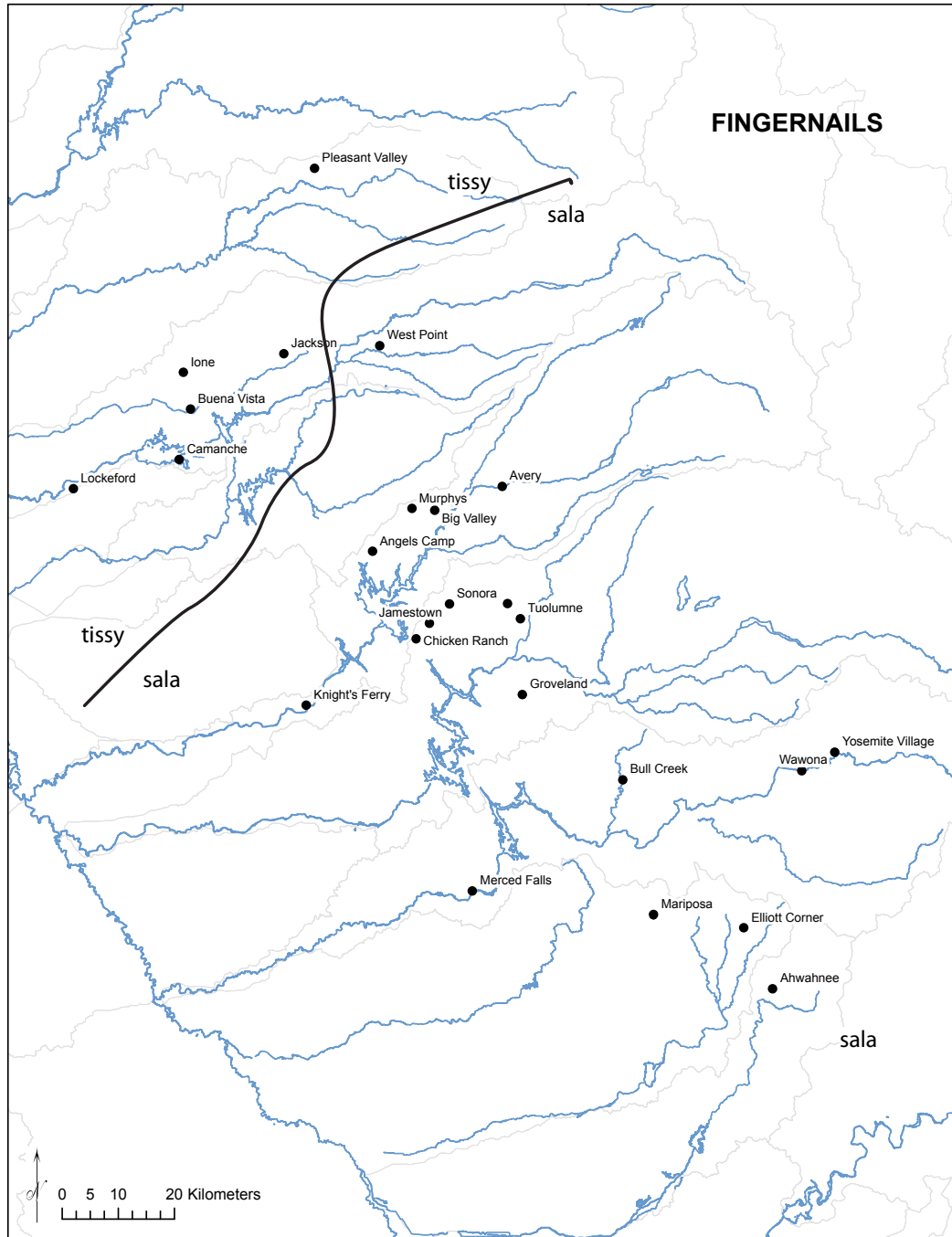


Figure 5.25: Eastern Miwok isogloss map: Fingernails

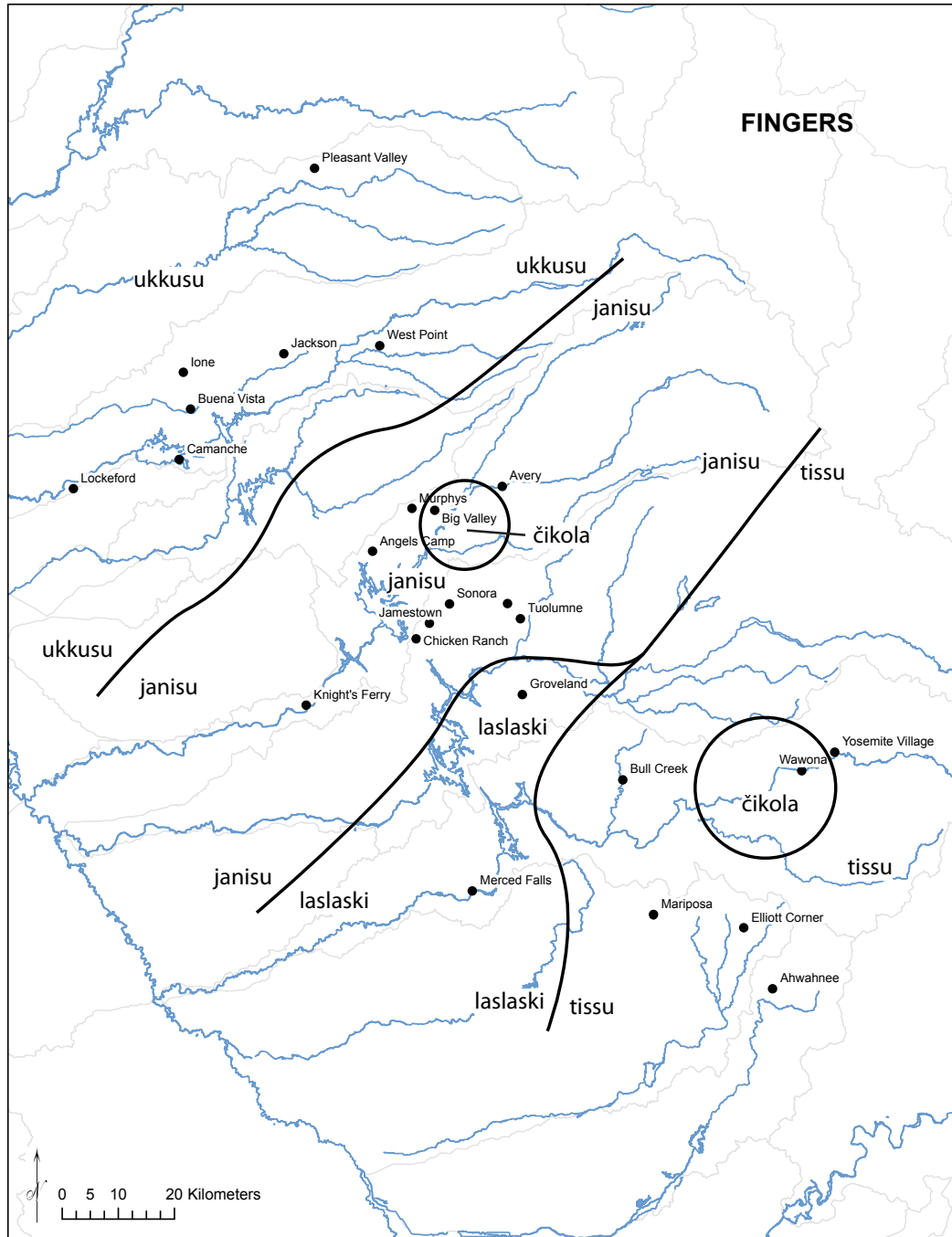


Figure 5.26: Eastern Miwok isogloss map: Fingers

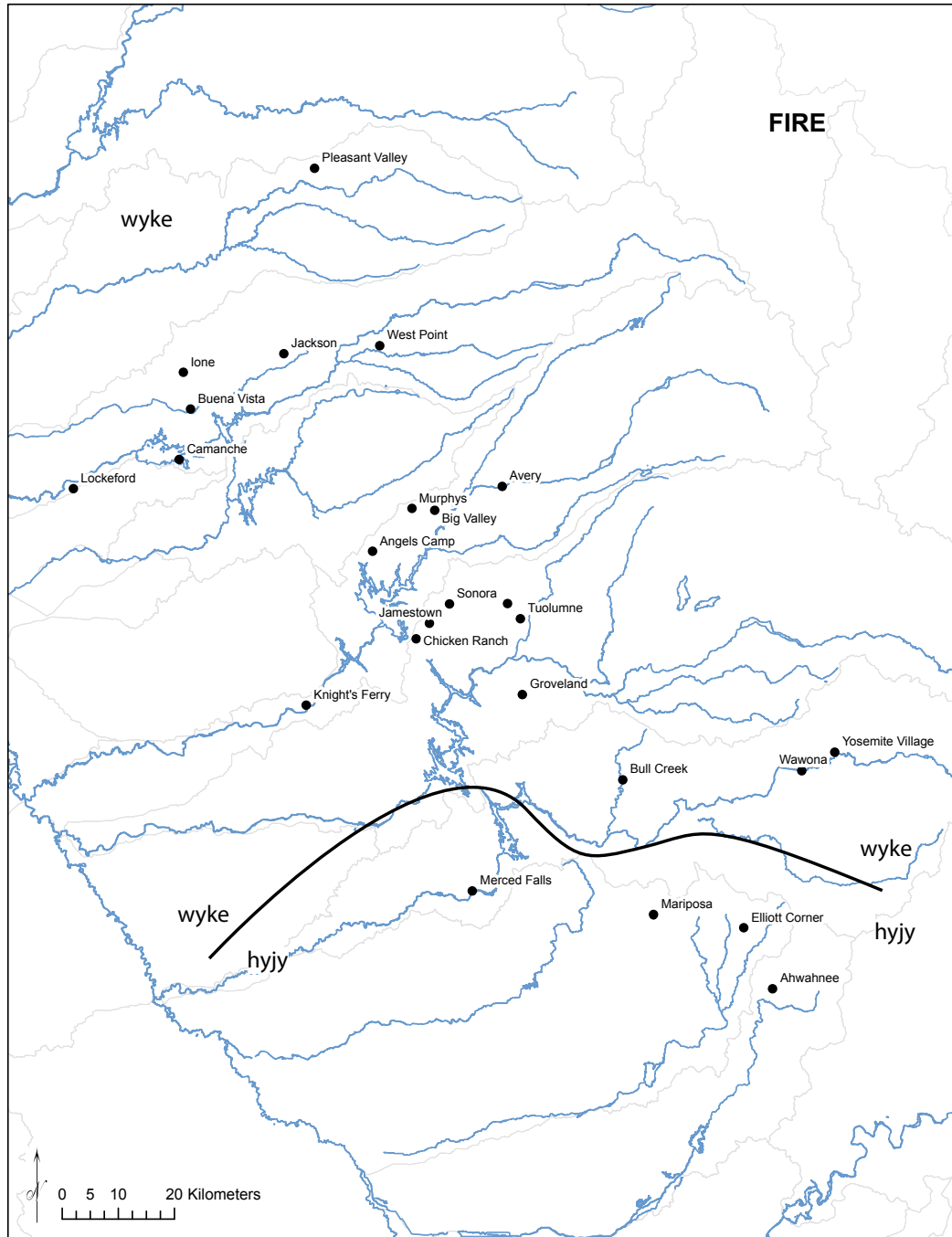


Figure 5.27: Eastern Miwok isogloss map: Fire

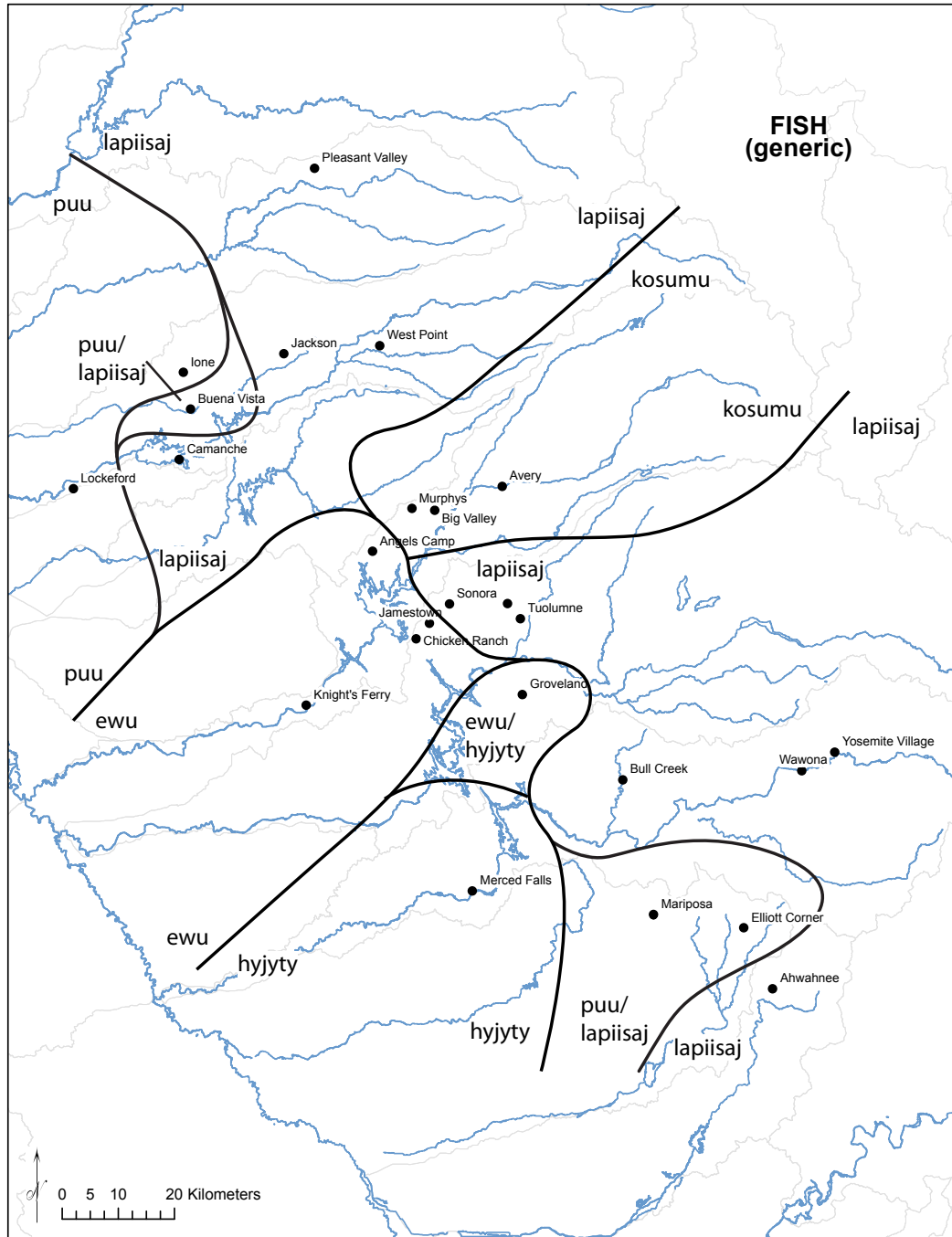


Figure 5.28: Eastern Miwok isogloss map: Fish (generic)

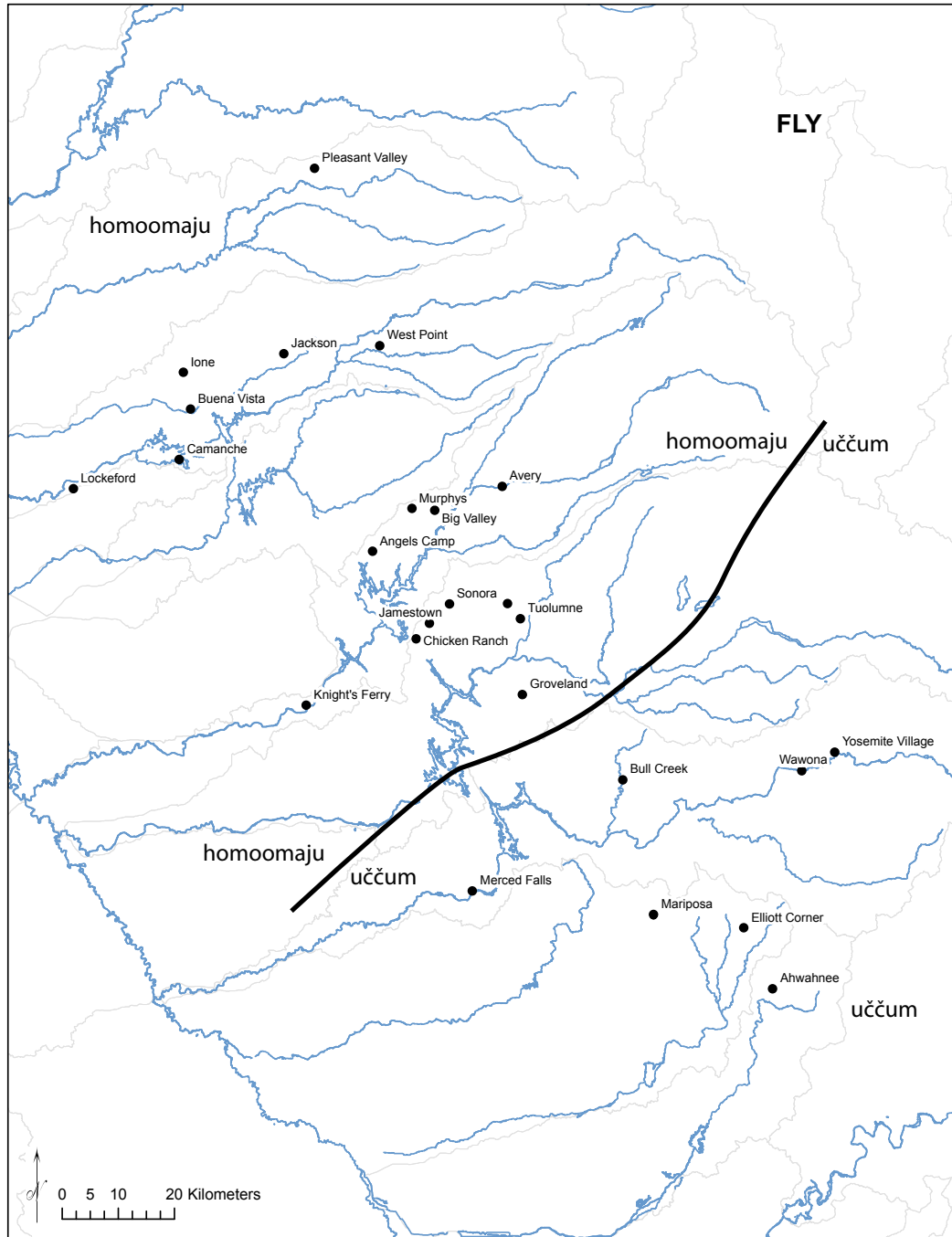


Figure 5.29: Eastern Miwok isogloss map: Fly

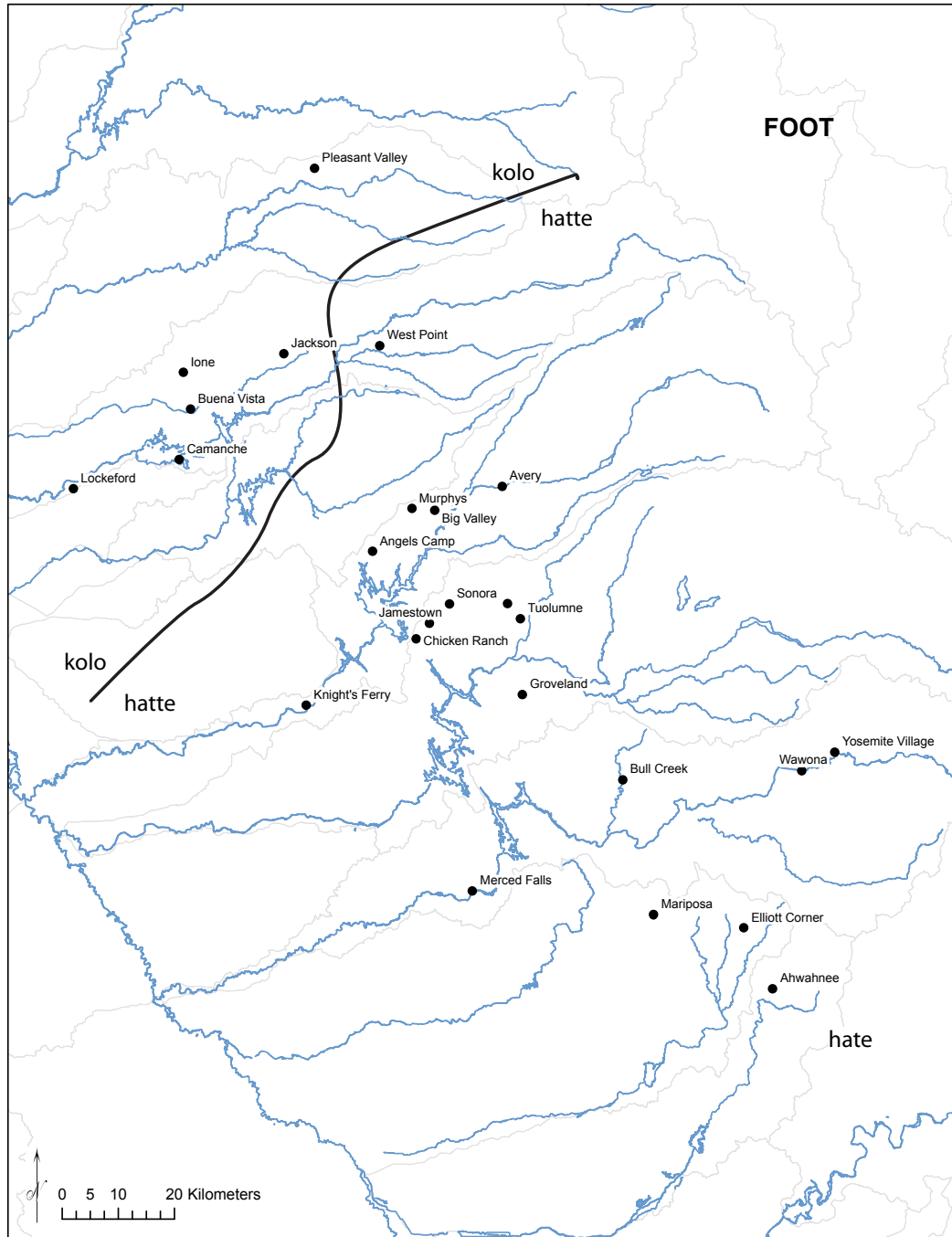


Figure 5.30: Eastern Miwok isogloss map: Foot

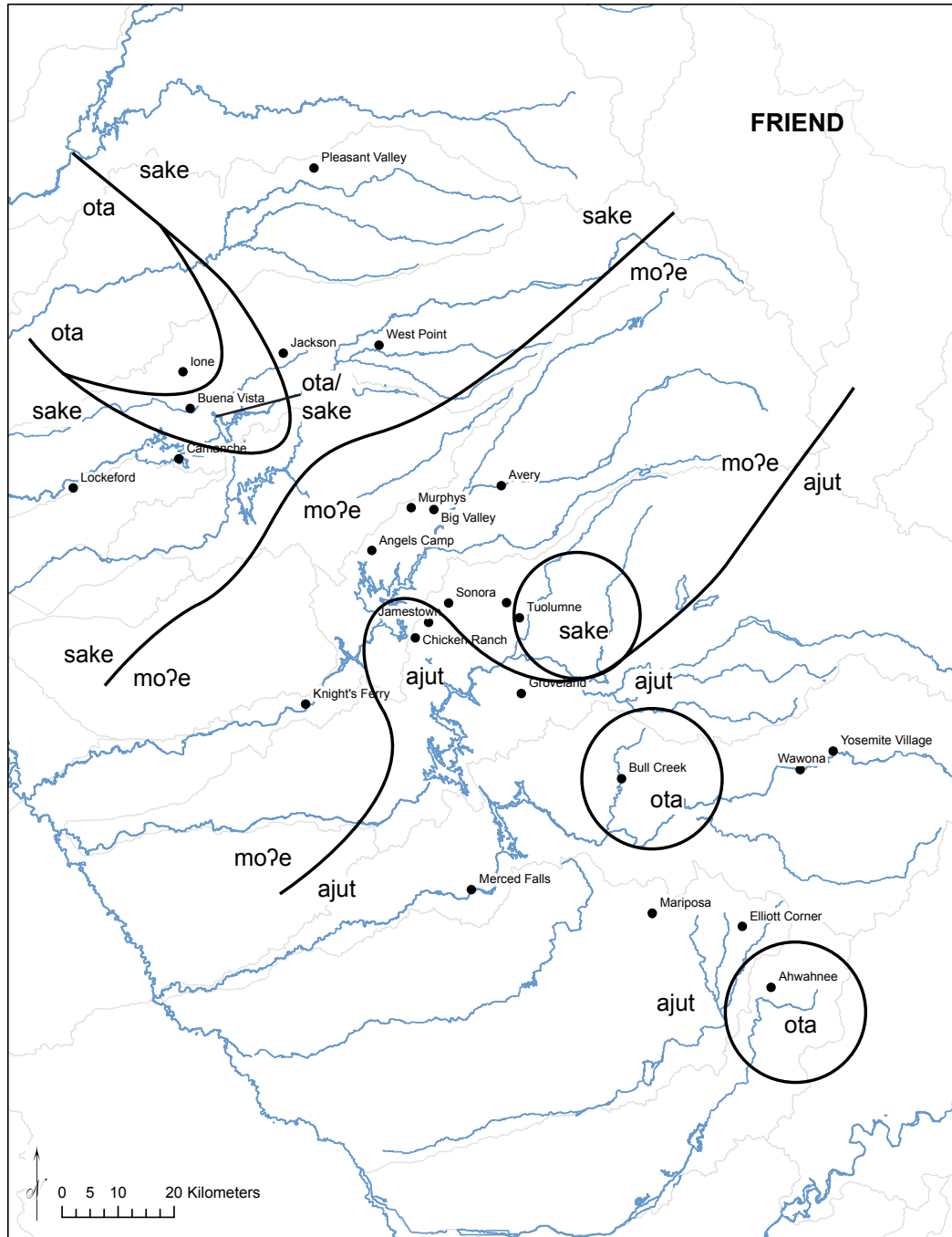


Figure 5.31: Eastern Miwok isogloss map: Friend

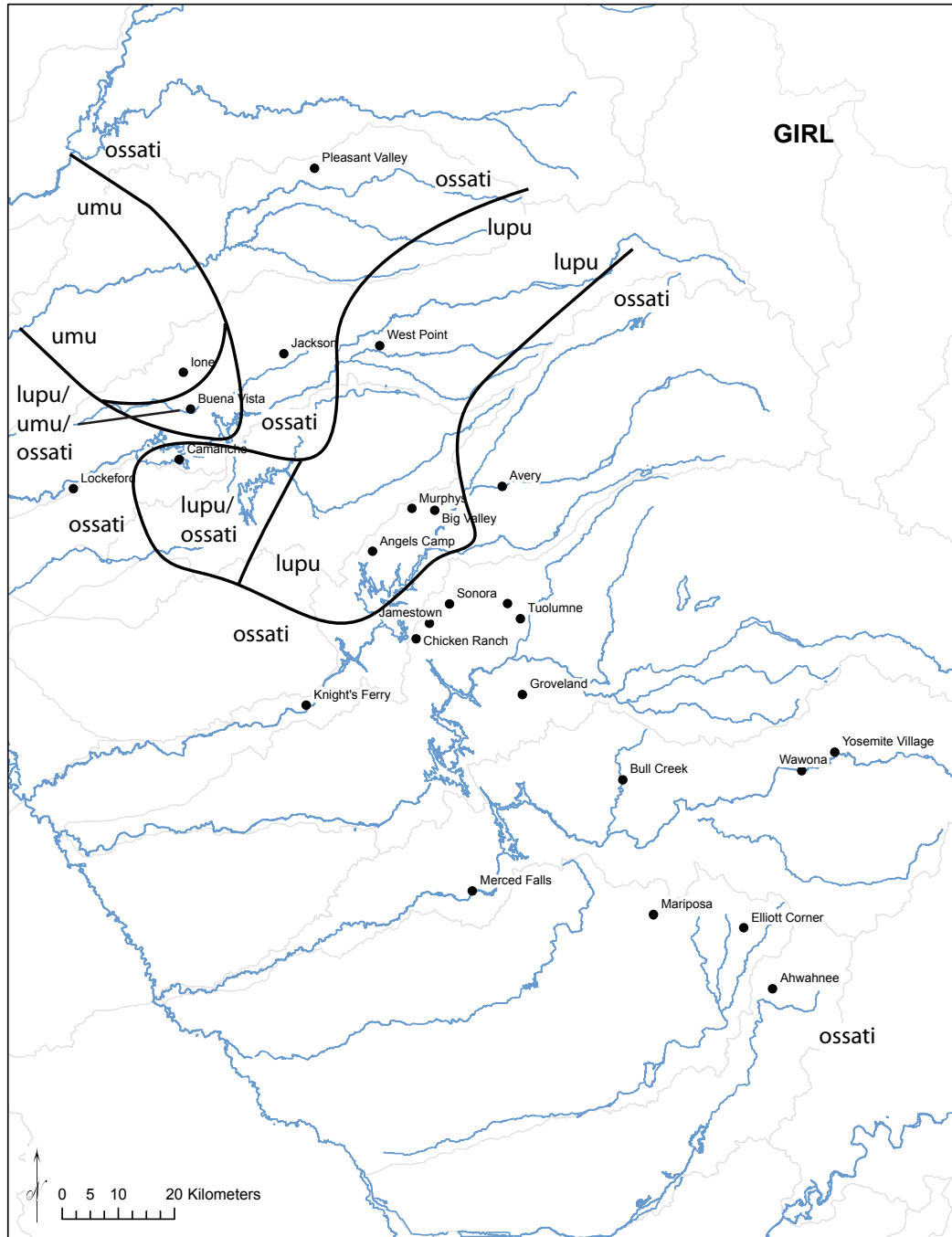


Figure 5.32: Eastern Miwok isogloss map: Girl

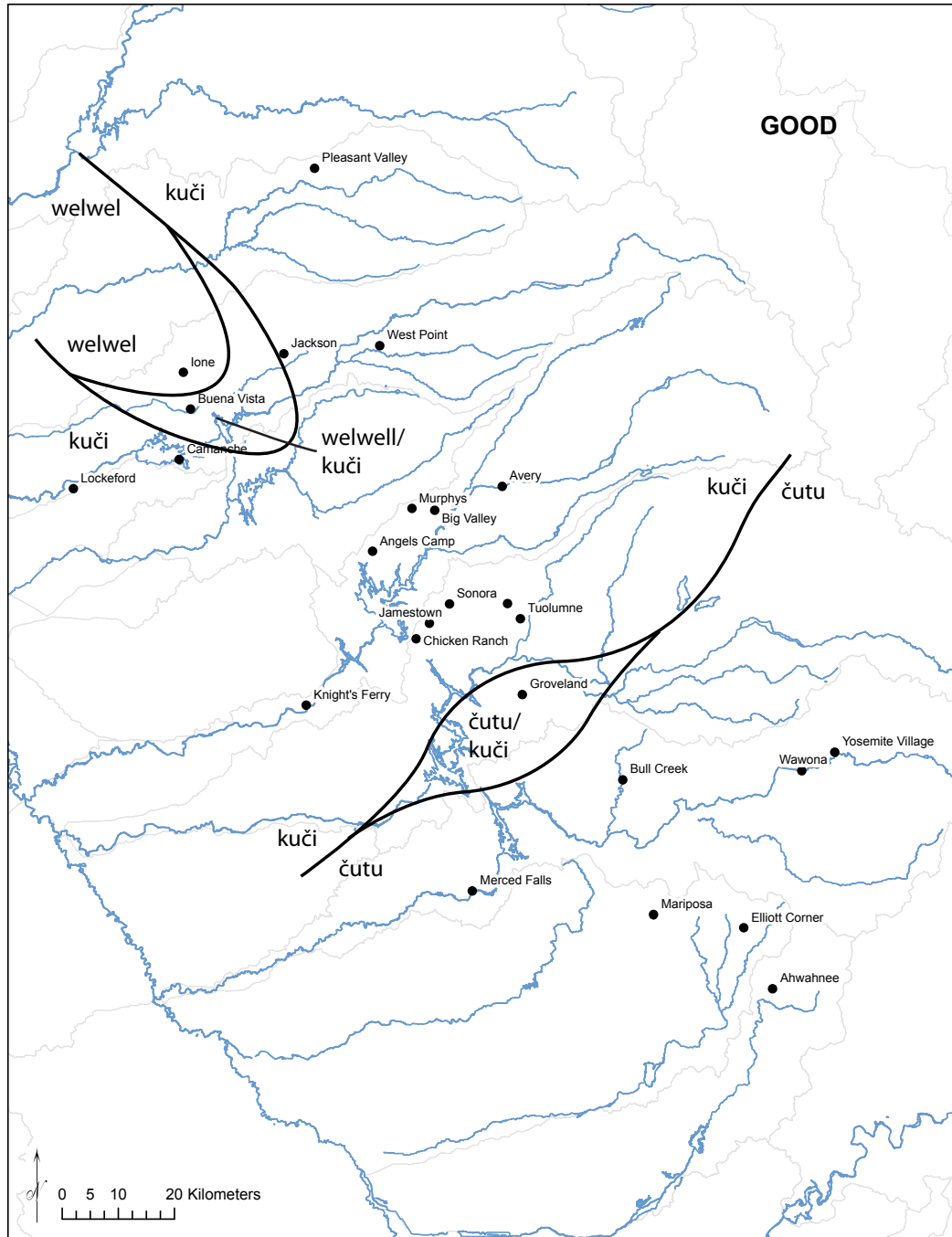


Figure 5.33: Eastern Miwok isogloss map: Good

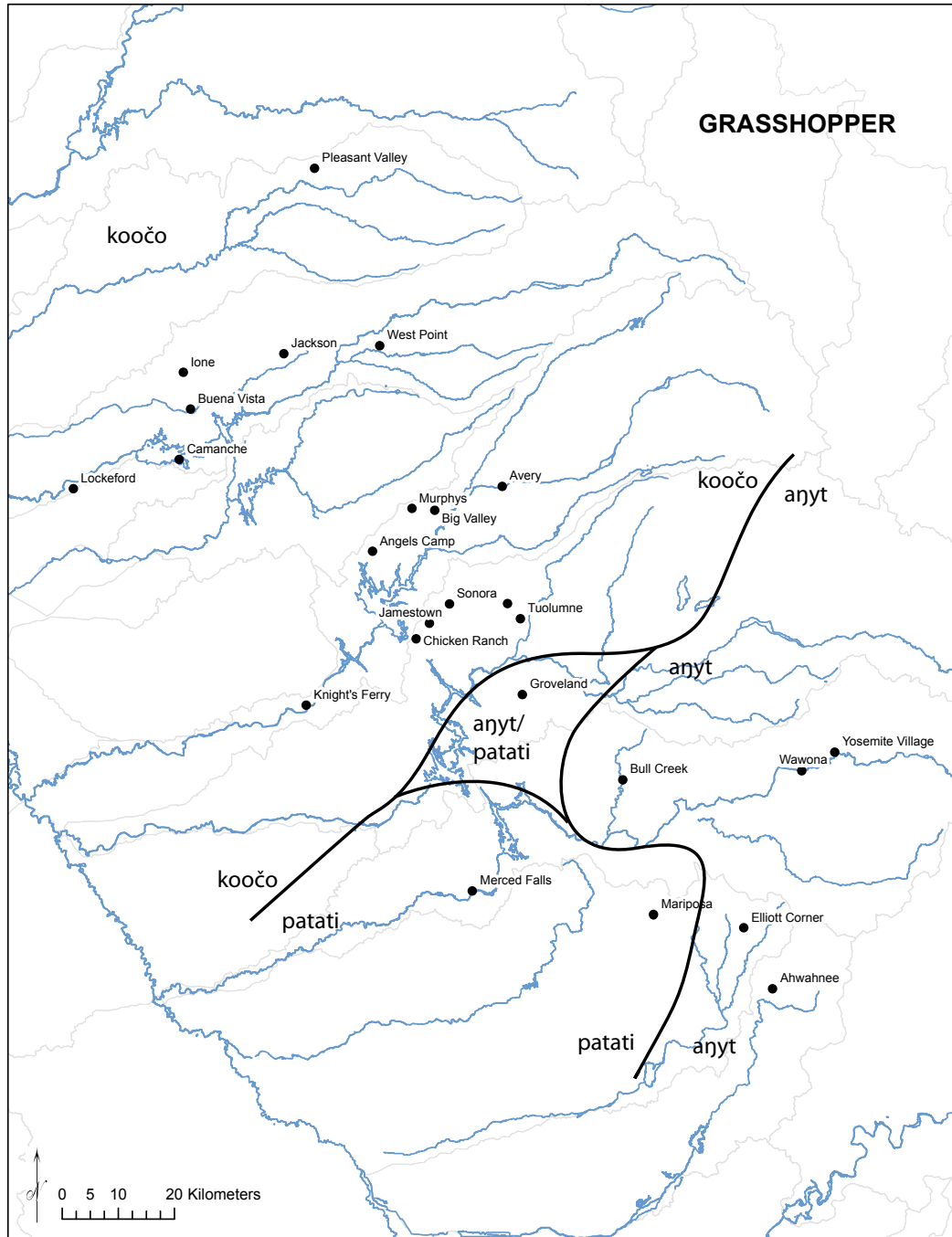


Figure 5.34: Eastern Miwok isogloss map: Grasshopper

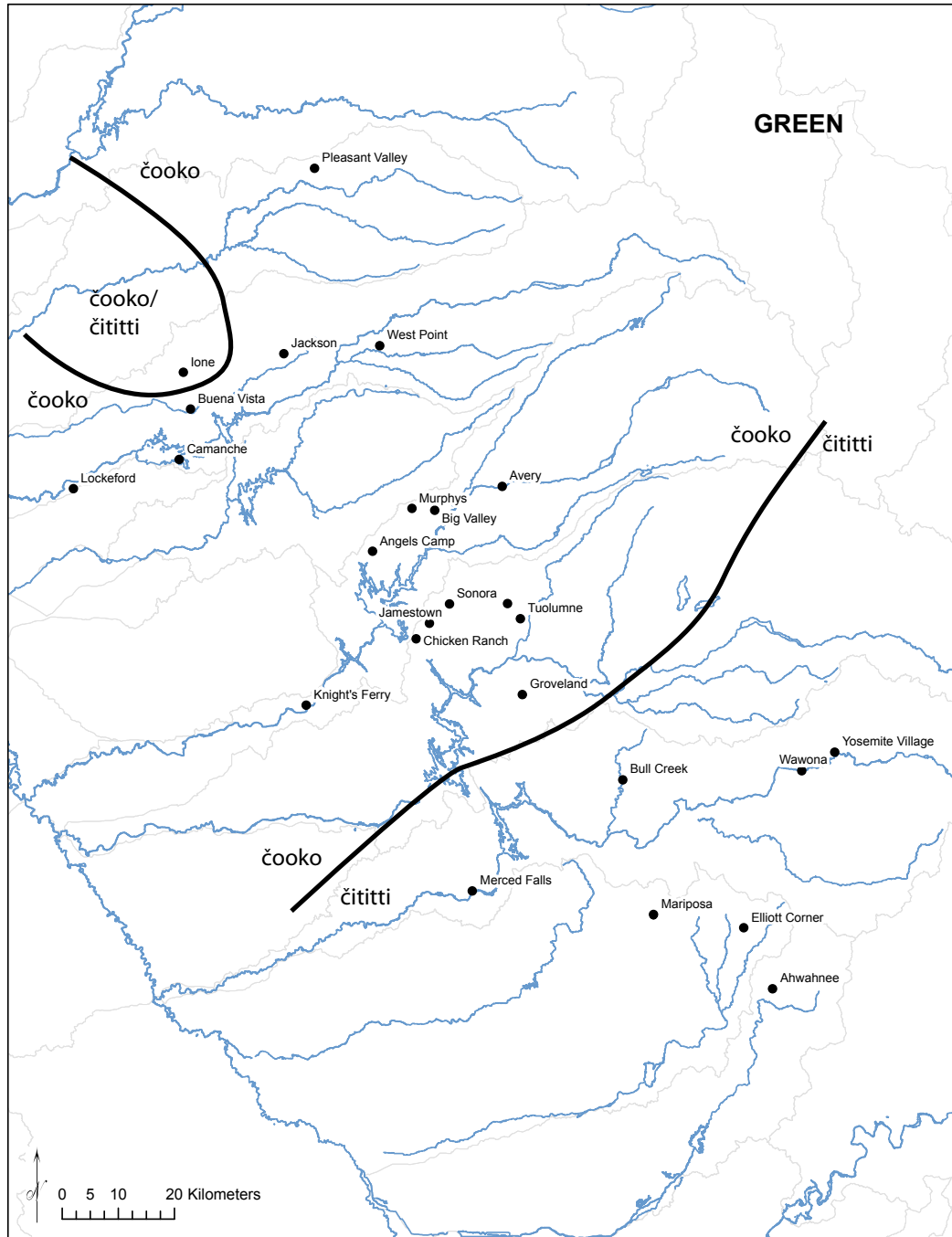


Figure 5.35: Eastern Miwok isogloss map: Green

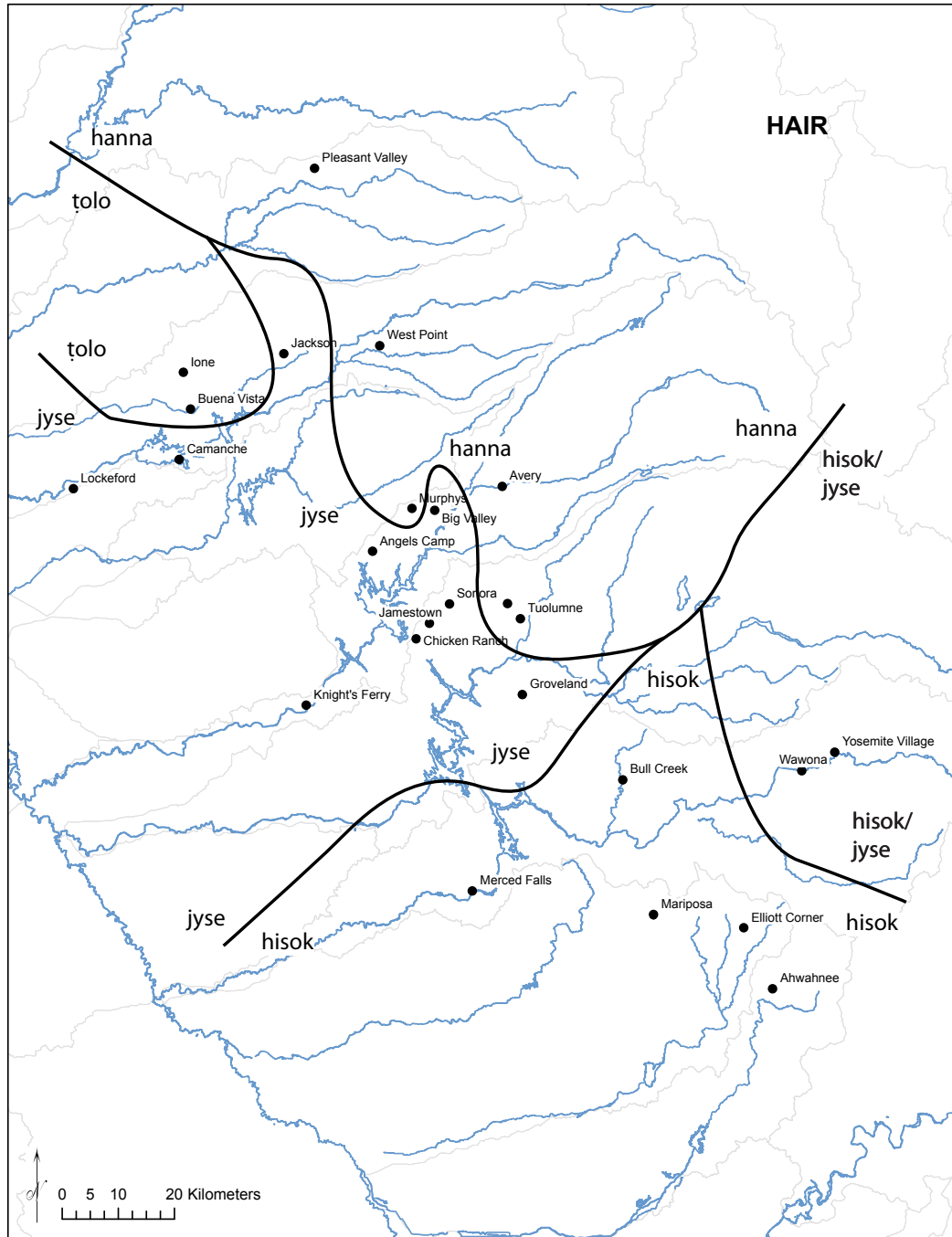


Figure 5.36: Eastern Miwok isogloss map: Hair

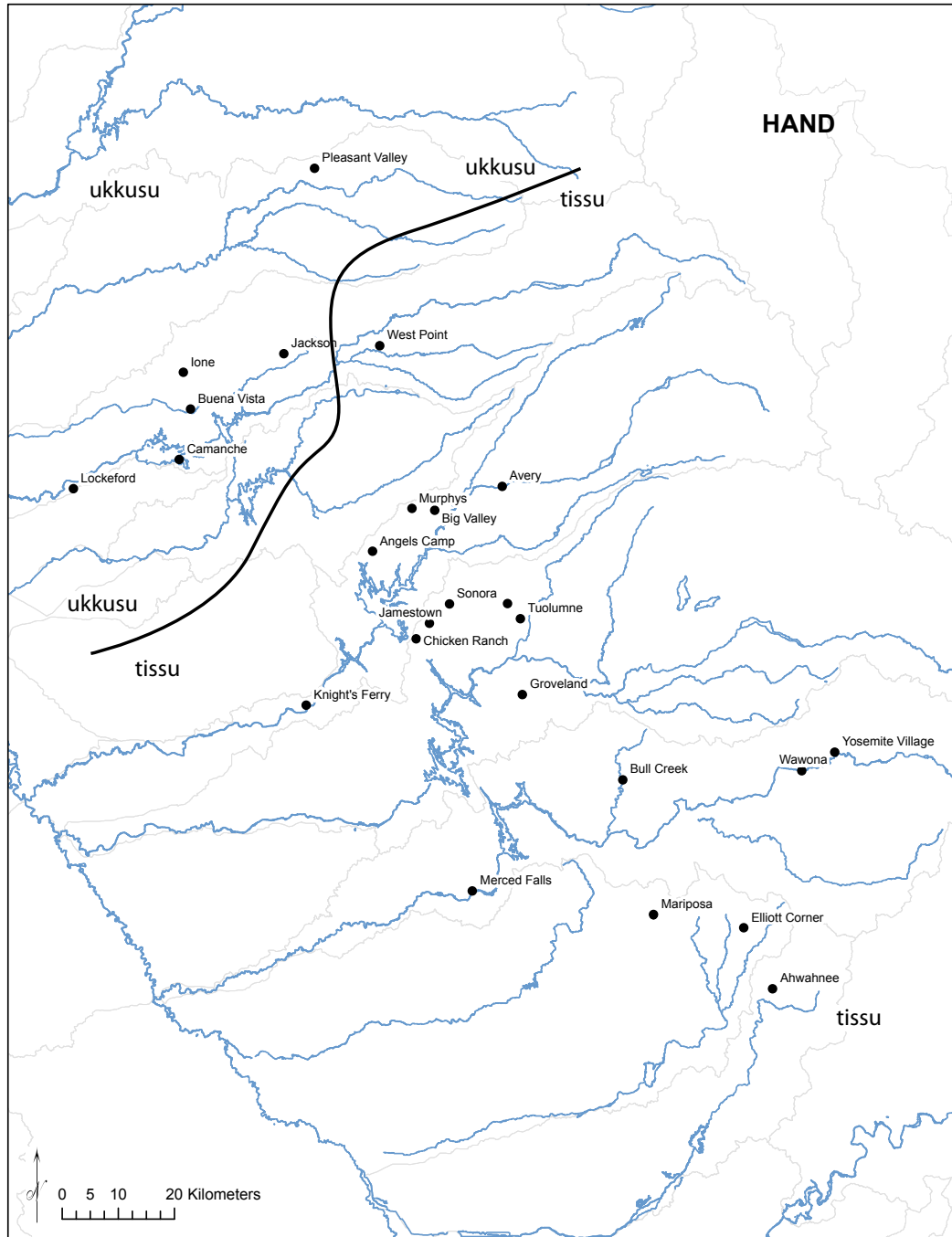


Figure 5.37: Eastern Miwok isogloss map: Hand

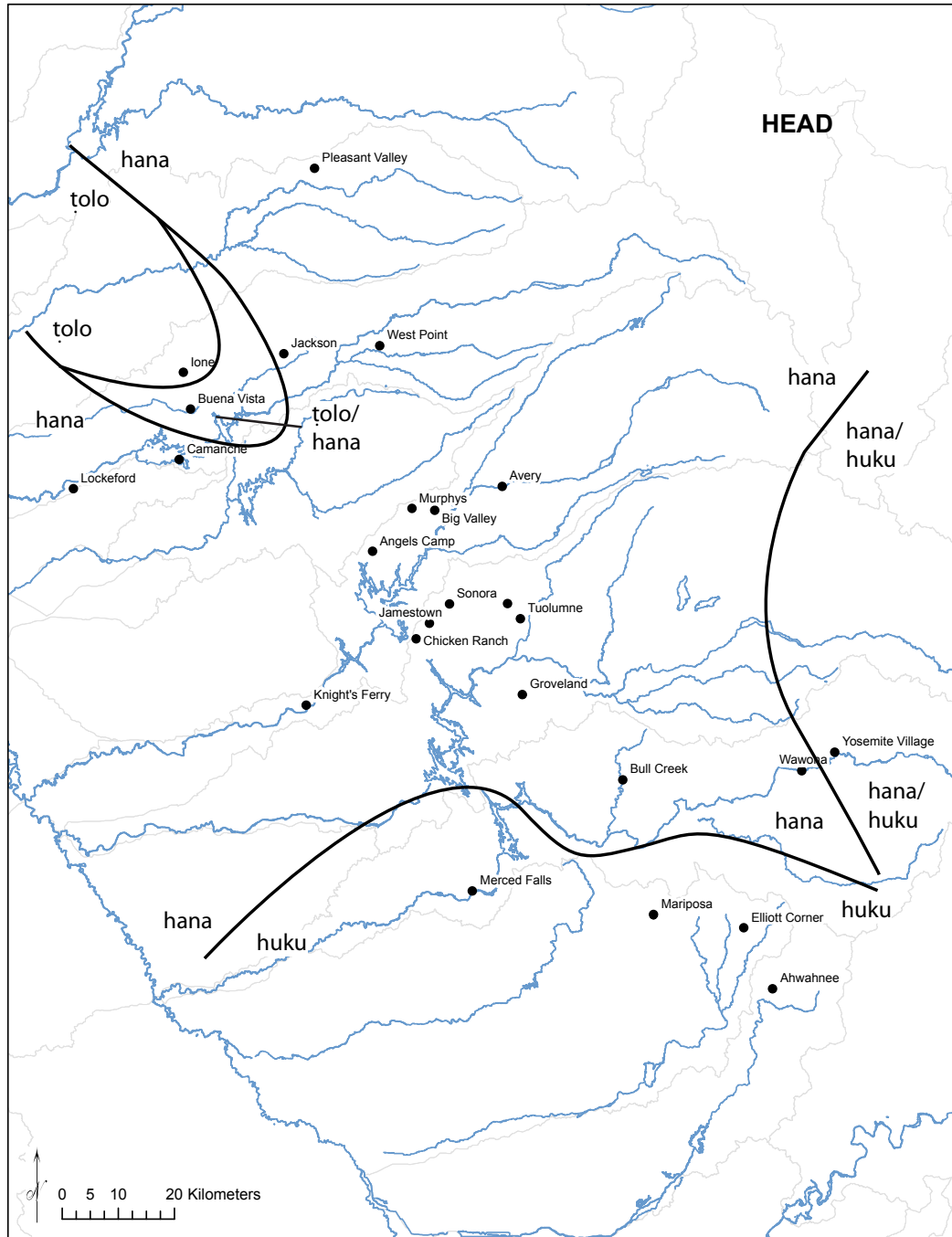


Figure 5.38: Eastern Miwok isogloss map: Head

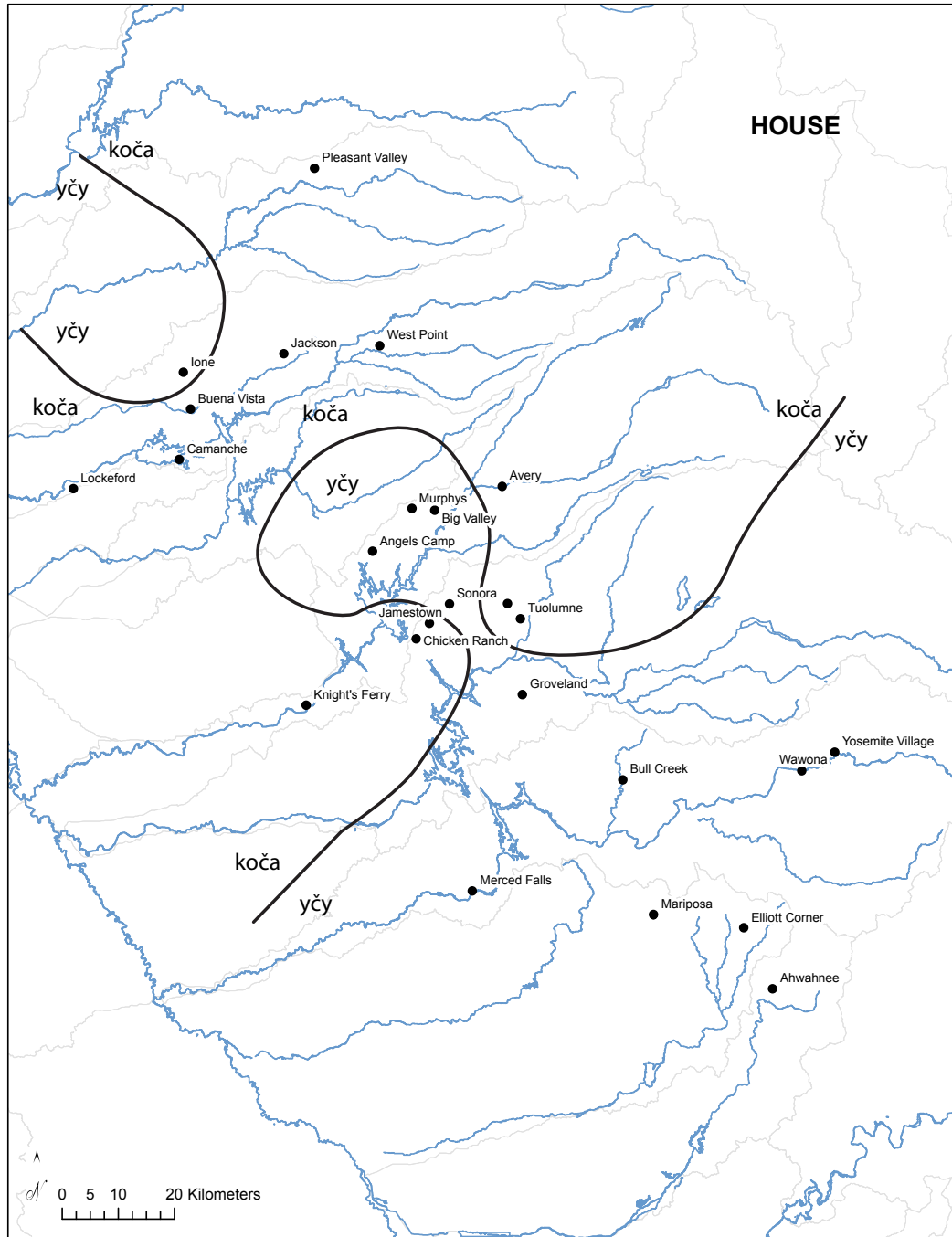


Figure 5.39: Eastern Miwok isogloss map: House

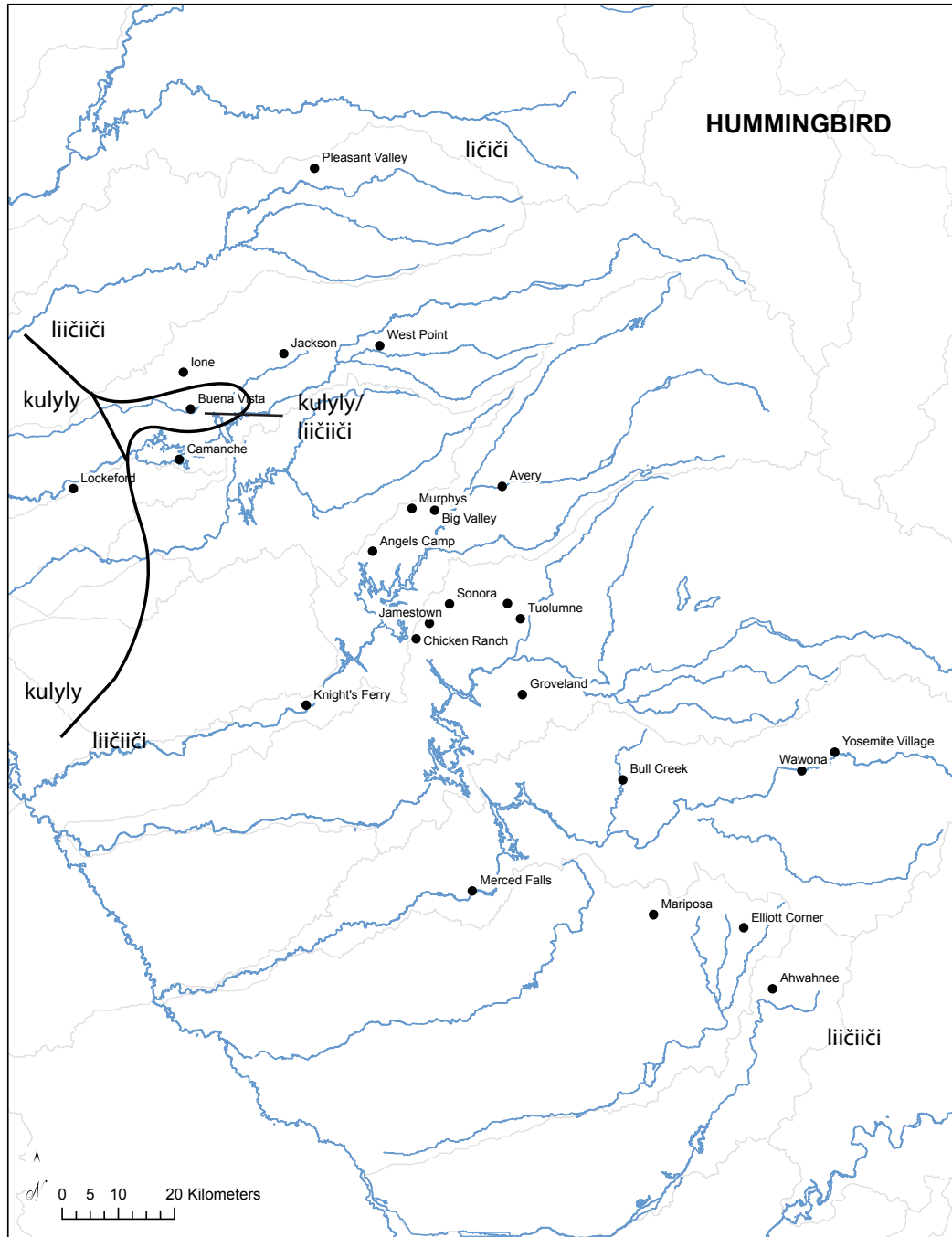


Figure 5.40: Eastern Miwok isogloss map: Hummingbird

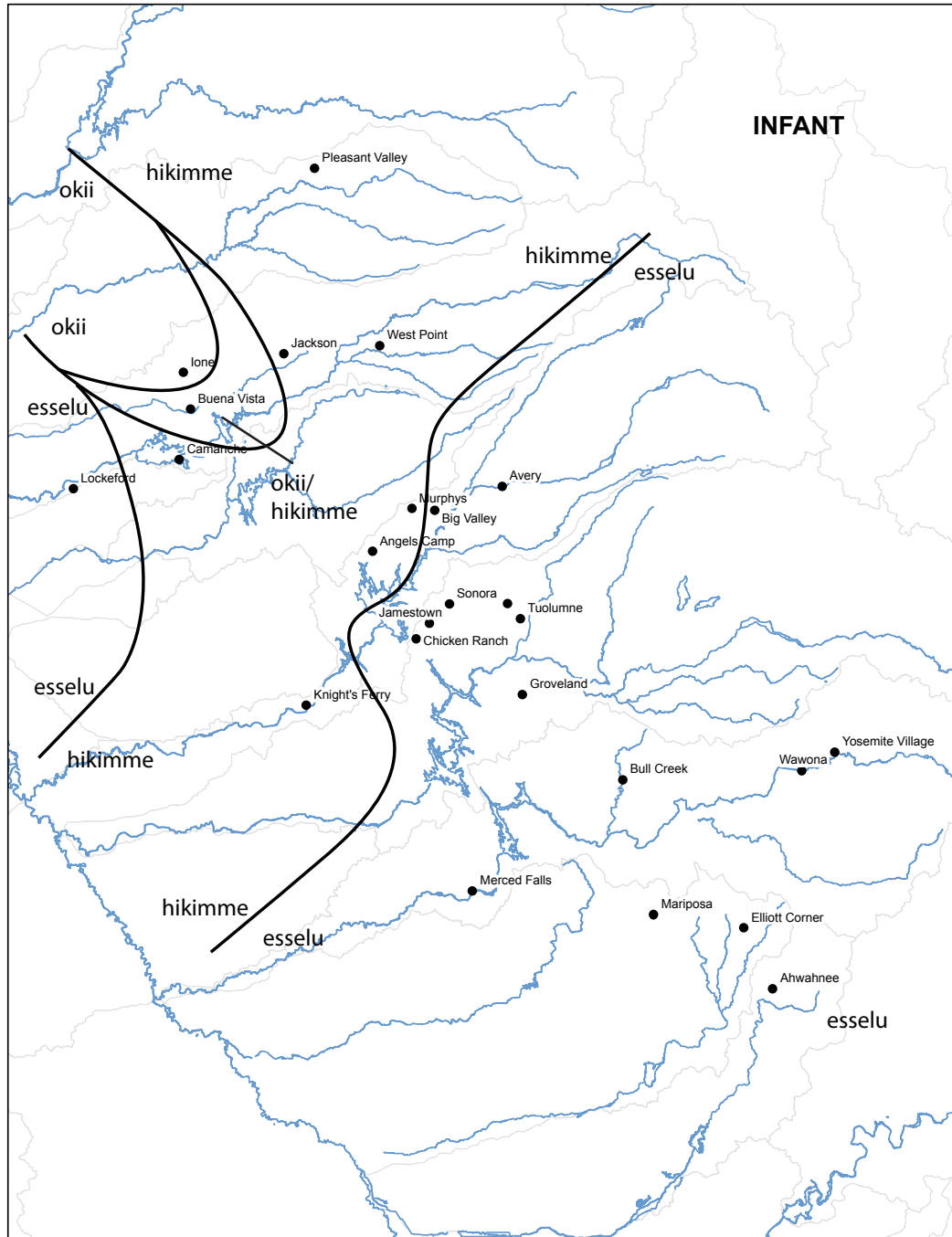


Figure 5.41: Eastern Miwok isogloss map: Infant

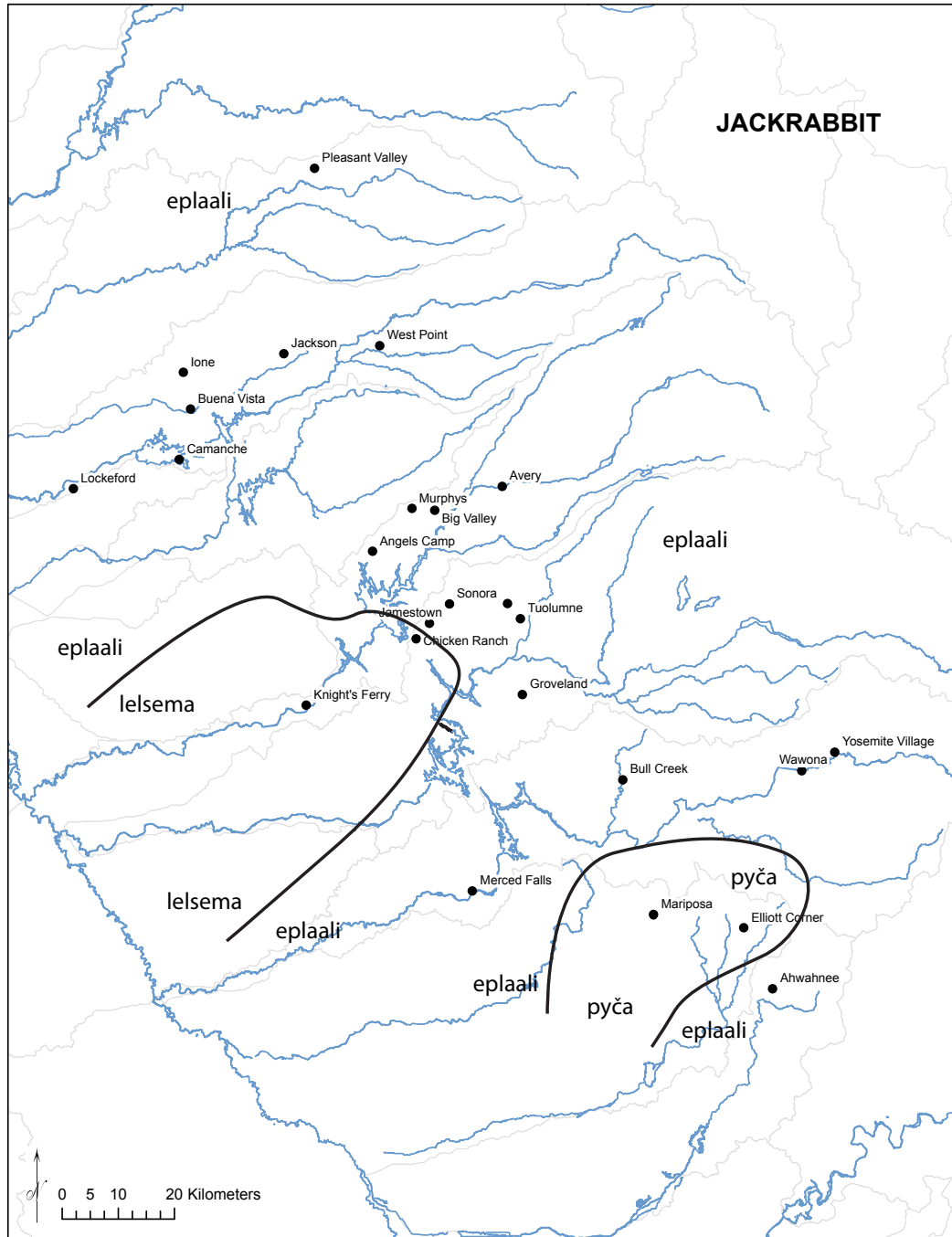


Figure 5.42: Eastern Miwok isogloss map: Jackrabbbit

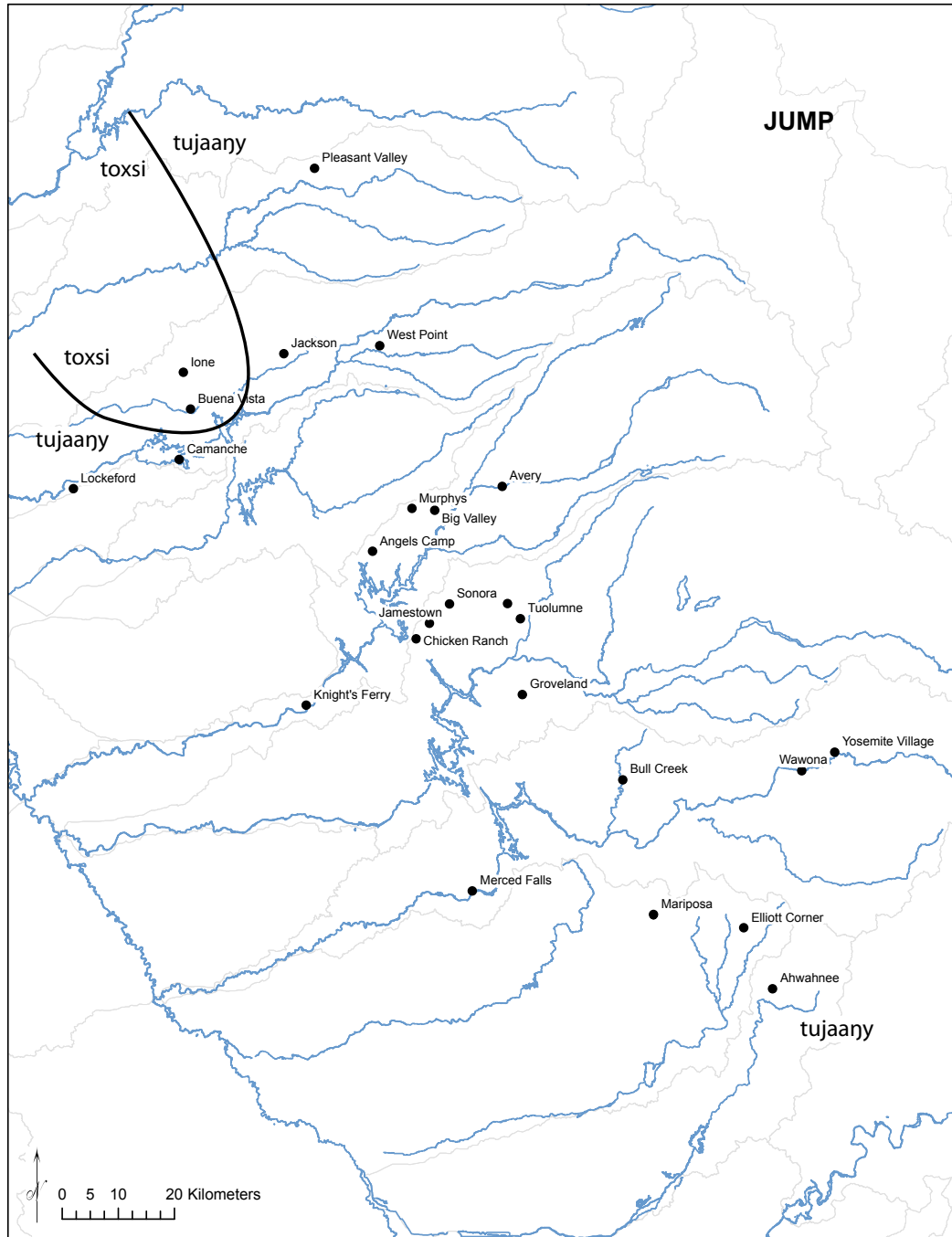


Figure 5.43: Eastern Miwok isogloss map: Jump

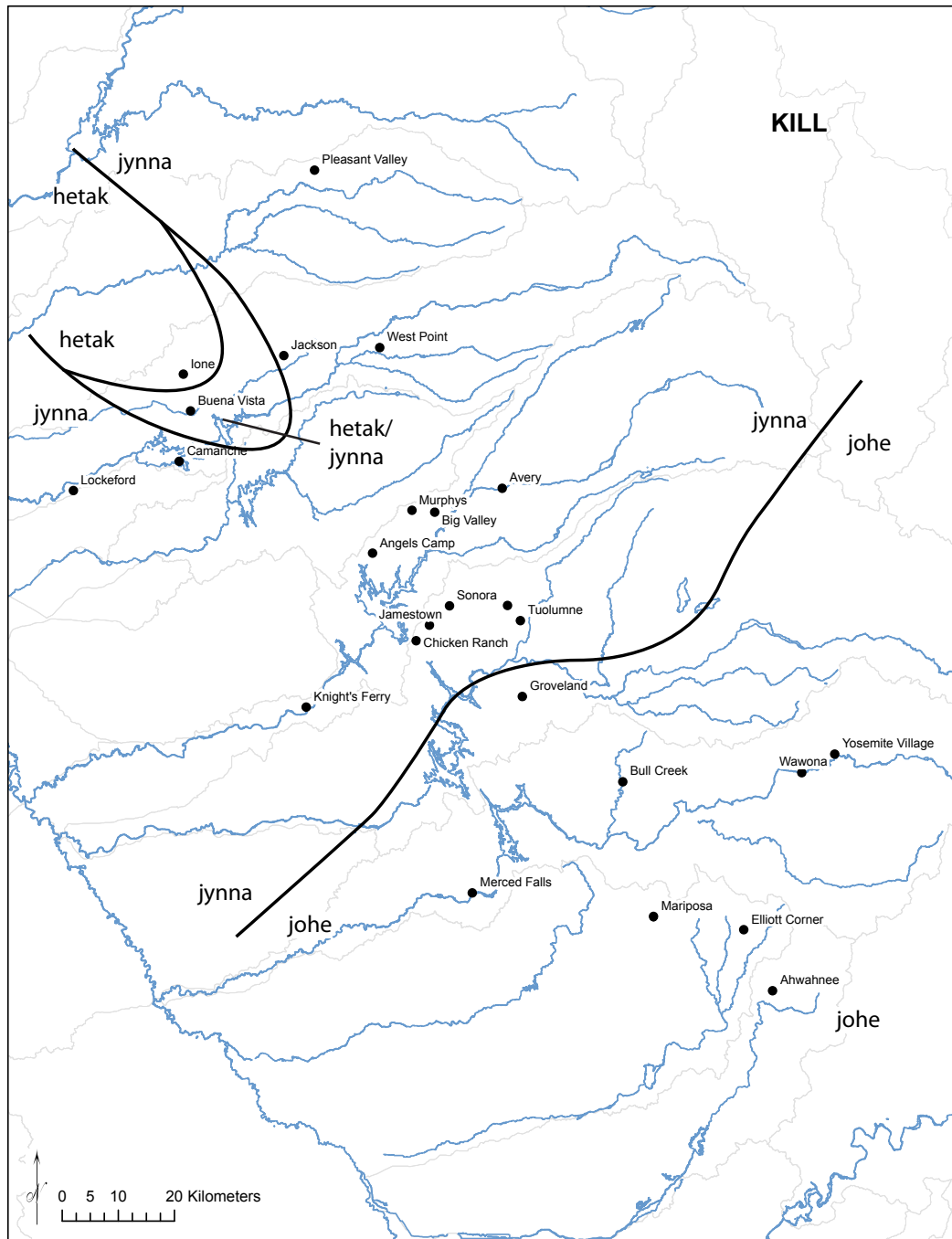


Figure 5.44: Eastern Miwok isogloss map: Kill

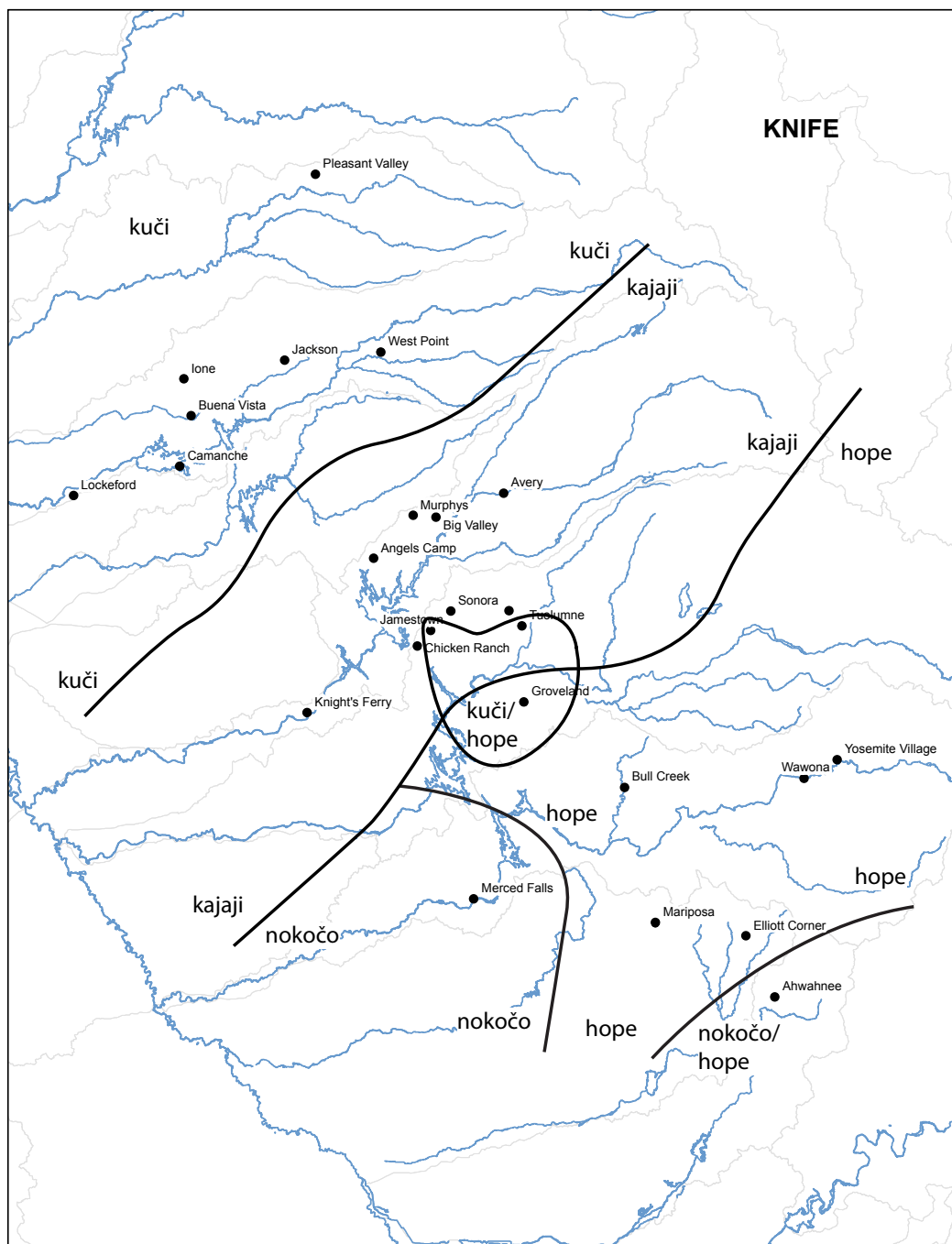


Figure 5.45: Eastern Miwok isogloss map: Knife

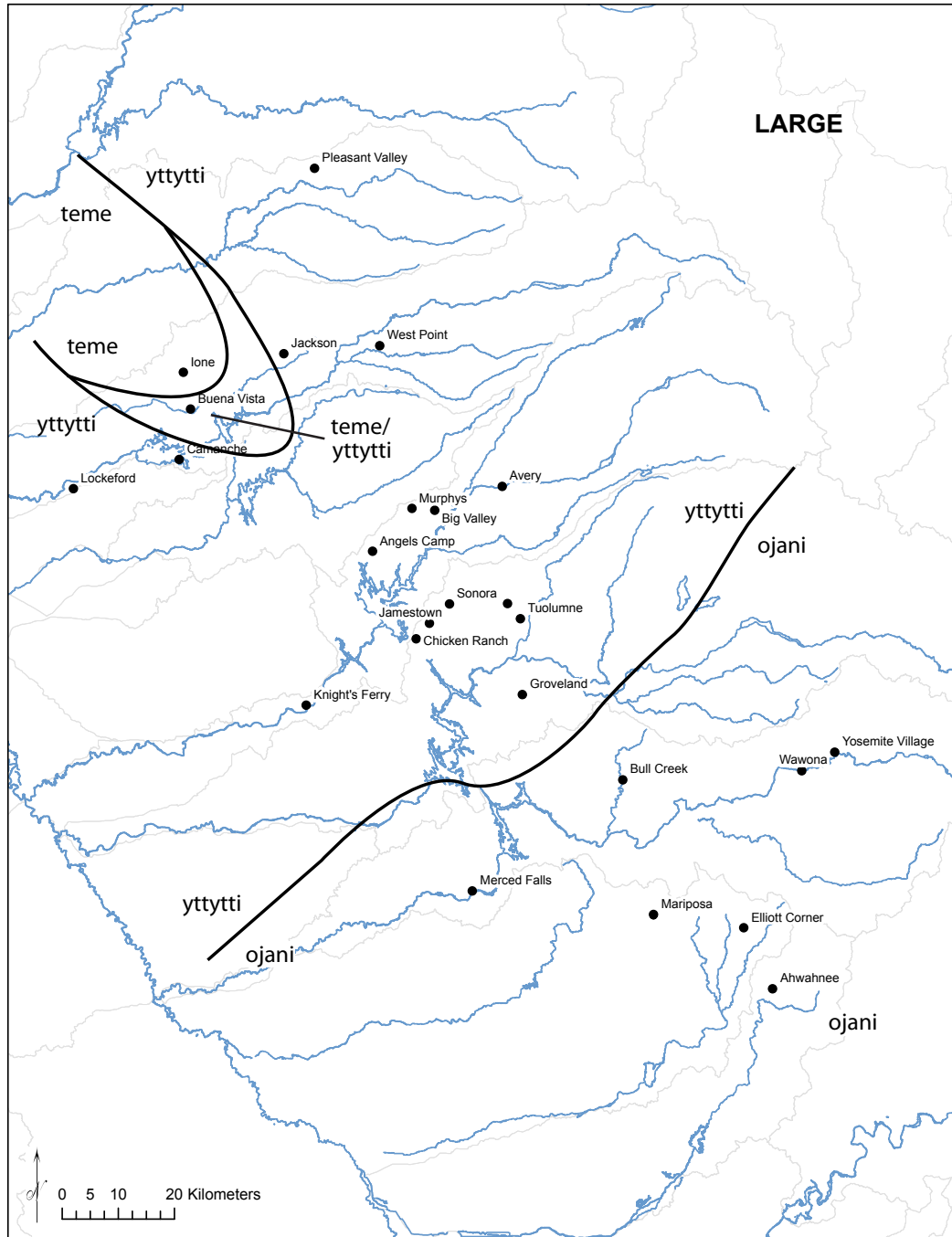


Figure 5.46: Eastern Miwok isogloss map: Large

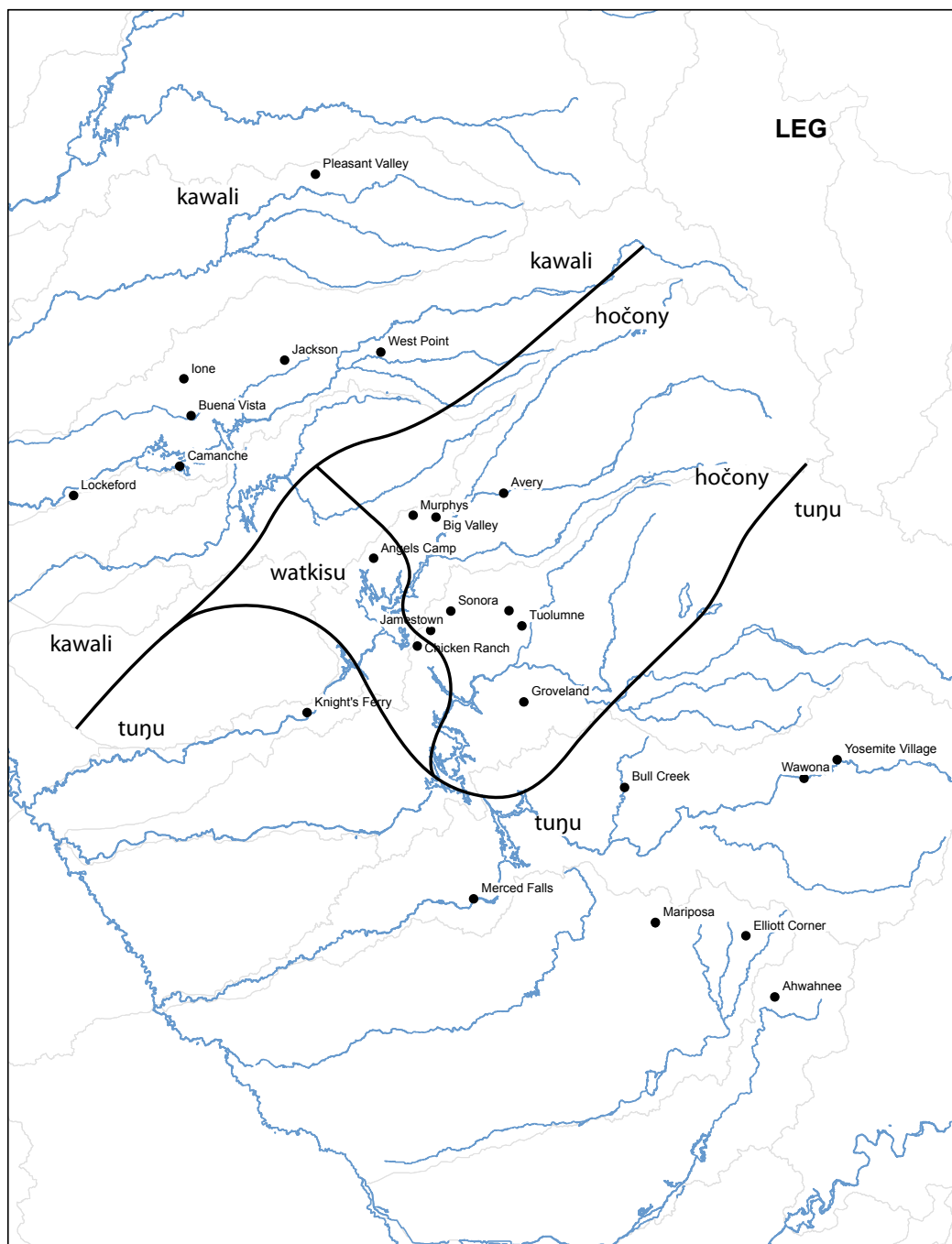


Figure 5.47: Eastern Miwok isogloss map: Leg

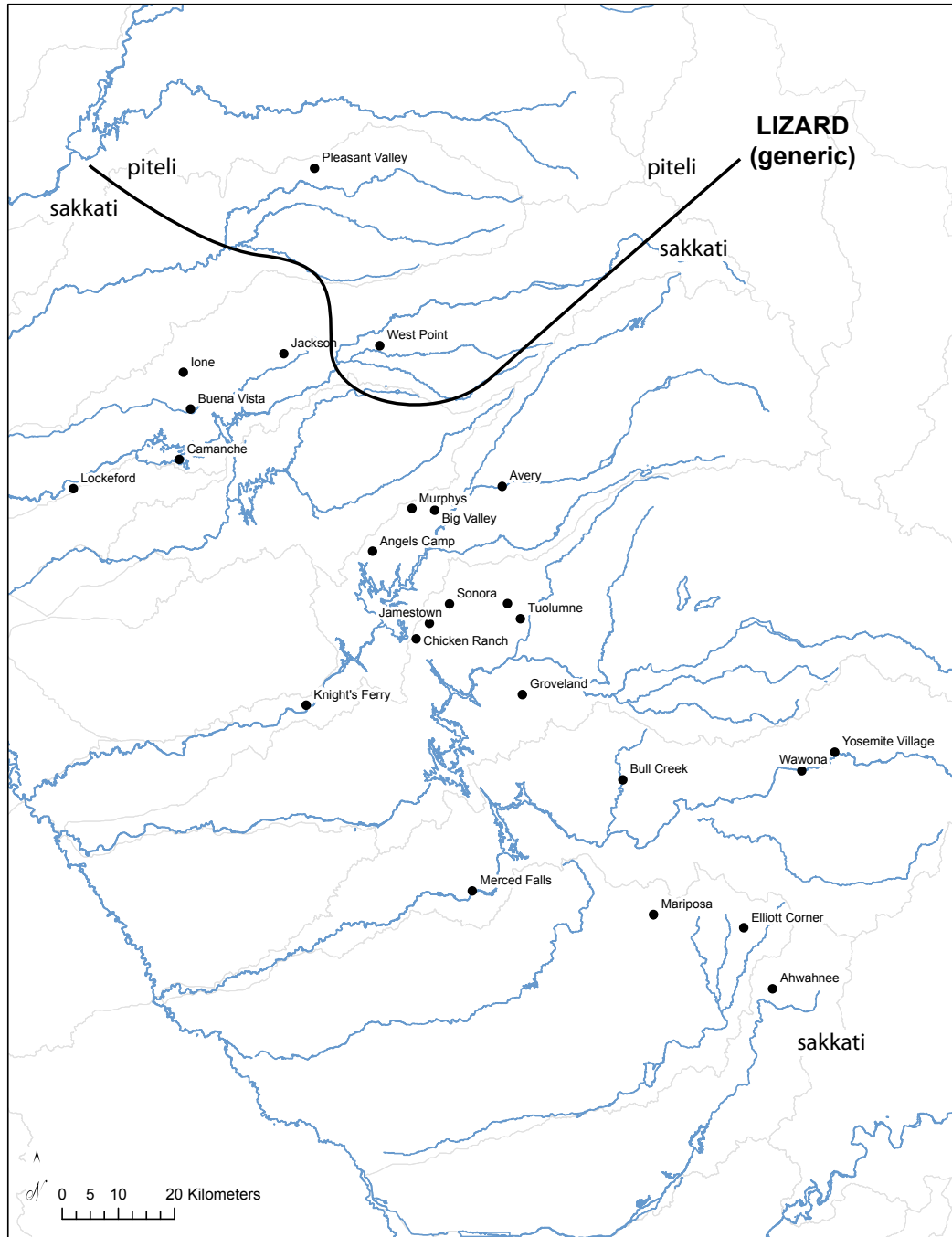


Figure 5.48: Eastern Miwok isogloss map: Lizard (generic)

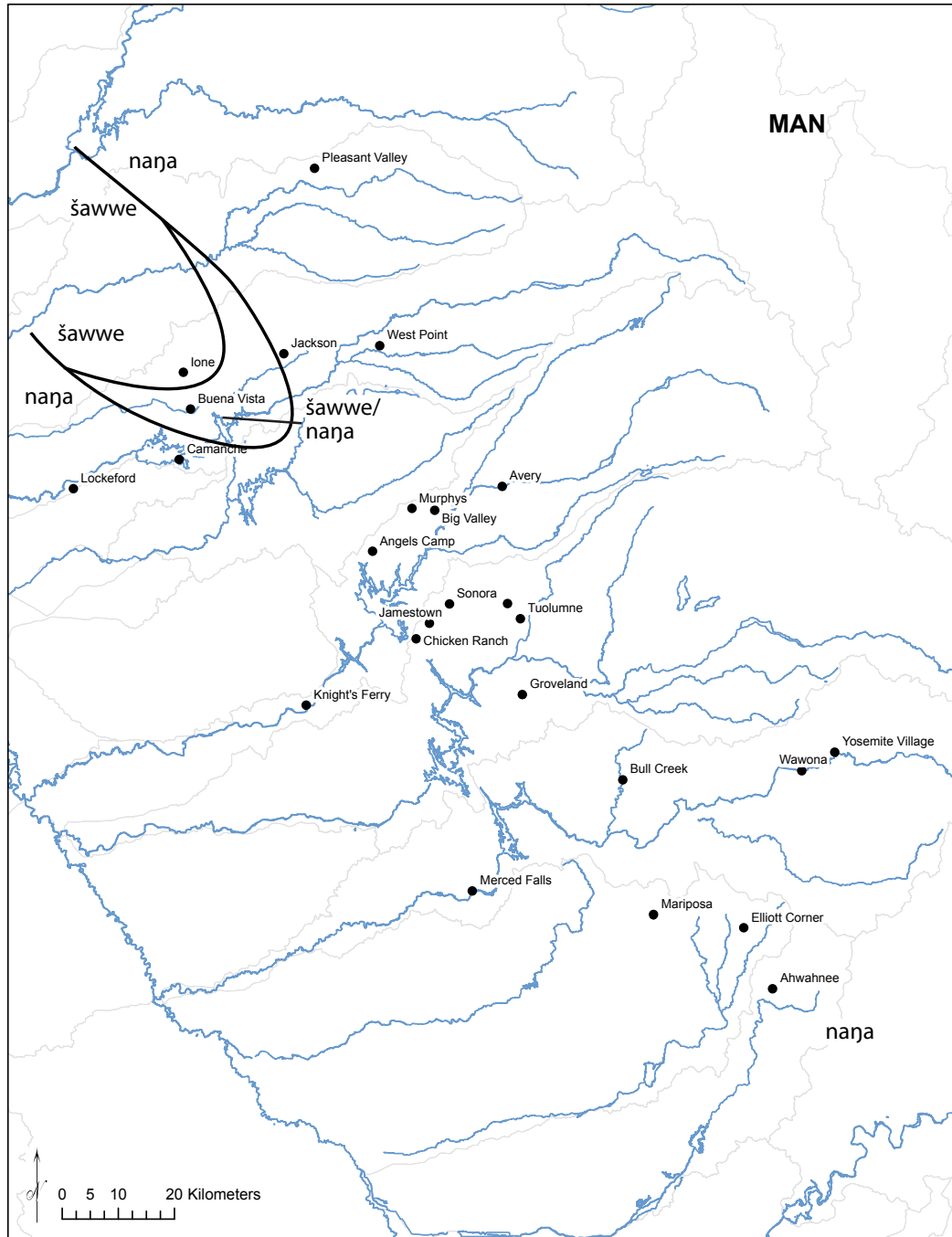


Figure 5.49: Eastern Miwok isogloss map: Man

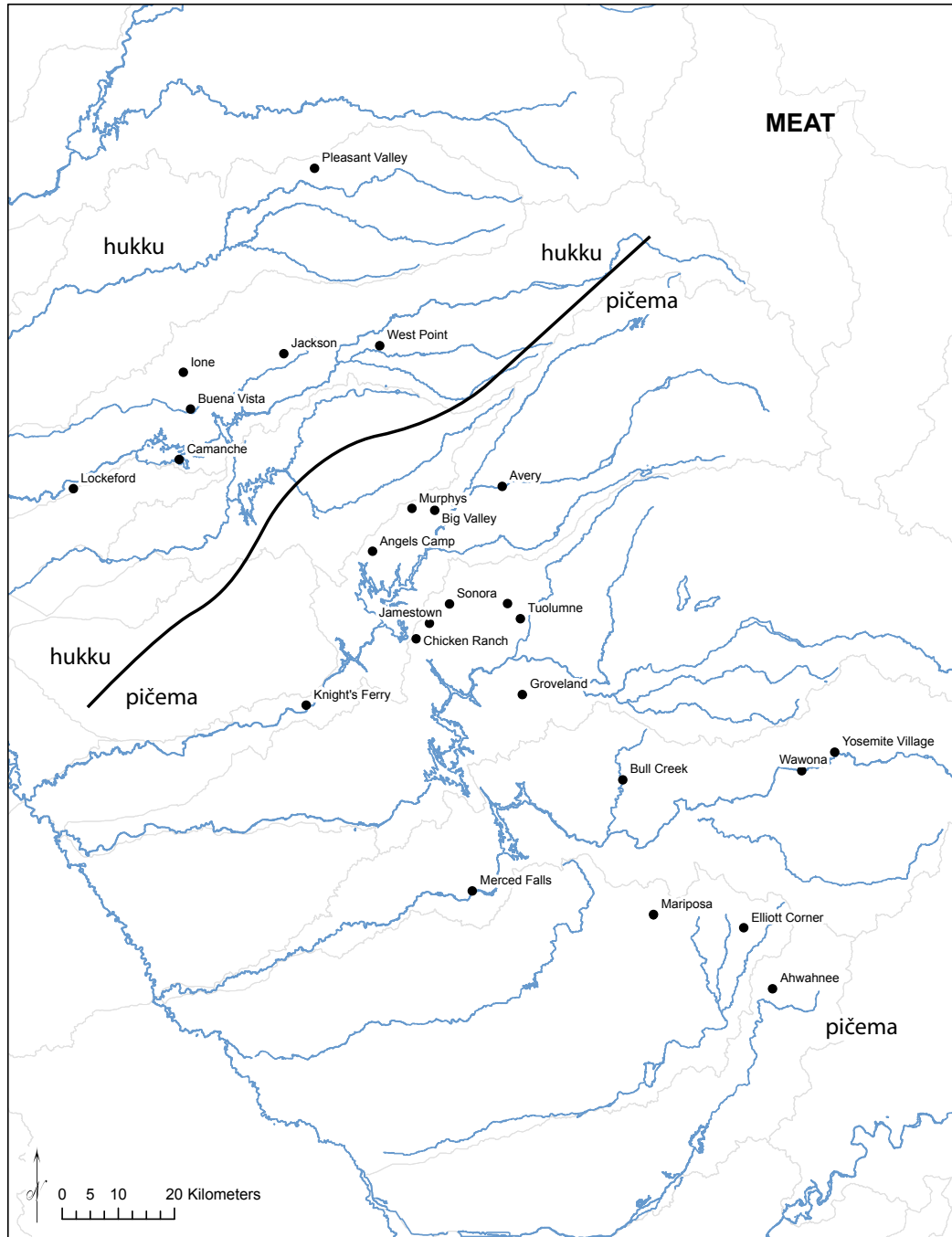


Figure 5.50: Eastern Miwok isogloss map: Meat

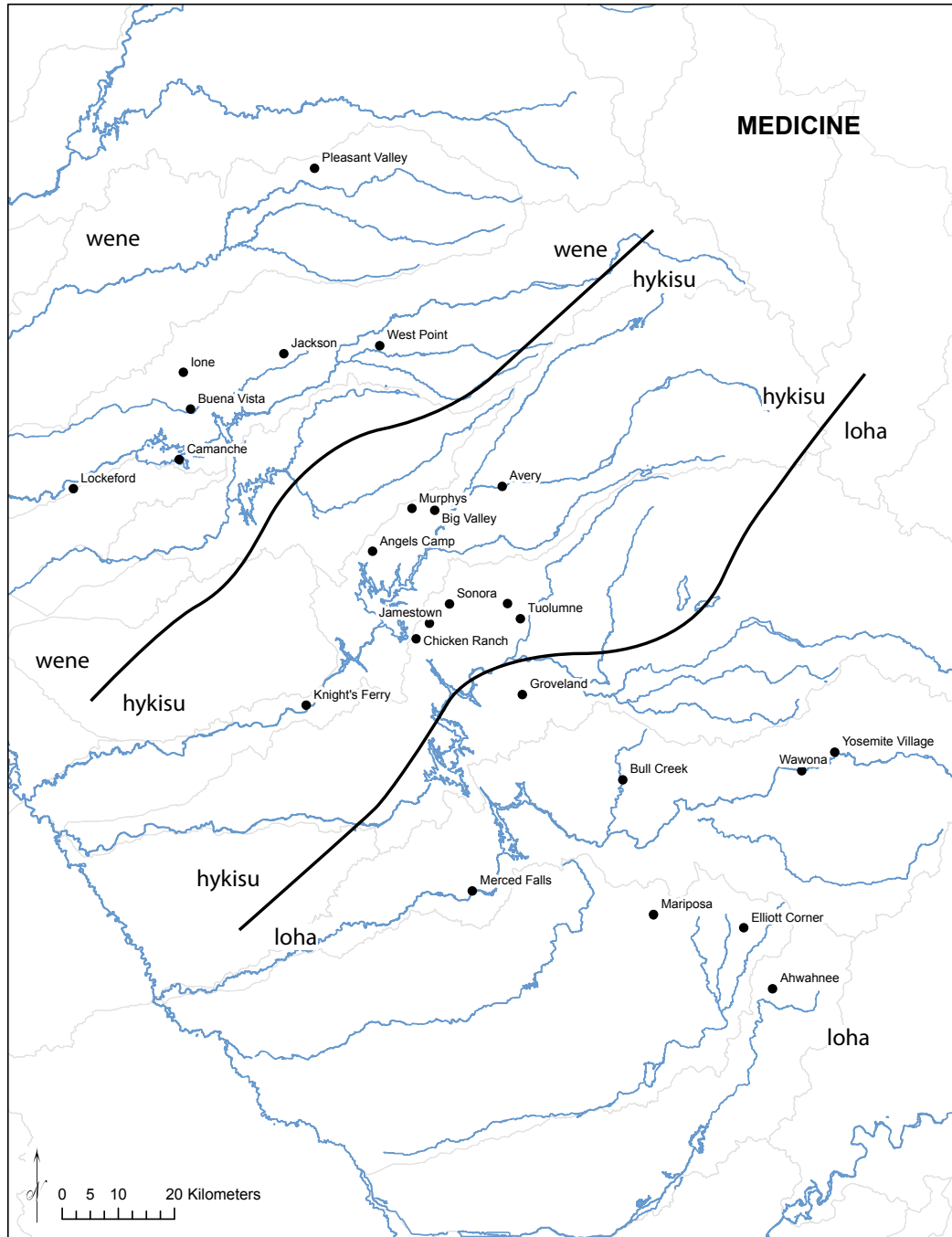


Figure 5.51: Eastern Miwok isogloss map: Medicine

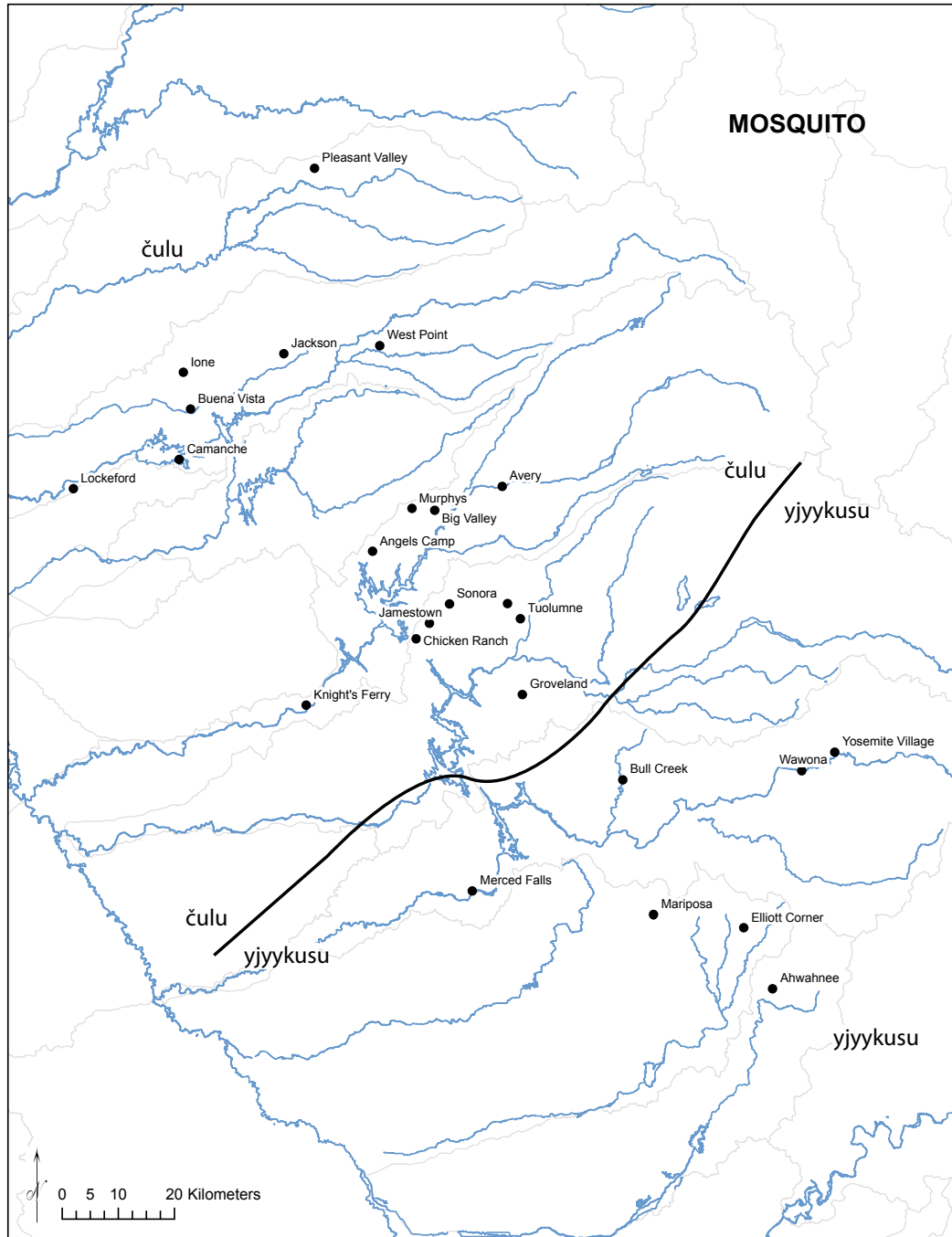


Figure 5.52: Eastern Miwok isogloss map: Mosquito

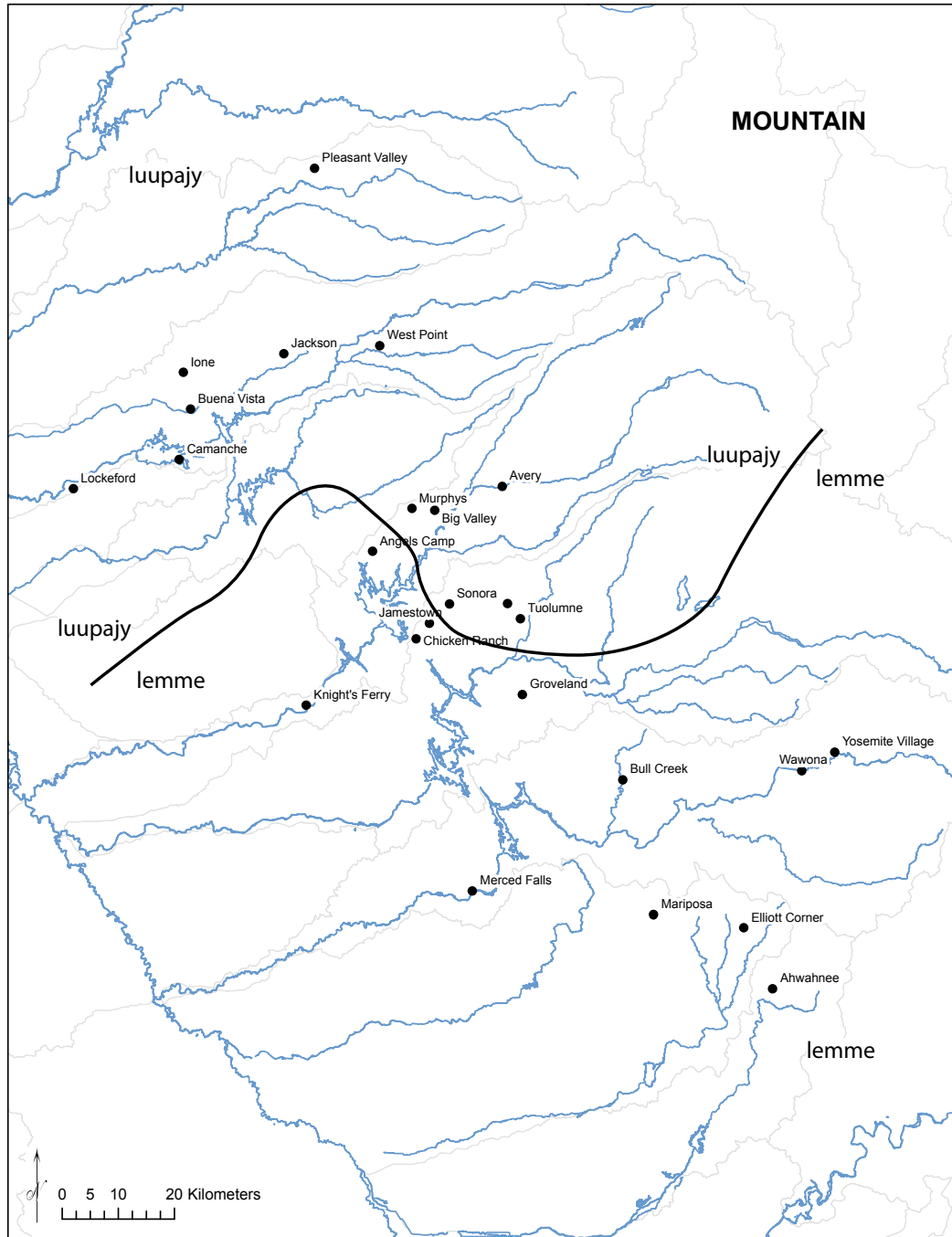


Figure 5.53: Eastern Miwok isogloss map: Mountain

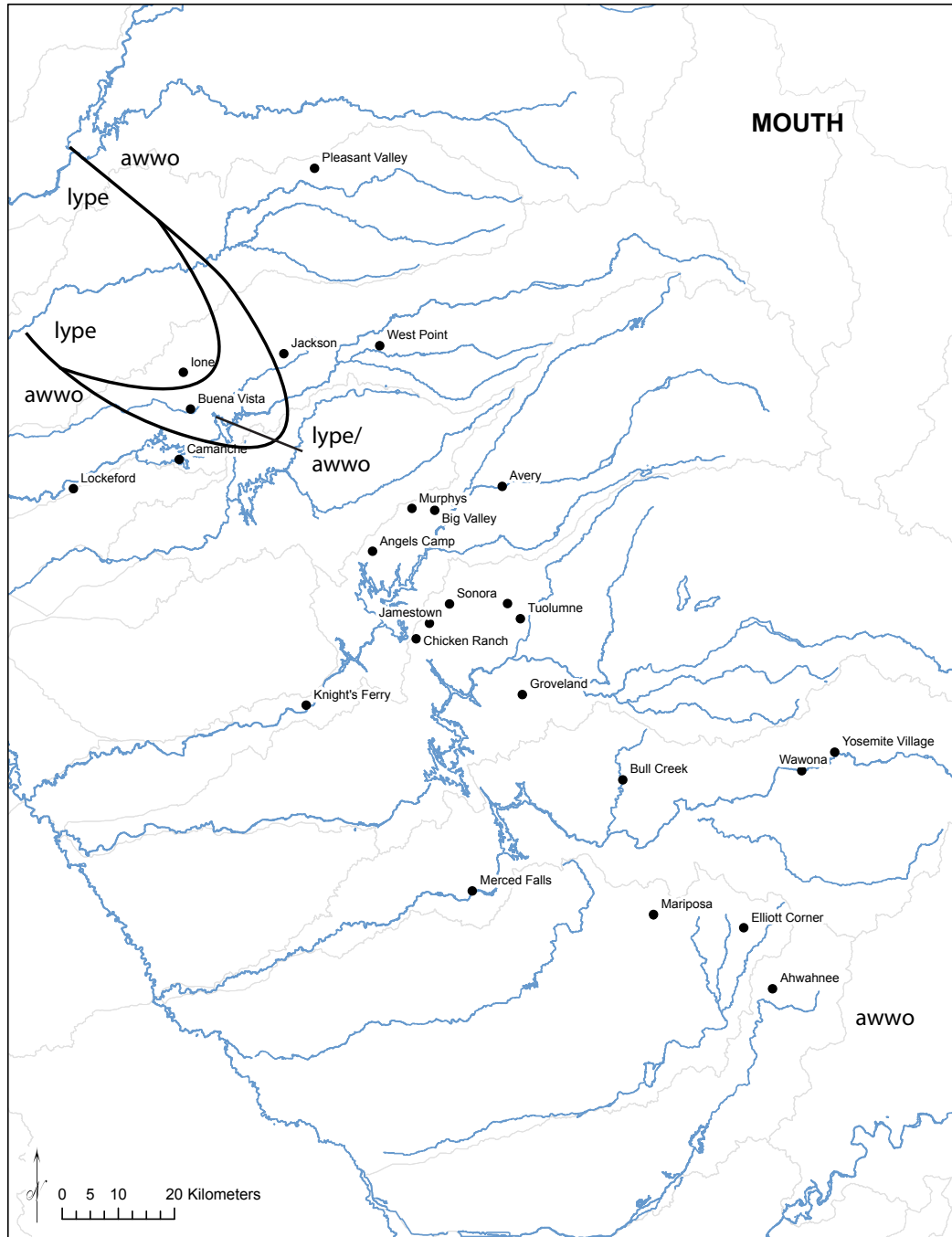


Figure 5.54: Eastern Miwok isogloss map: Mouth

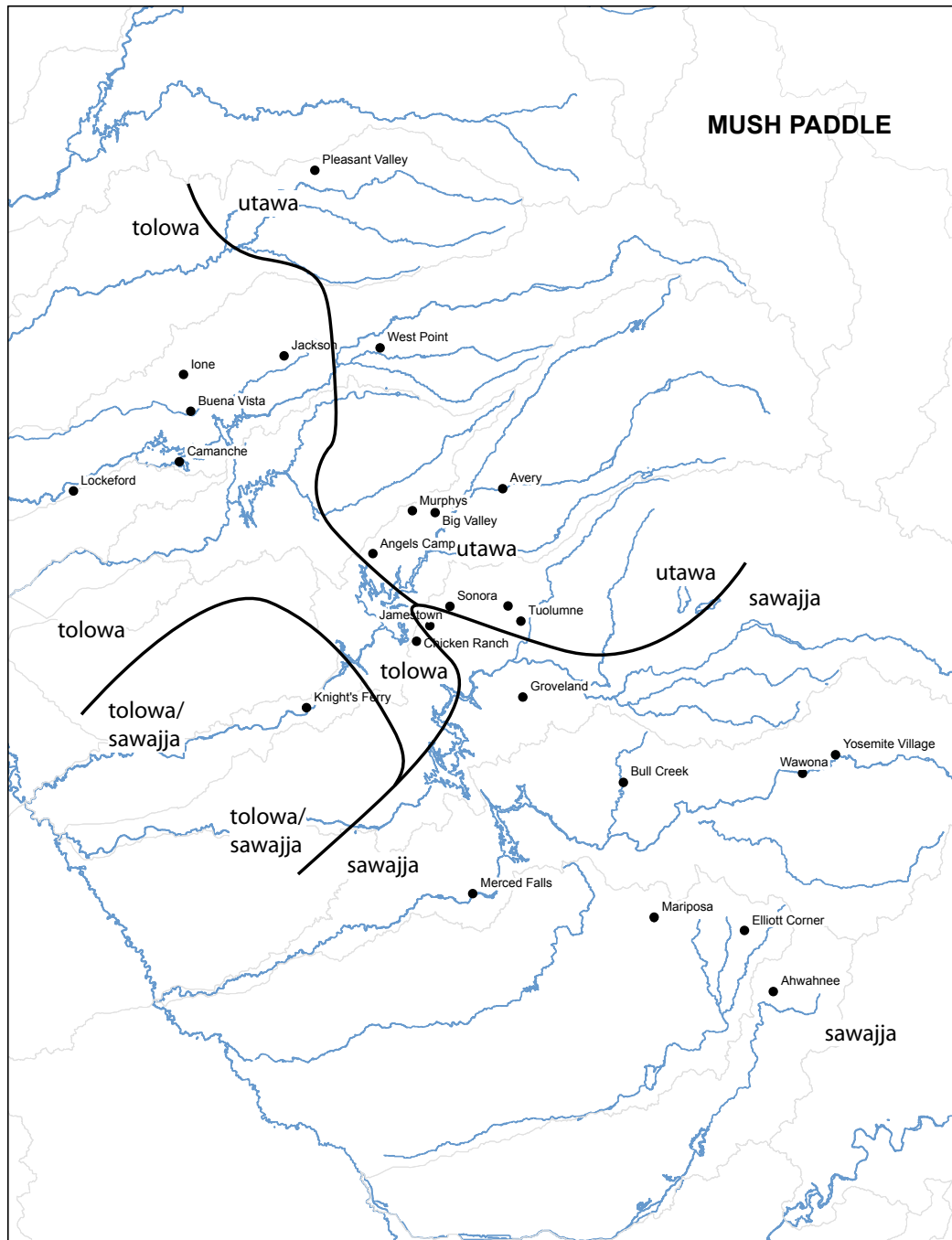


Figure 5.55: Eastern Miwok isogloss map: Mush Paddle

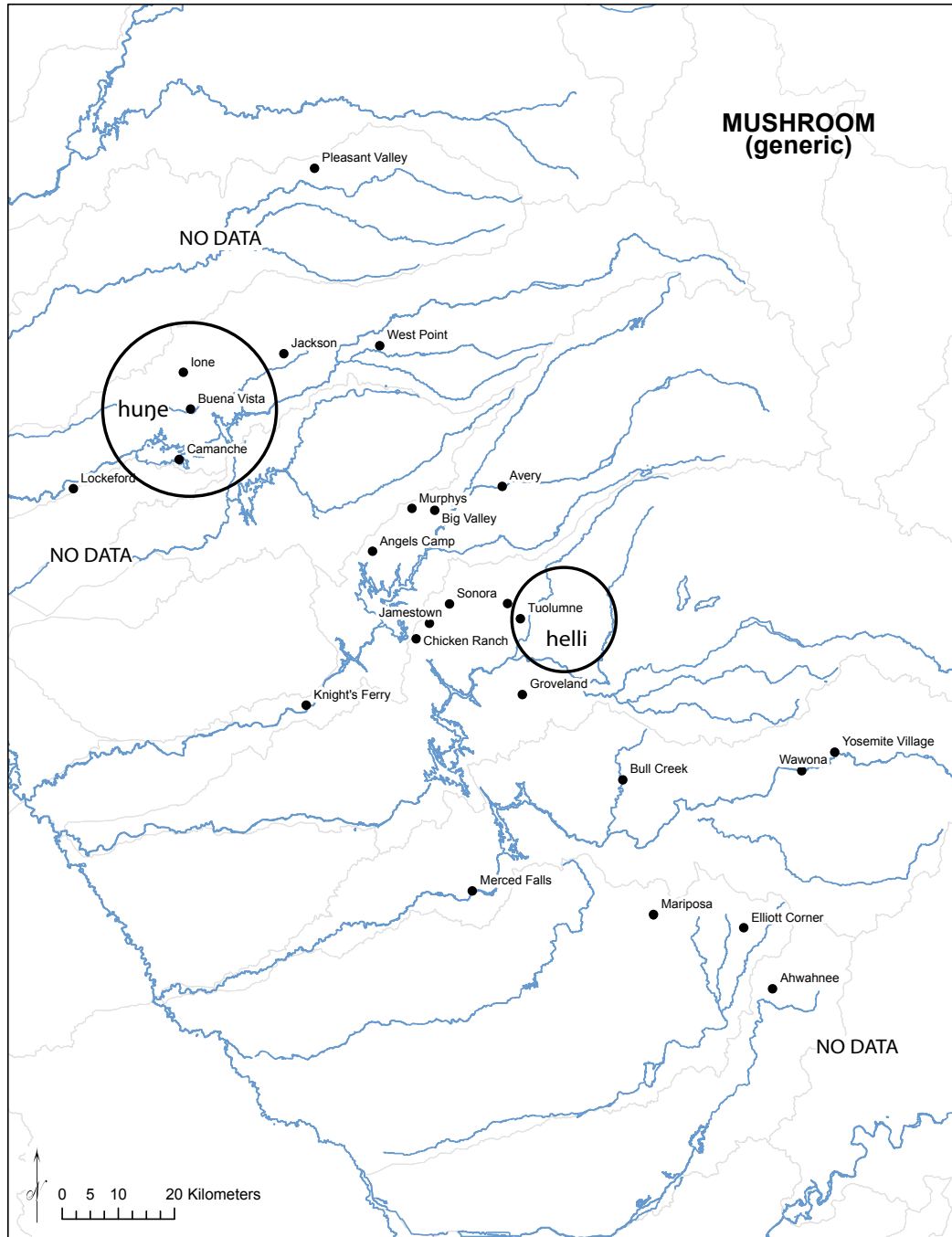


Figure 5.56: Eastern Miwok isogloss map: Mushroom (generic)

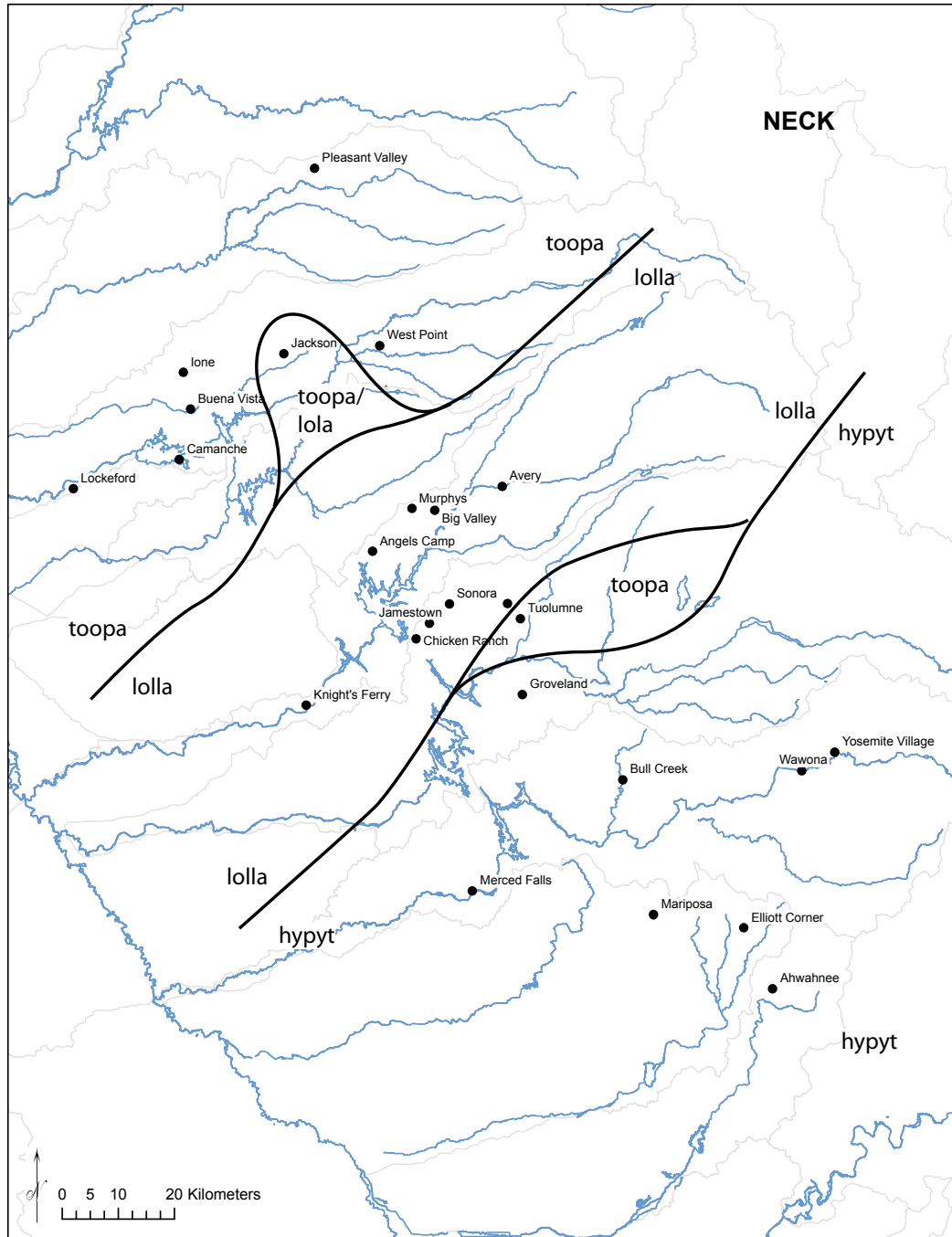


Figure 5.57: Eastern Miwok isogloss map: Neck

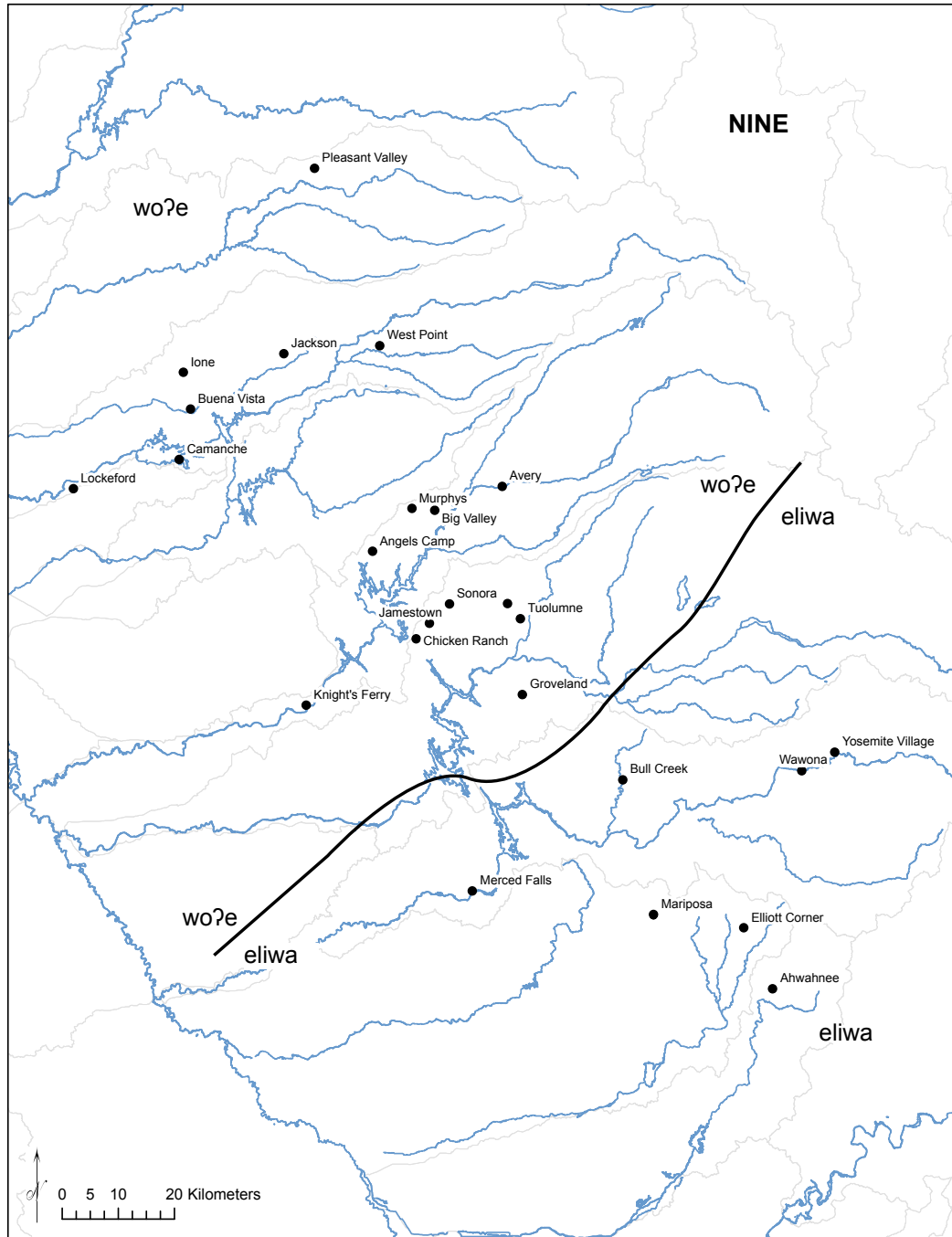


Figure 5.58: Eastern Miwok isogloss map: Nine

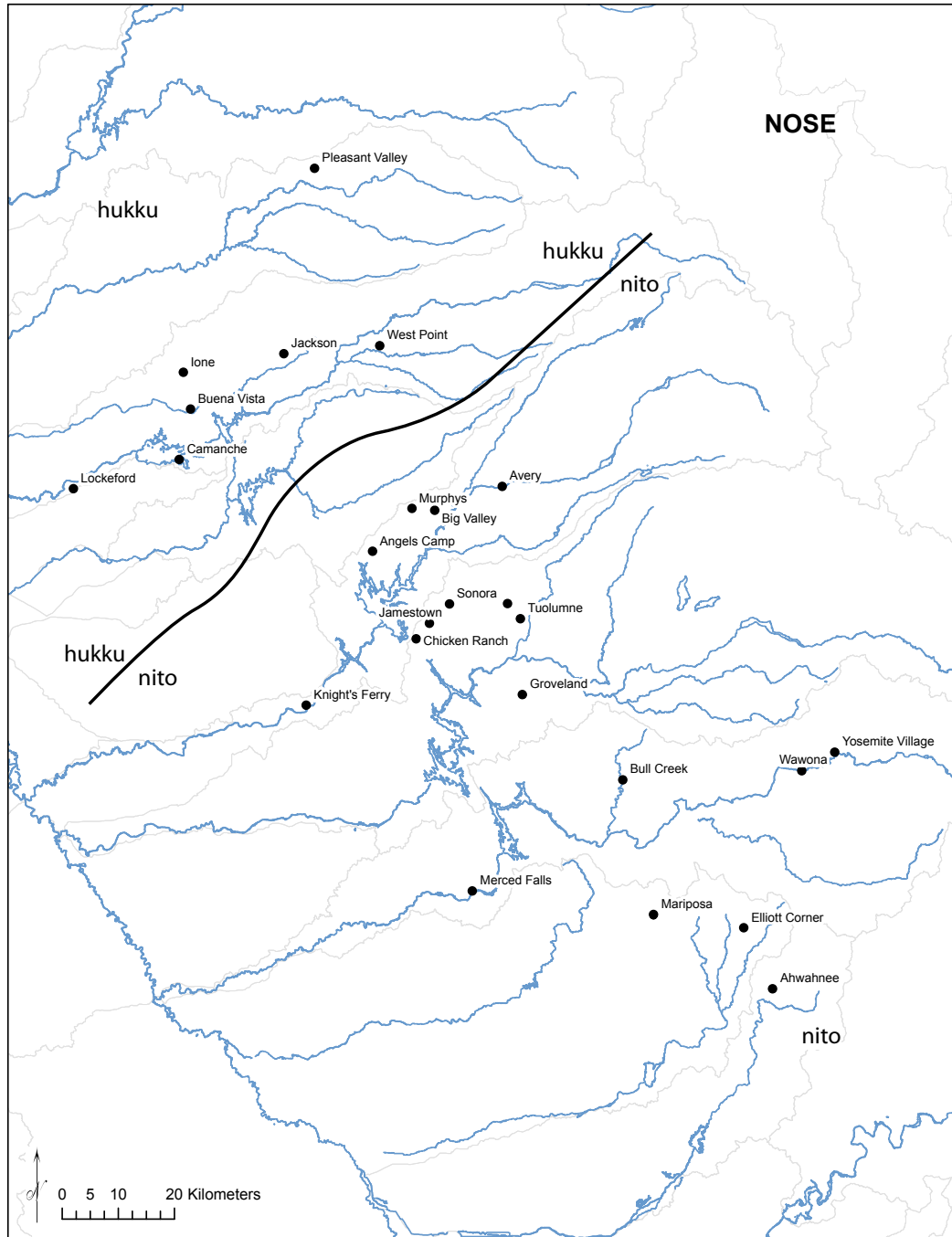


Figure 5.59: Eastern Miwok isogloss map: Nose

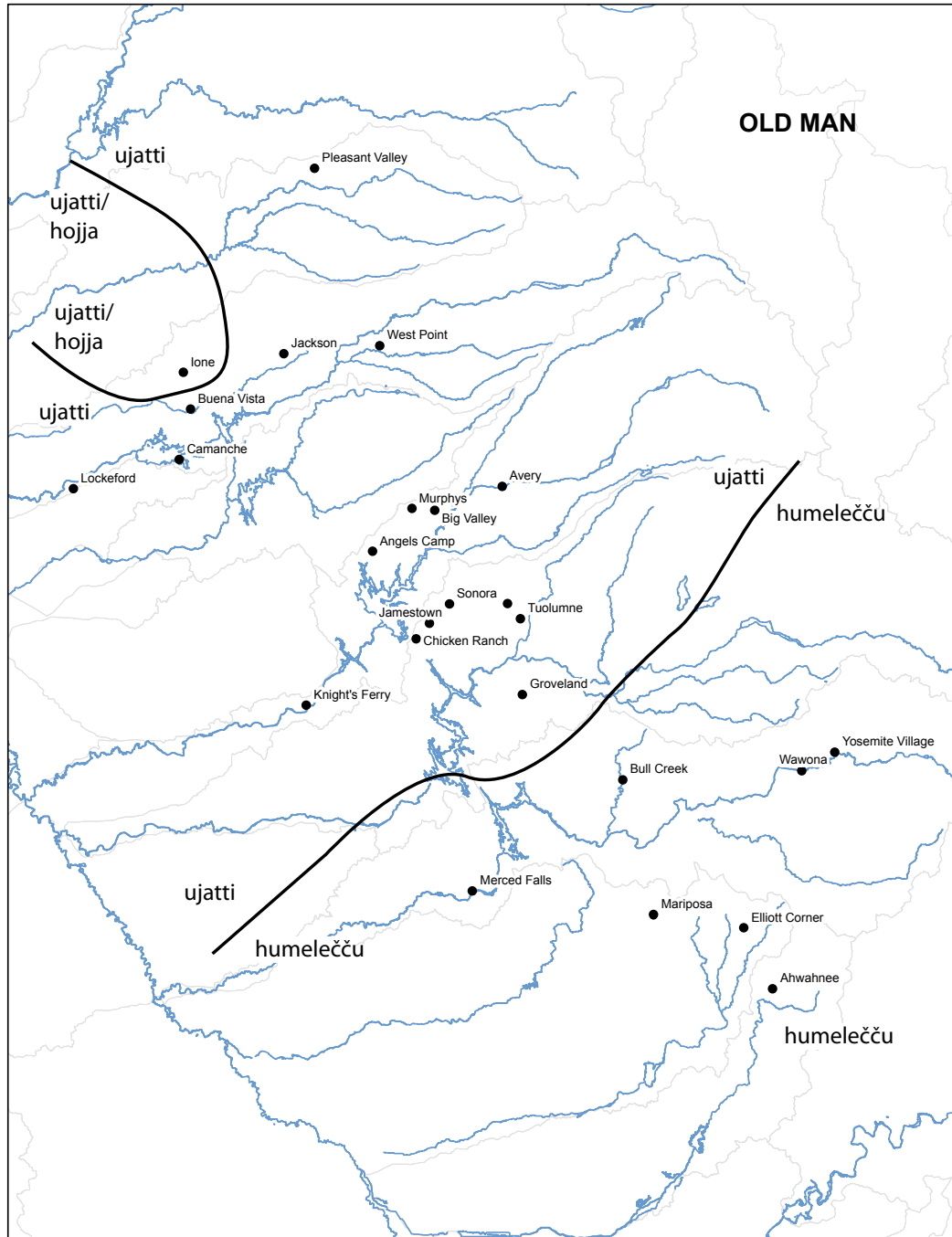


Figure 5.60: Eastern Miwok isogloss map: Old Man

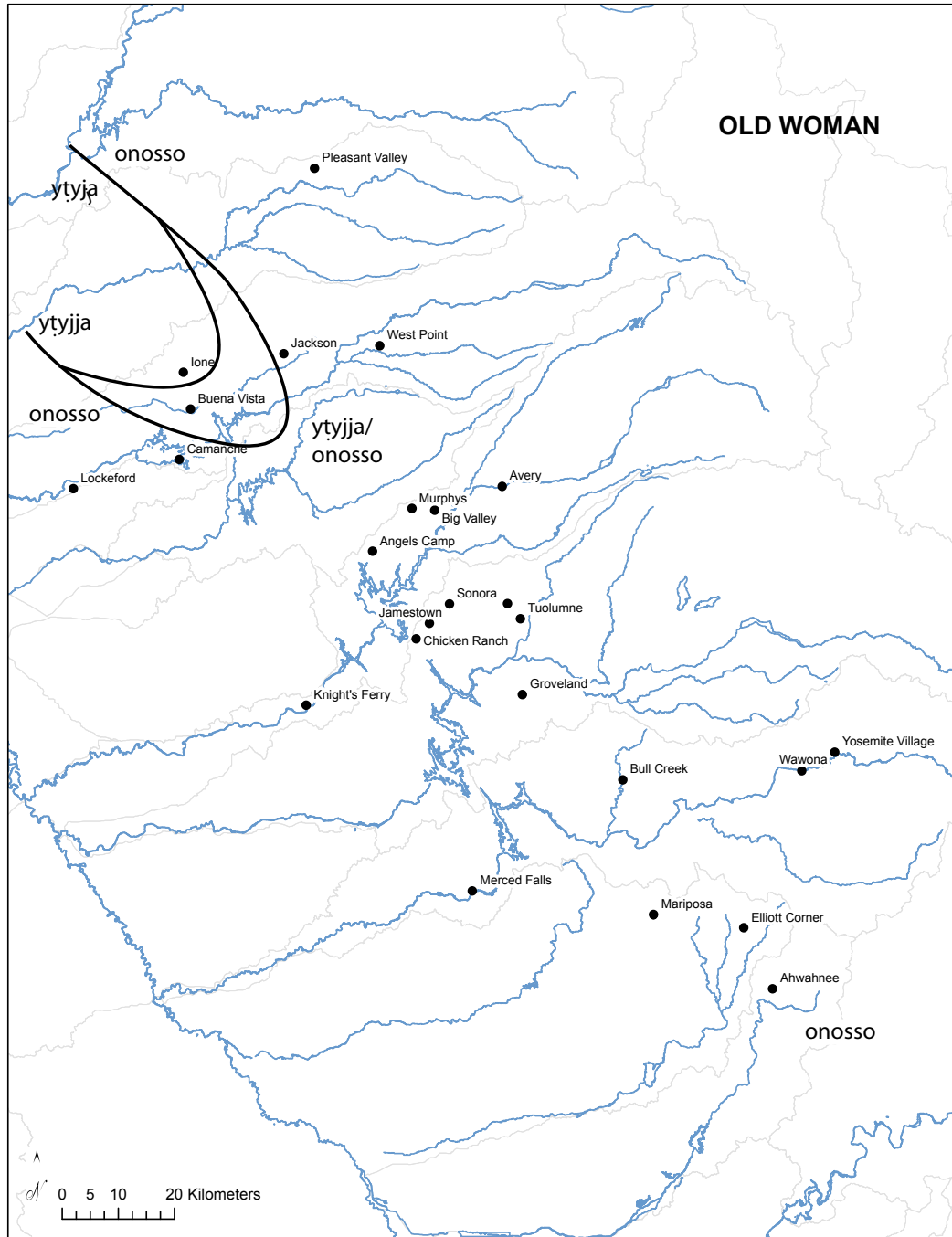


Figure 5.61: Eastern Miwok isogloss map: Old Woman

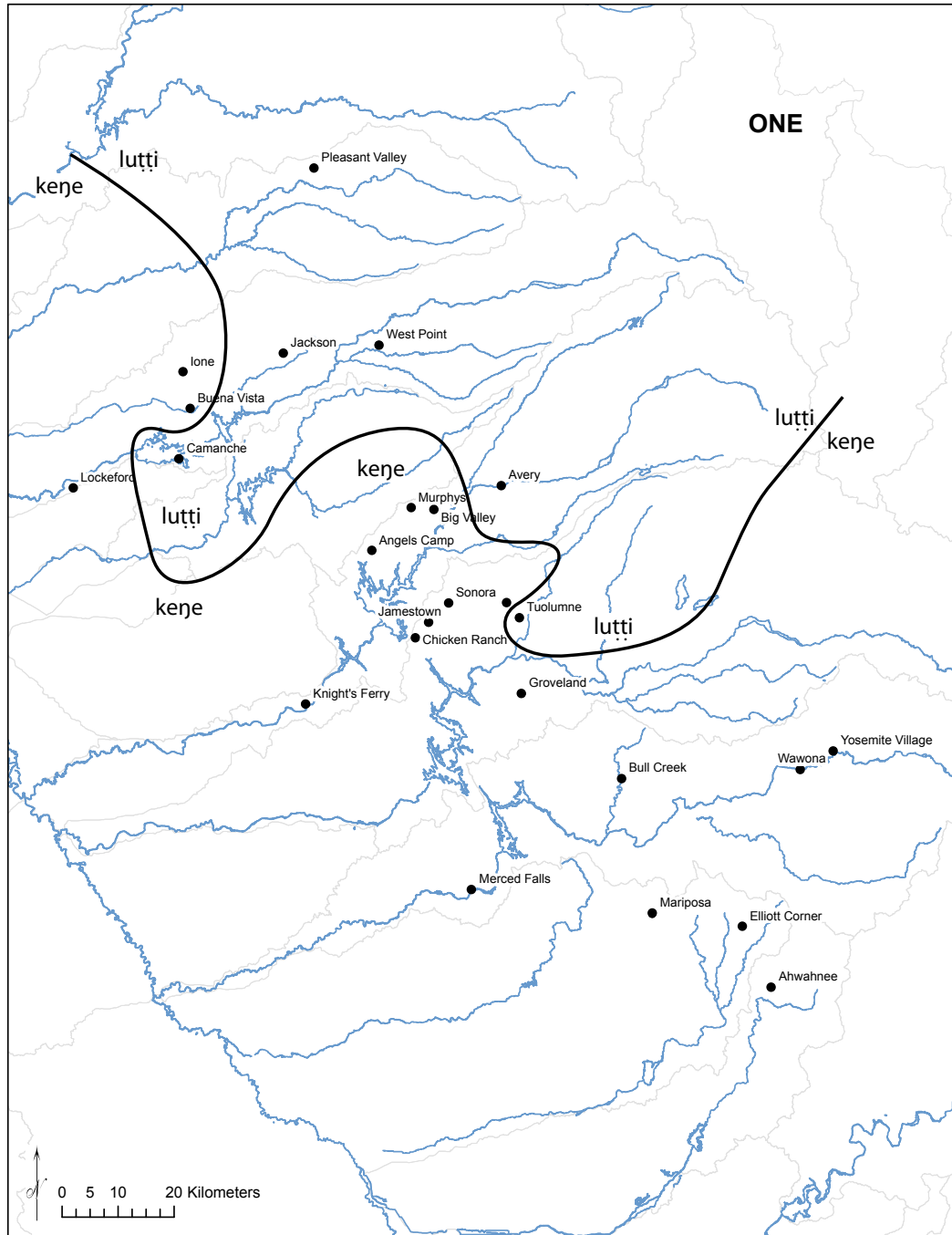


Figure 5.62: Eastern Miwok isogloss map: One

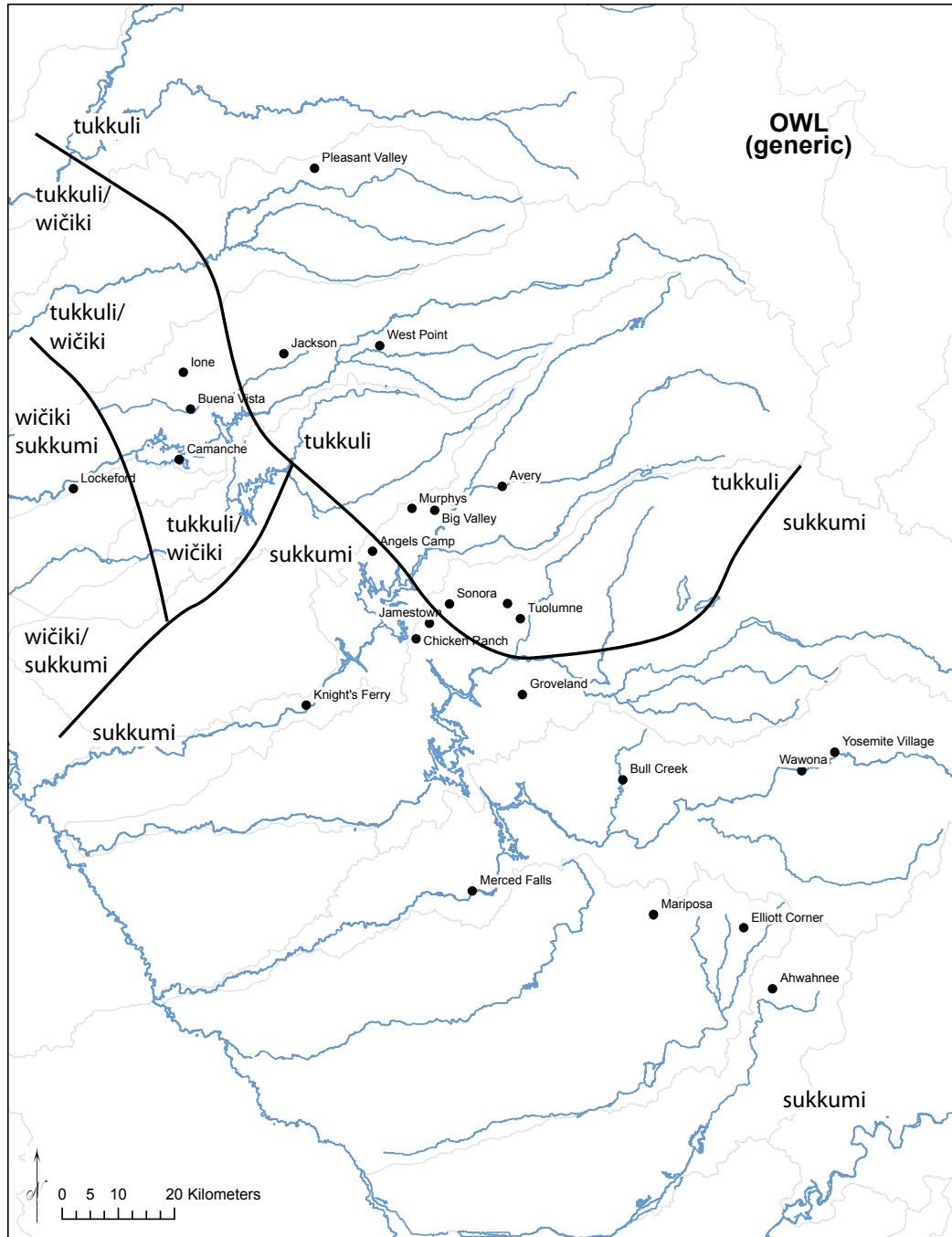


Figure 5.63: Eastern Miwok isogloss map: Owl (generic)

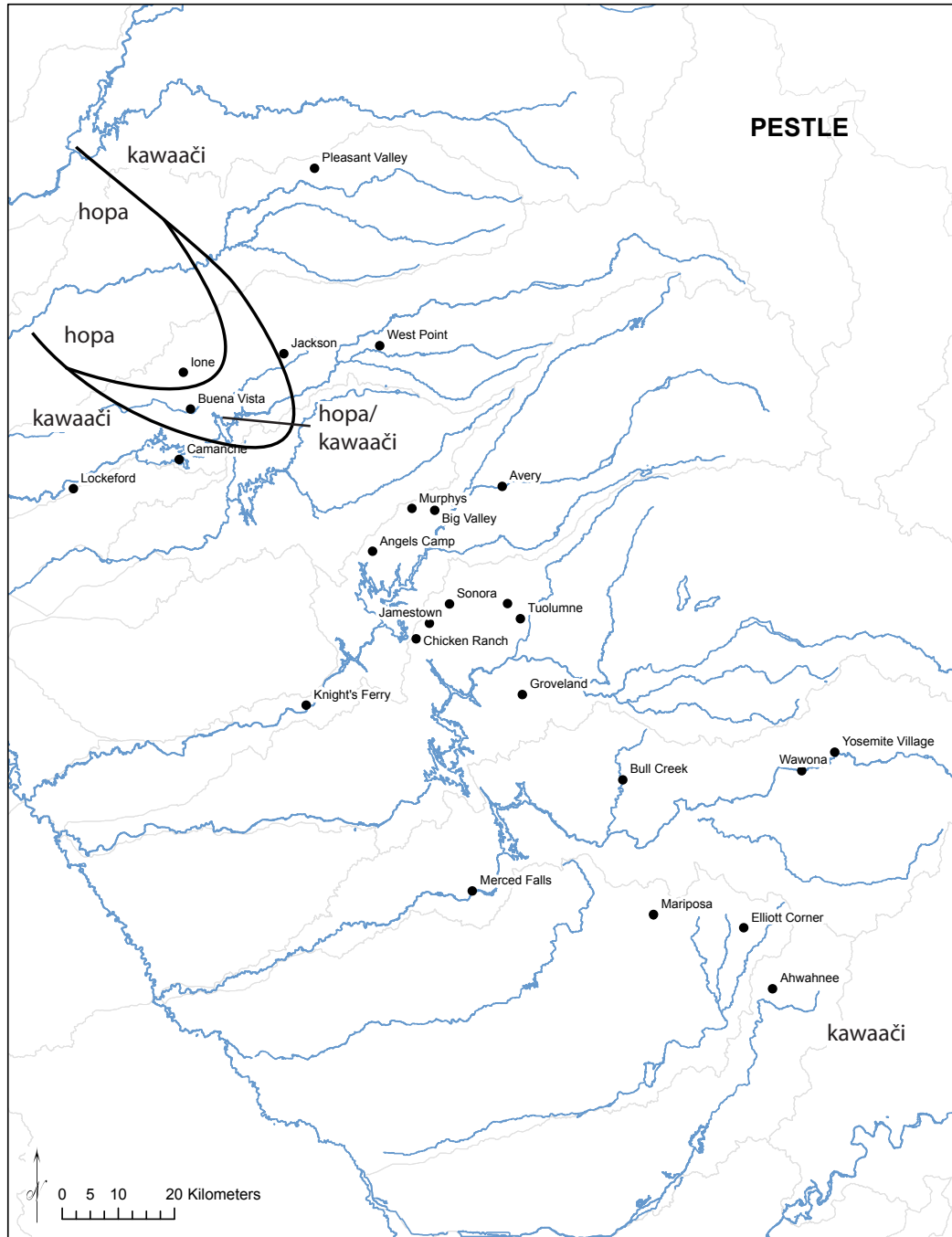


Figure 5.64: Eastern Miwok isogloss map: Pestle

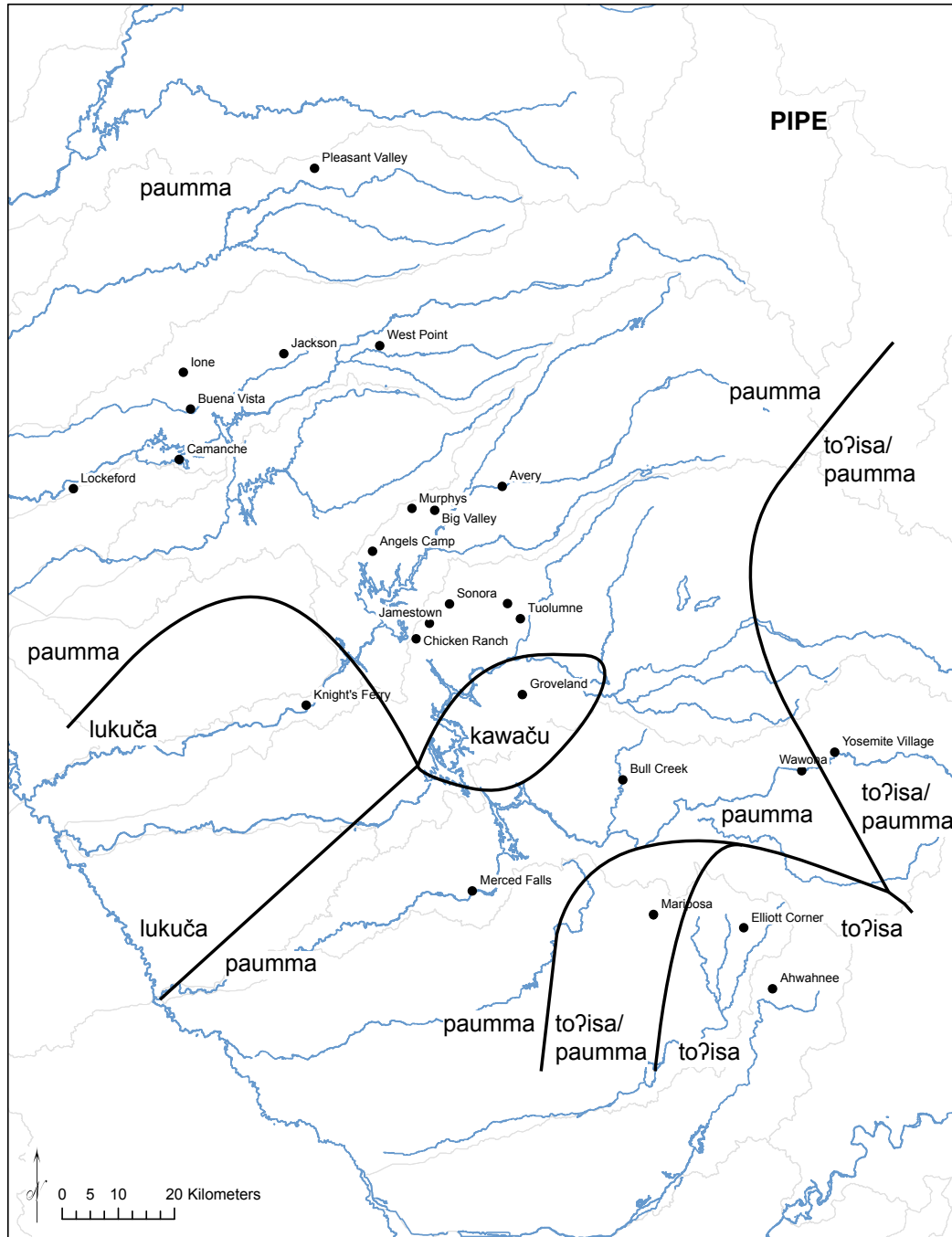


Figure 5.65: Eastern Miwok isogloss map: Pipe

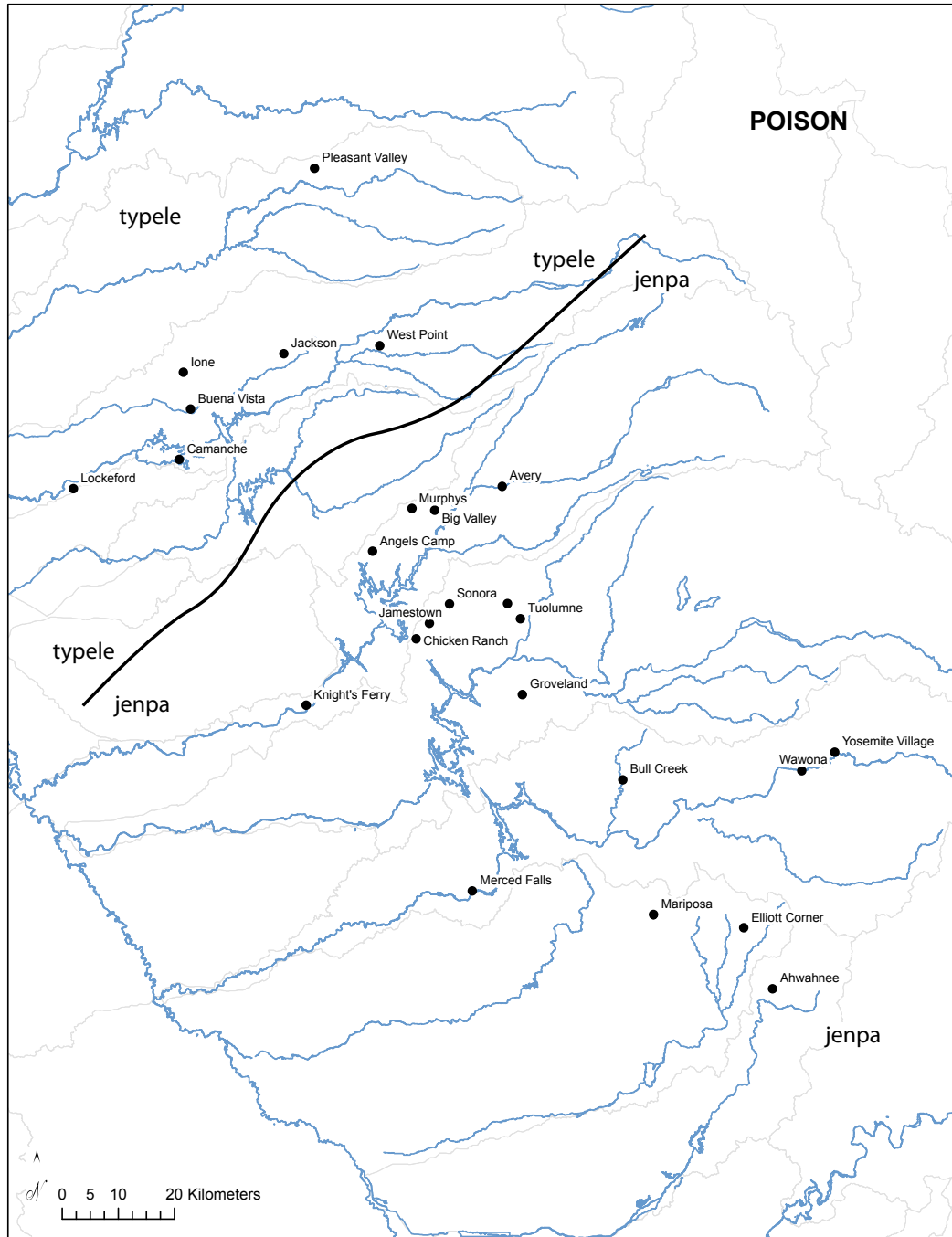


Figure 5.66: Eastern Miwok isogloss map: Poison

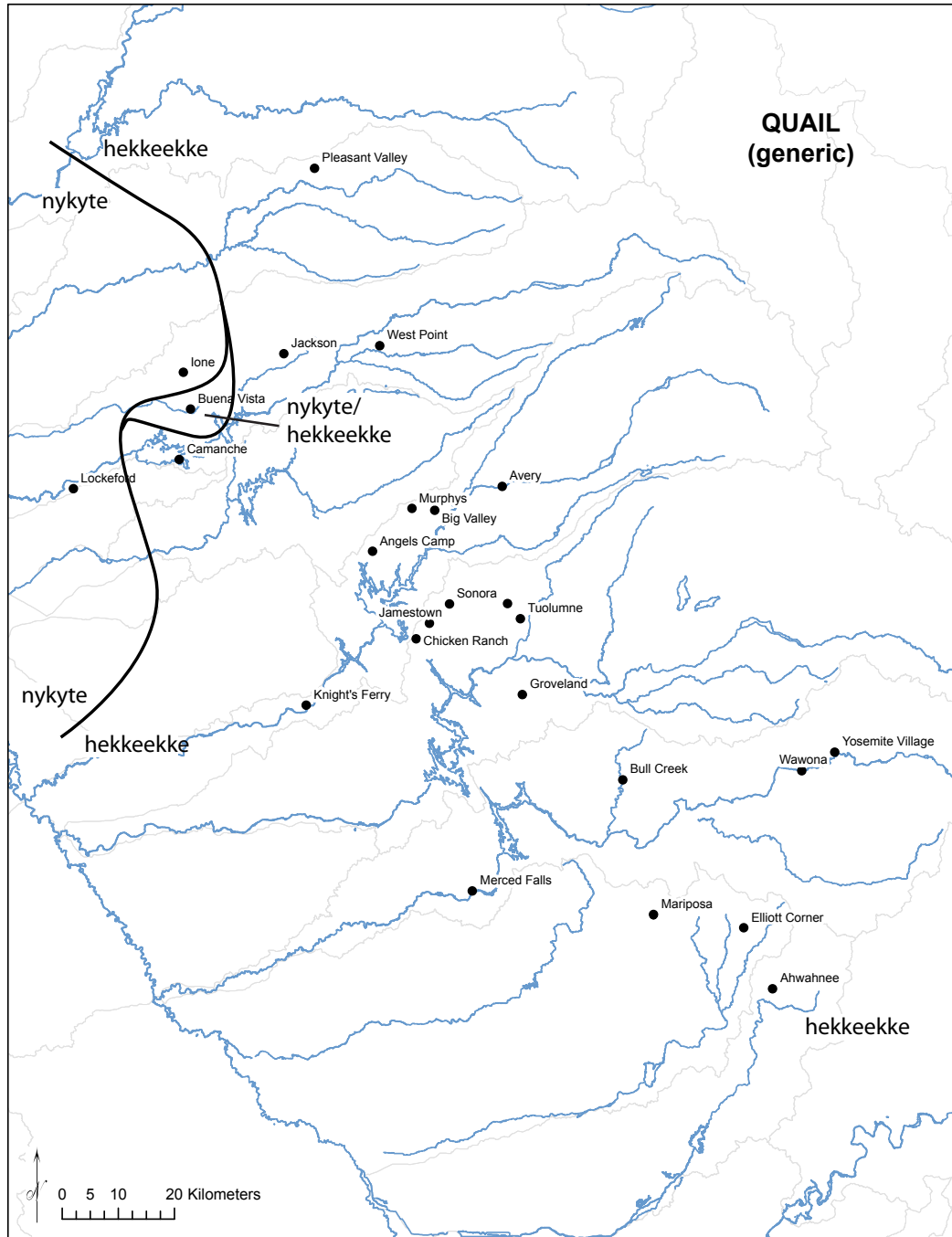


Figure 5.67: Eastern Miwok isogloss map: Quail (generic)

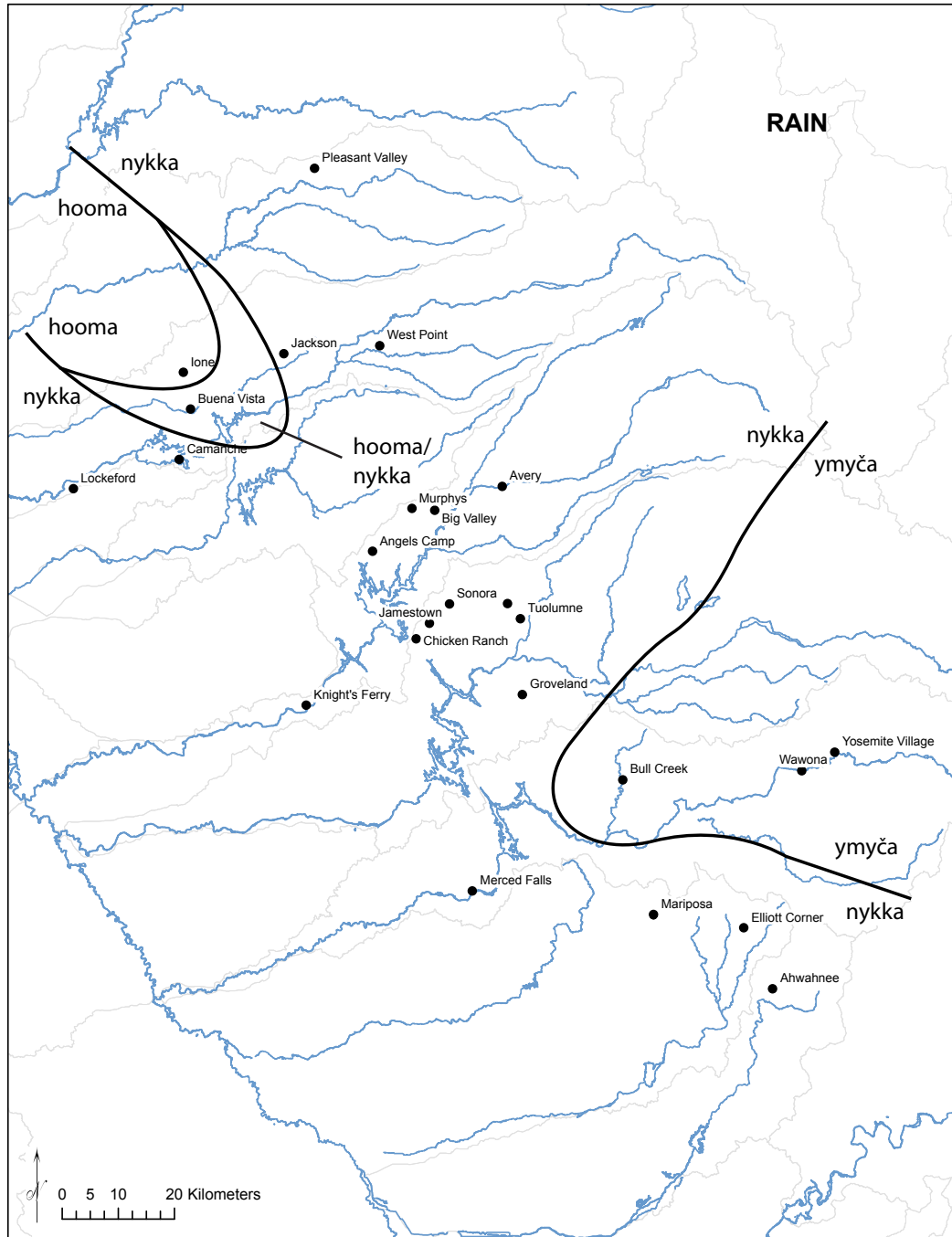


Figure 5.68: Eastern Miwok isogloss map: Rain

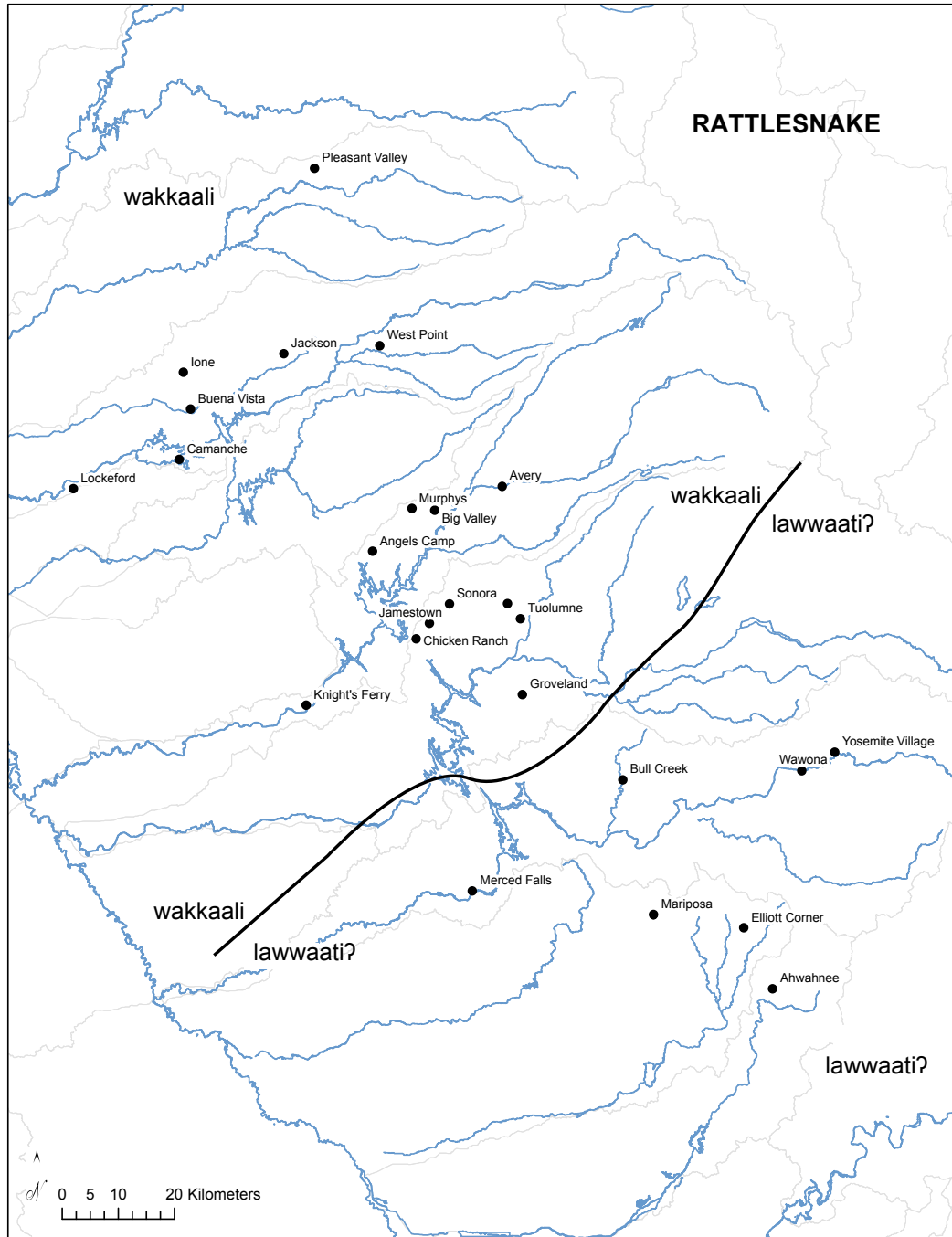


Figure 5.69: Eastern Miwok isogloss map: Rattlesnake

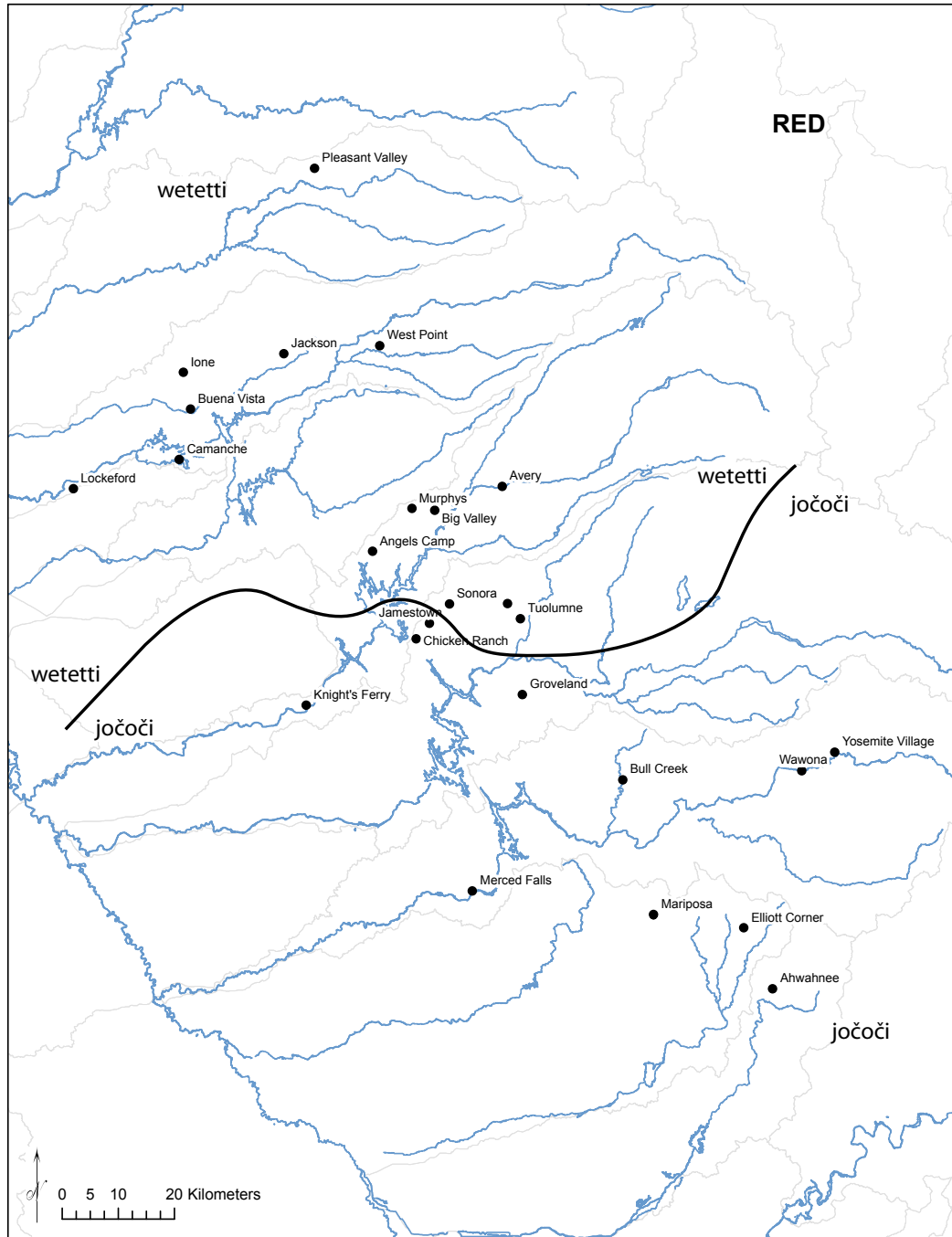


Figure 5.70: Eastern Miwok isogloss map: Red

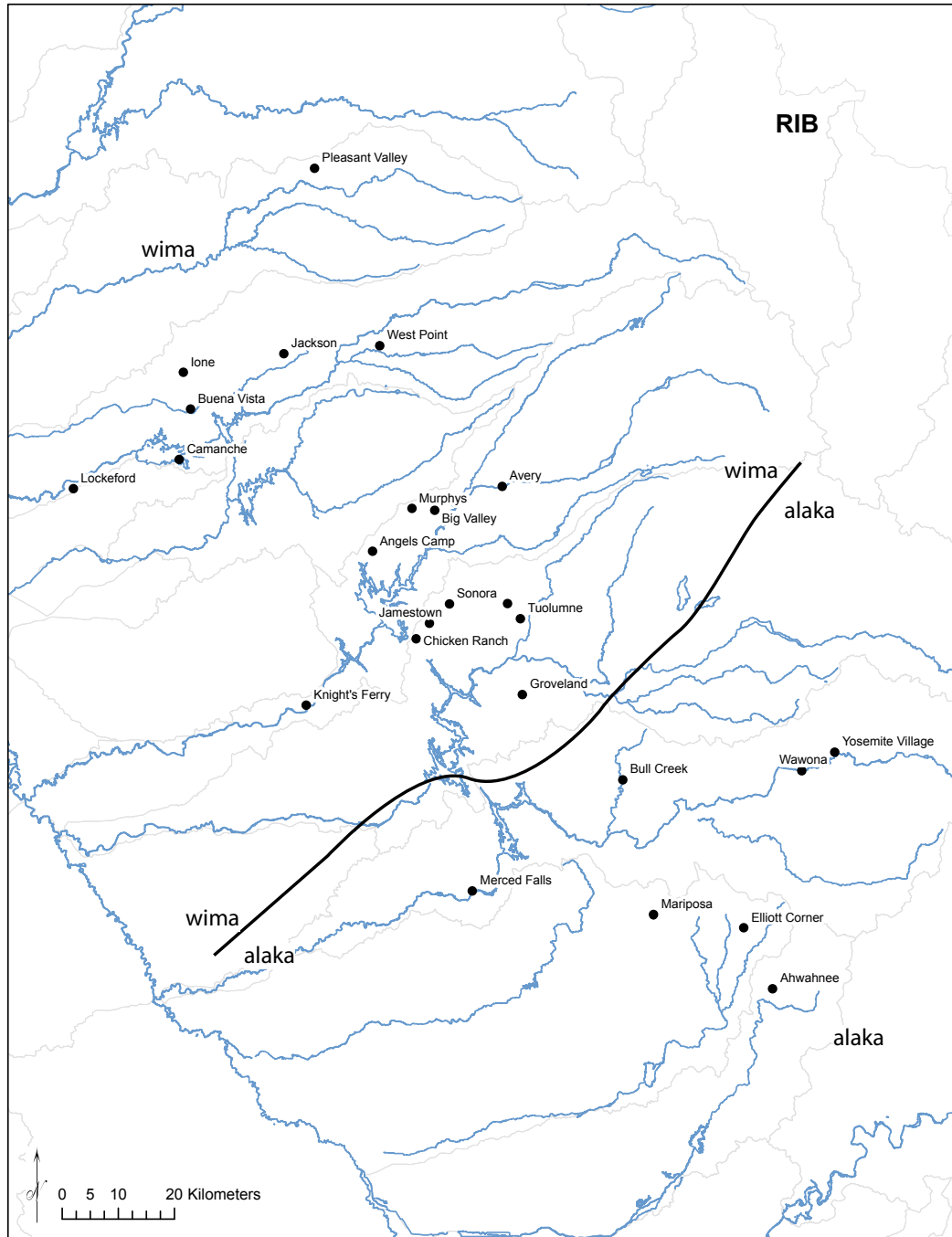


Figure 5.71: Eastern Miwok isogloss map: Rib

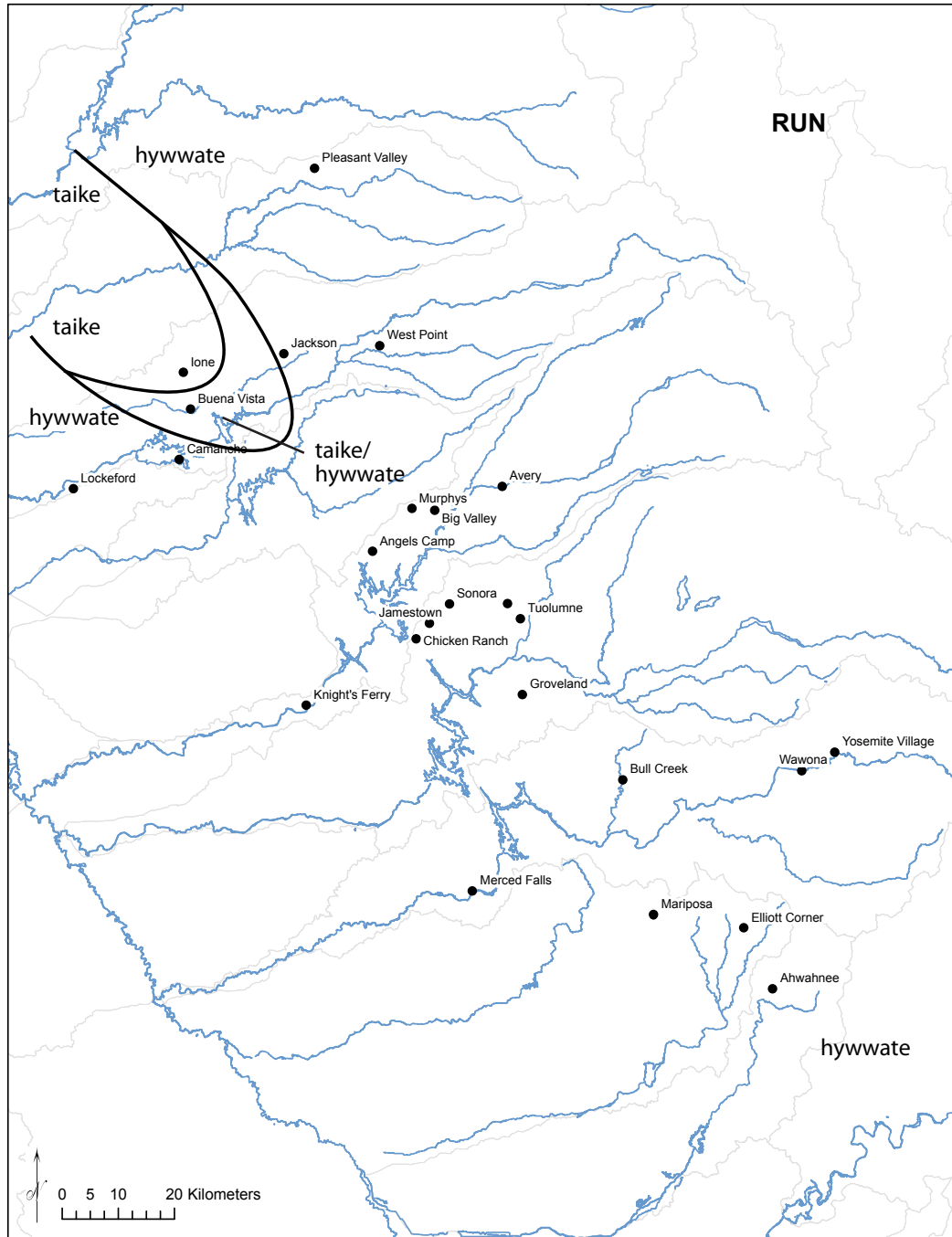


Figure 5.72: Eastern Miwok isogloss map: Run

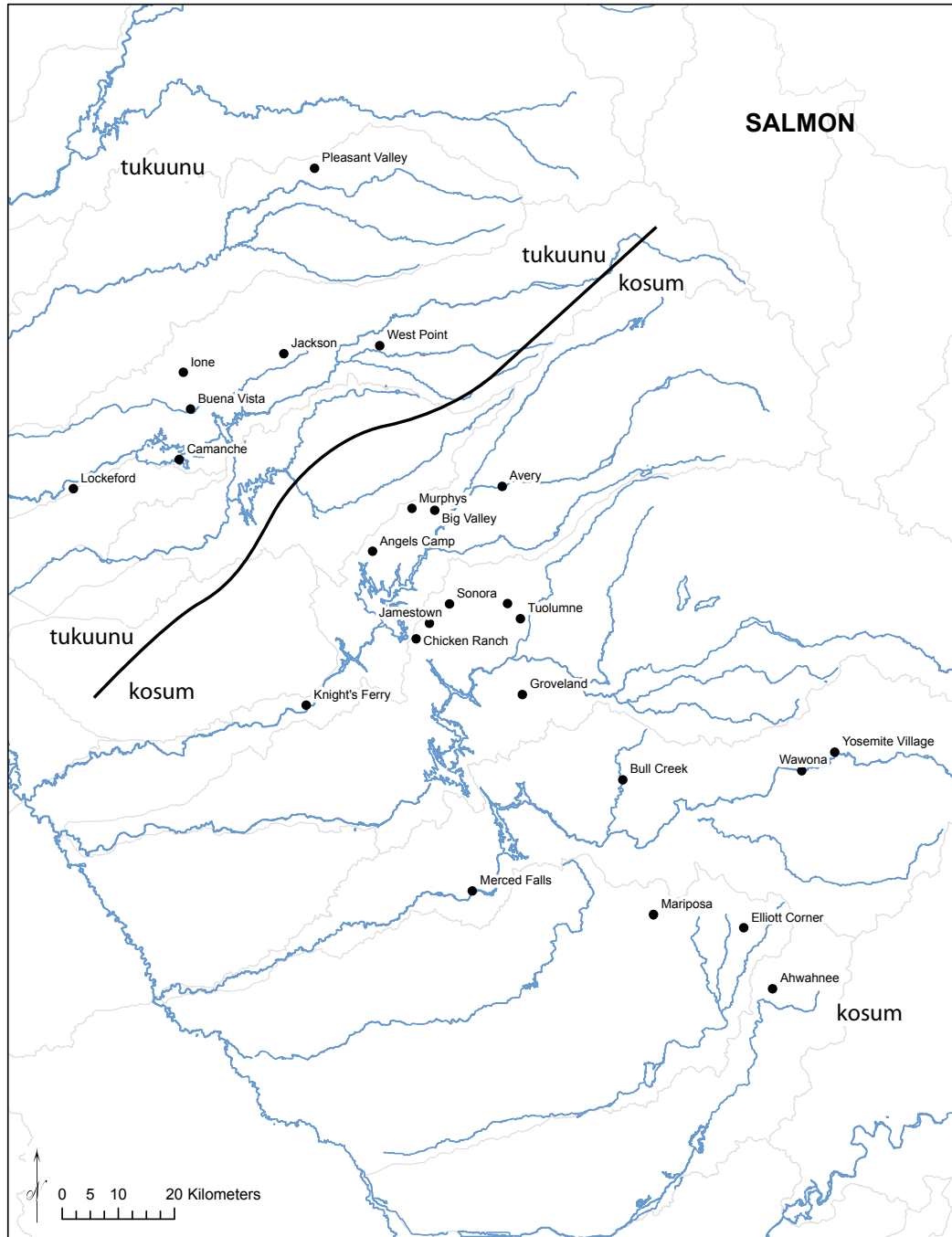


Figure 5.73: Eastern Miwok isogloss map: Salmon

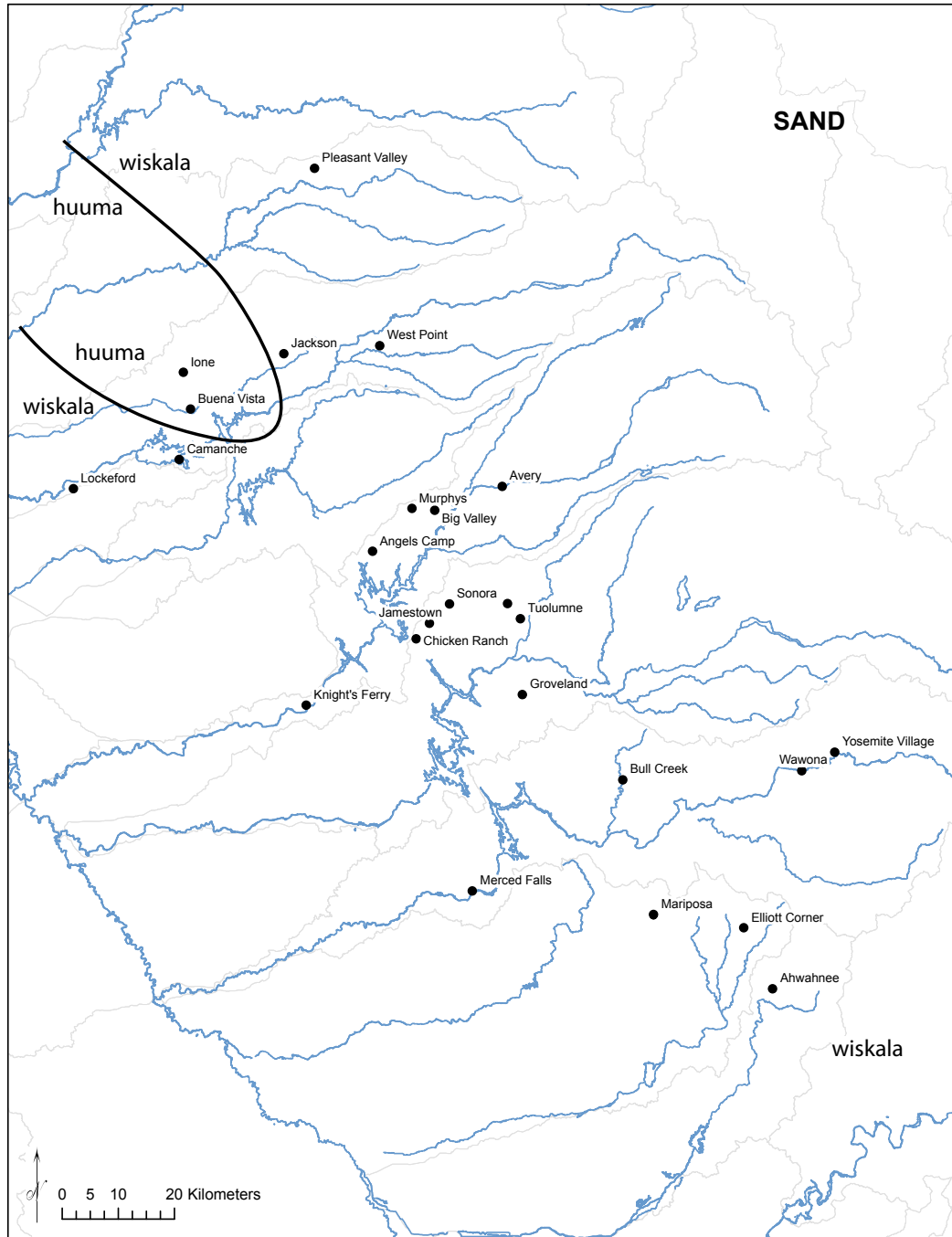


Figure 5.74: Eastern Miwok isogloss map: Sand

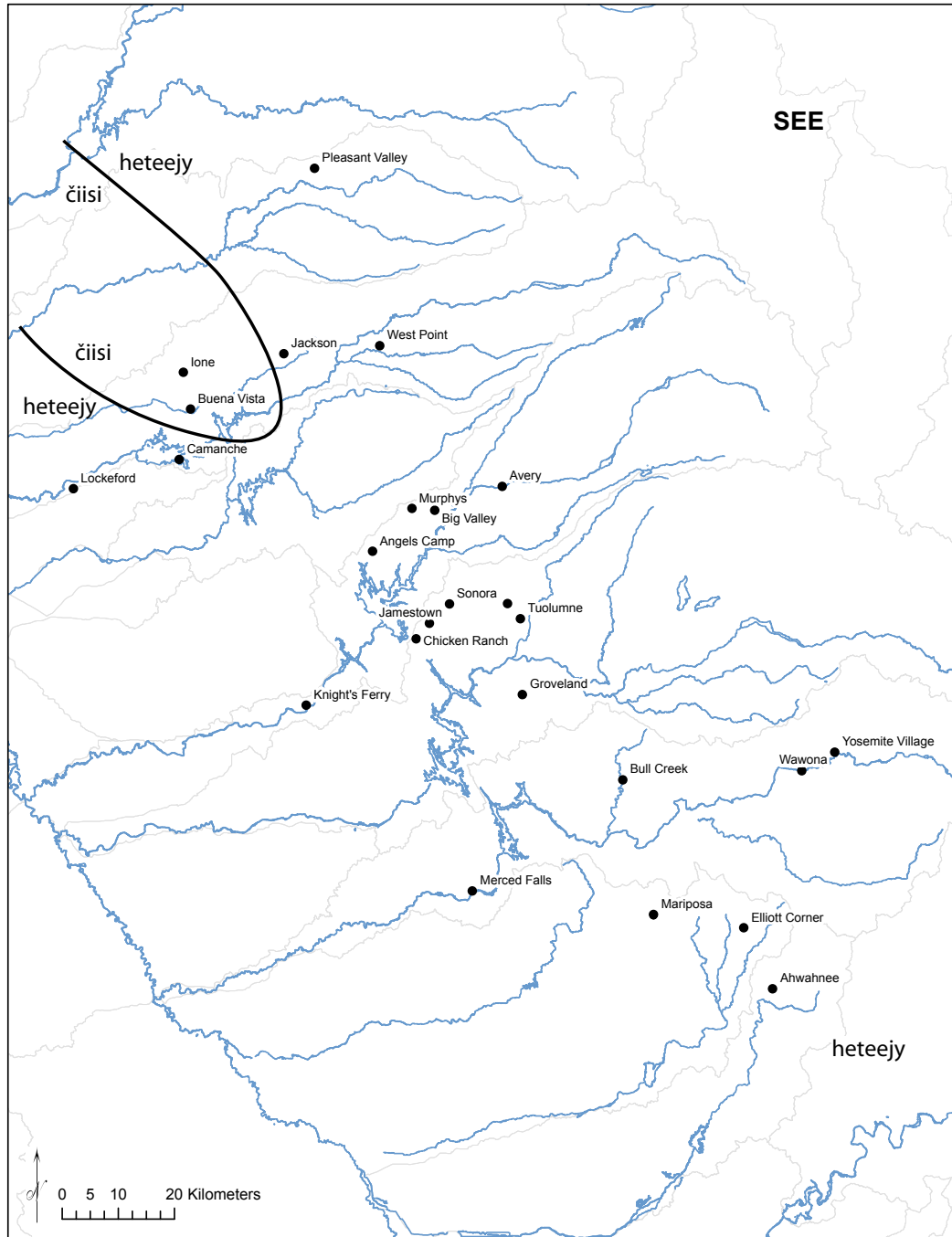


Figure 5.75: Eastern Miwok isogloss map: See

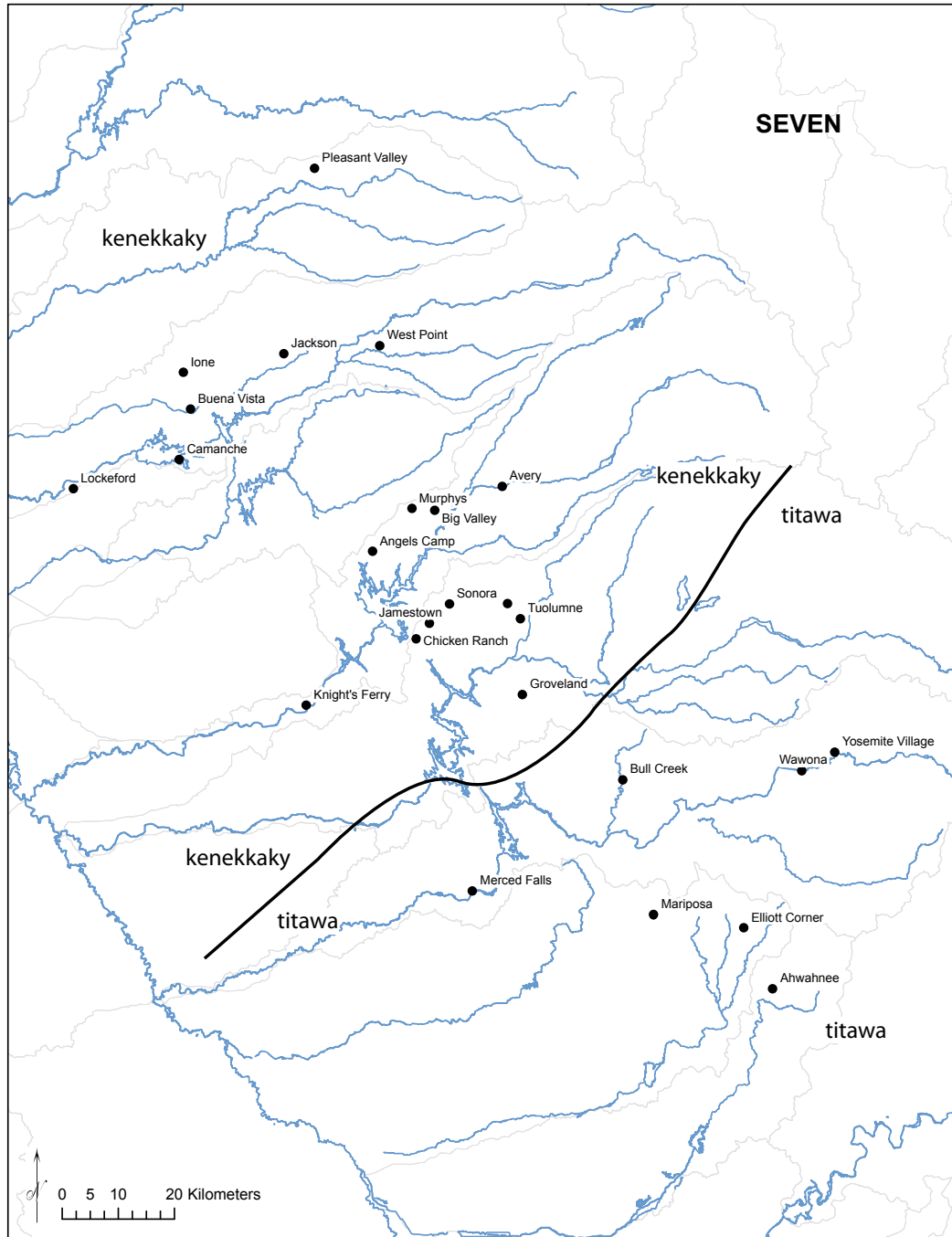


Figure 5.76: Eastern Miwok isogloss map: Seven

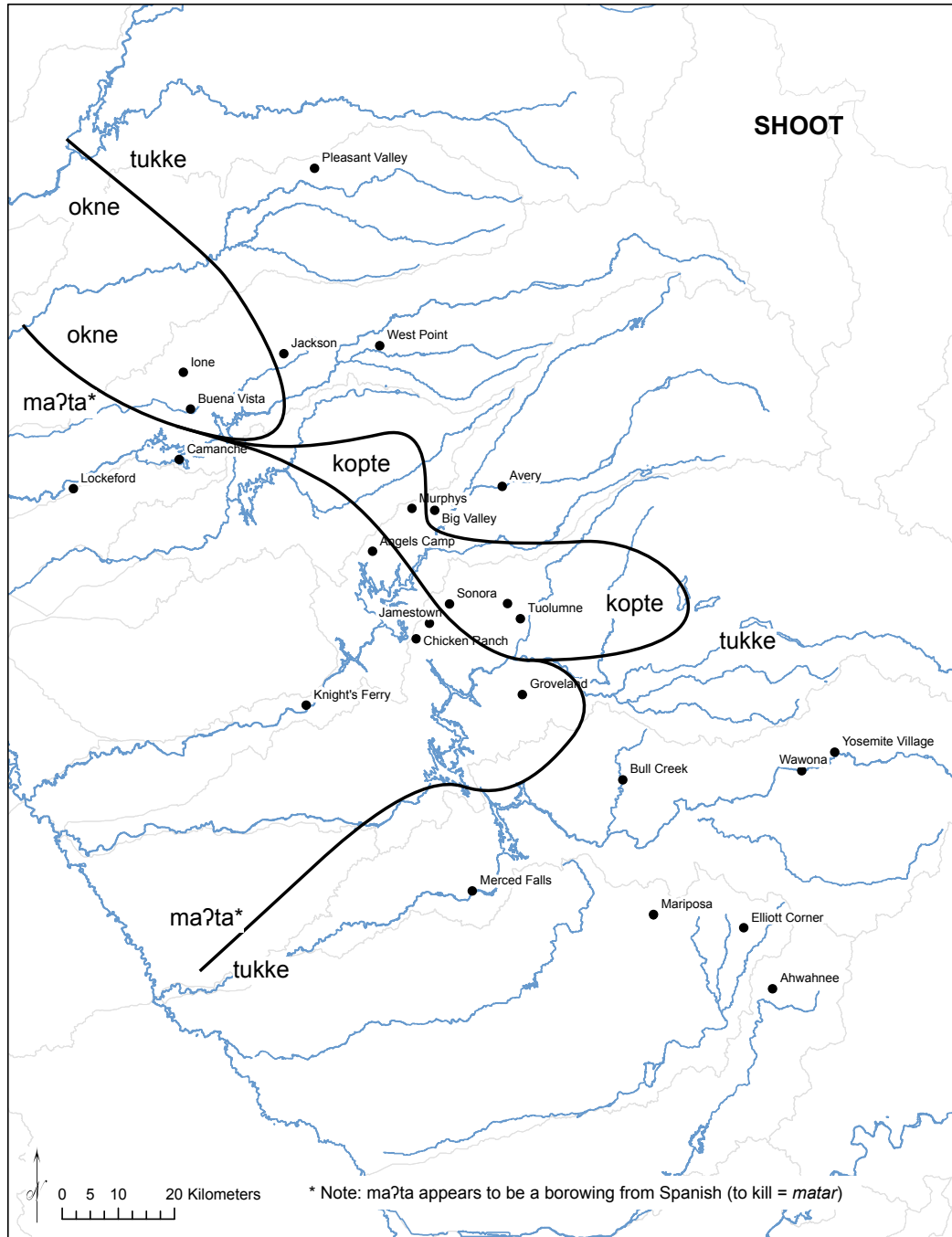


Figure 5.77: Eastern Miwok isogloss map: Shoot

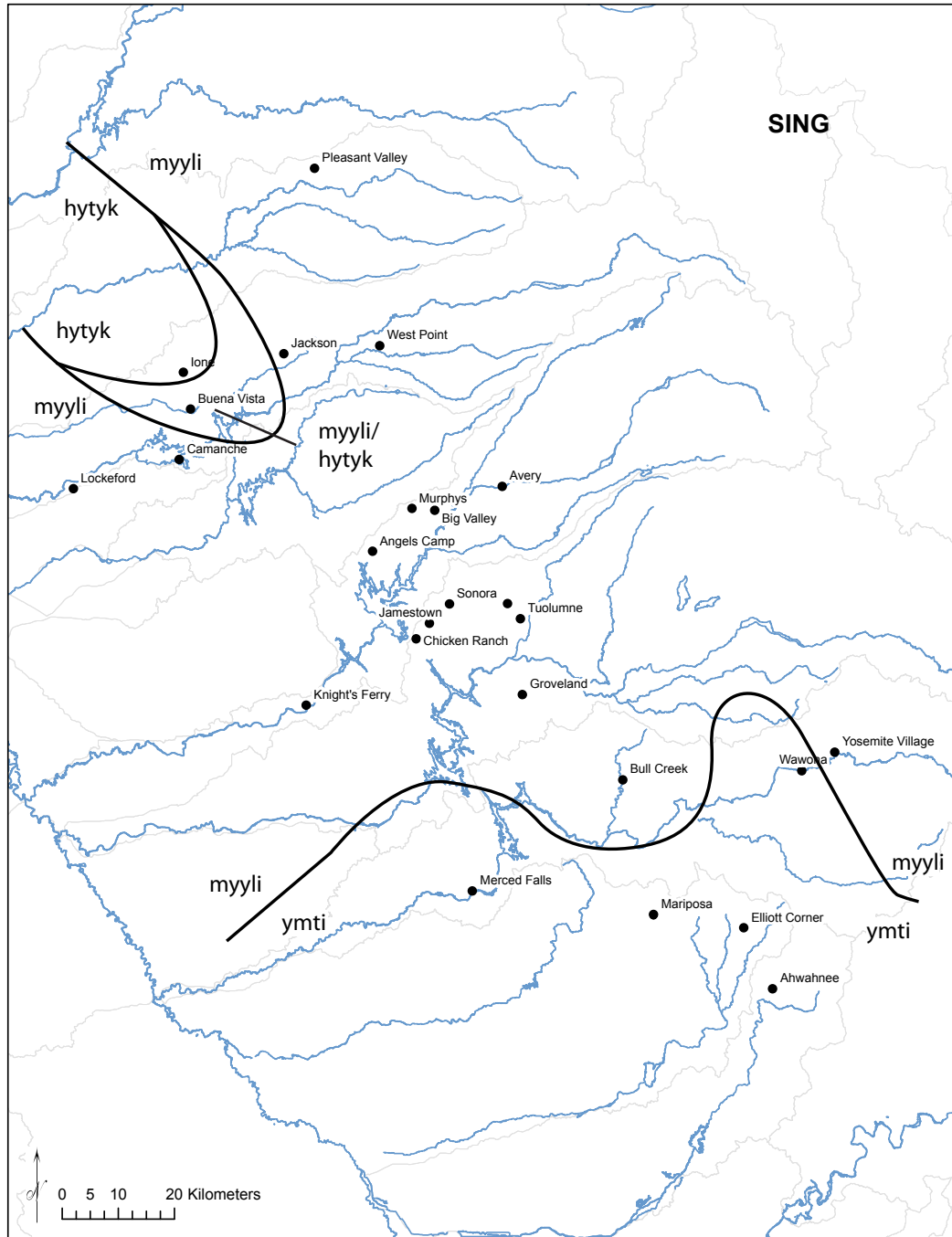


Figure 5.78: Eastern Miwok isogloss map: Sing

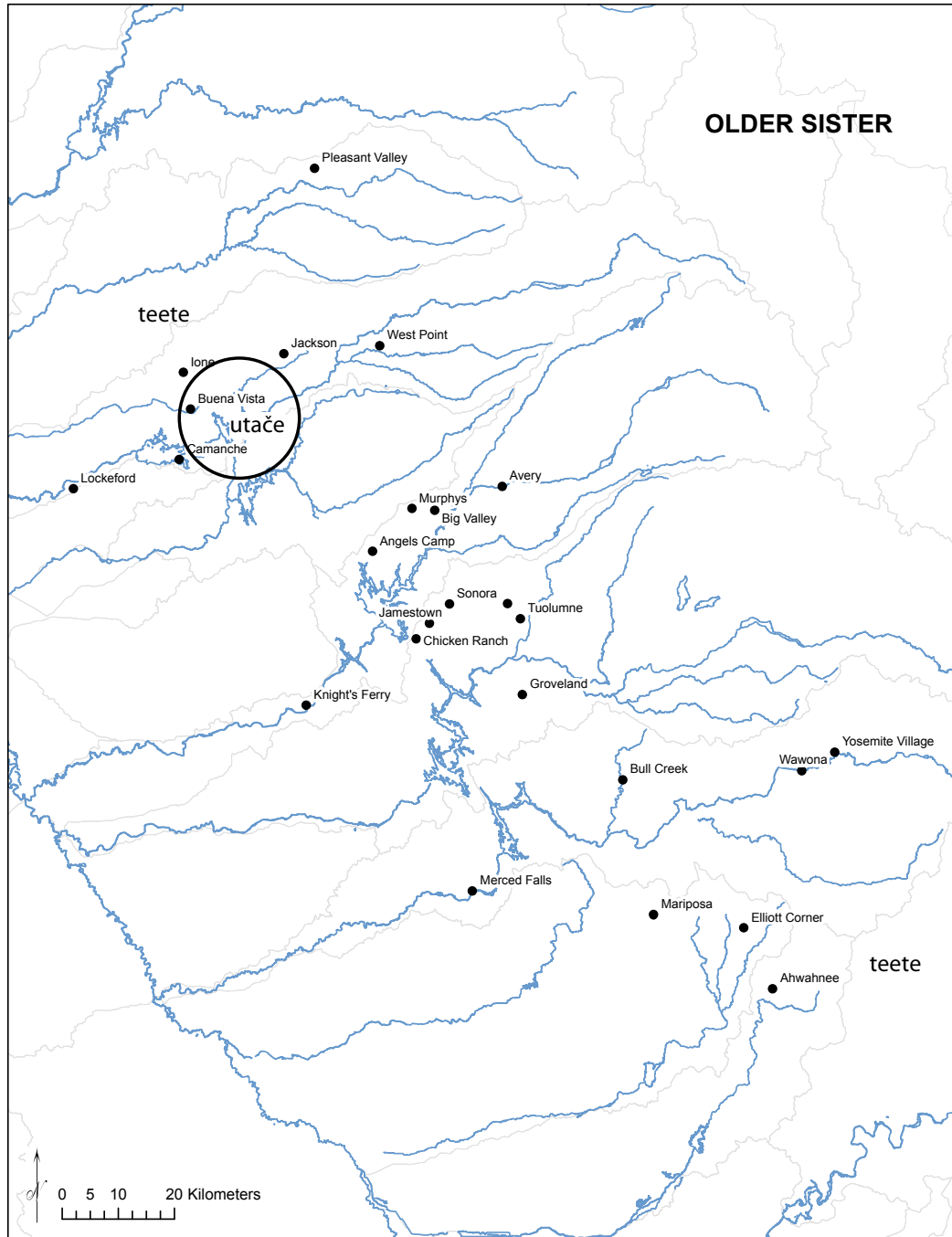


Figure 5.79: Eastern Miwok isogloss map: Older Sister

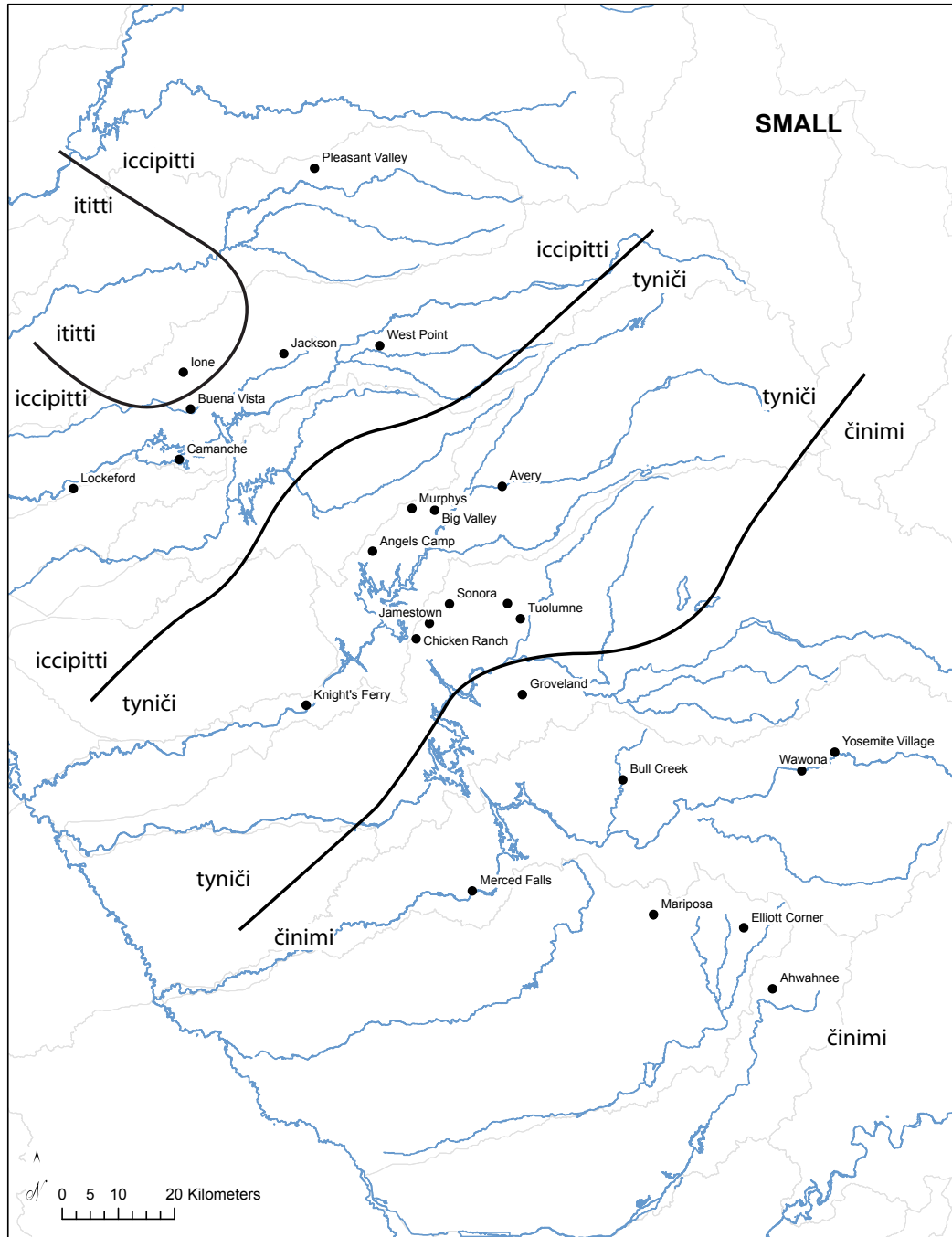


Figure 5.80: Eastern Miwok isogloss map: Small

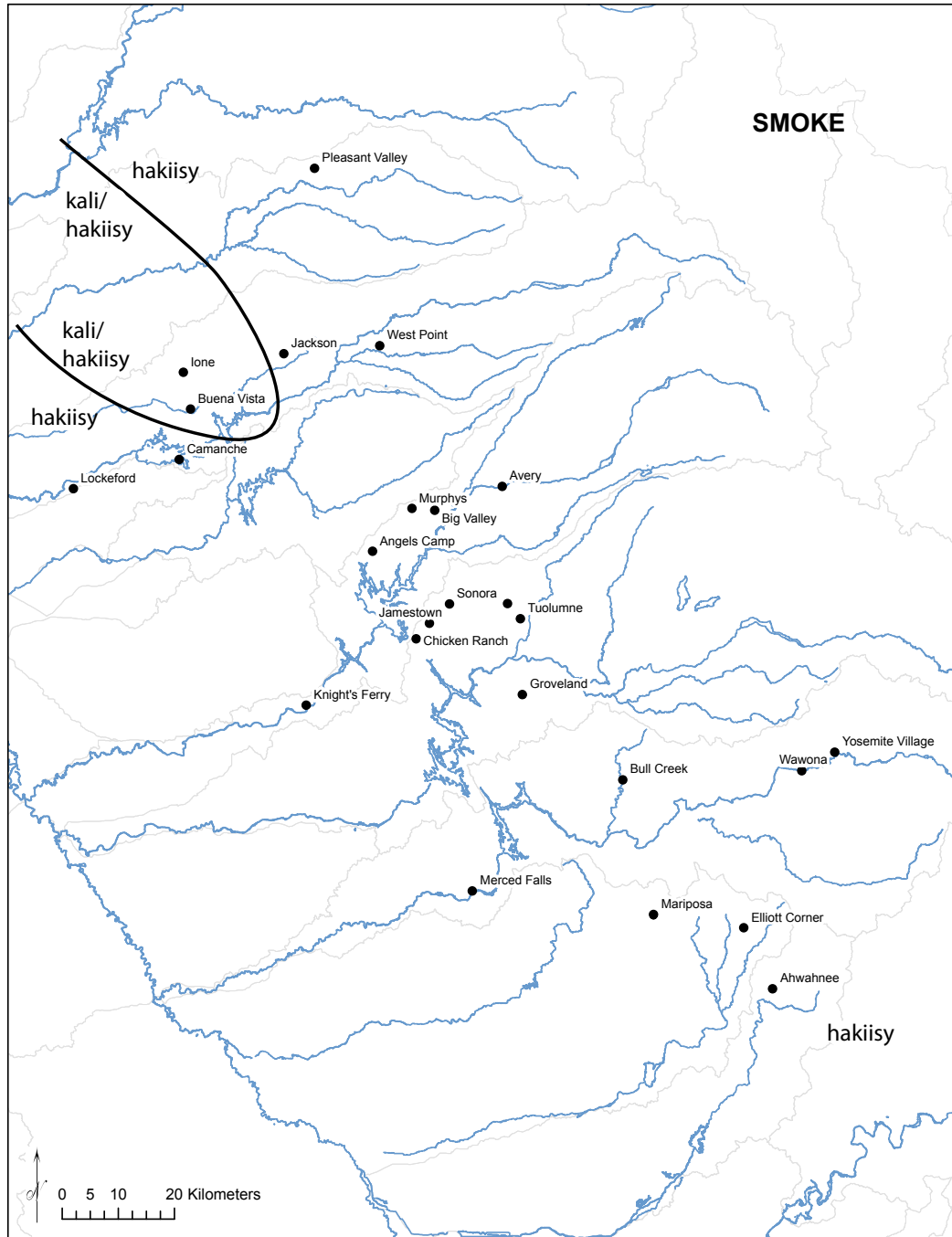


Figure 5.81: Eastern Miwok isogloss map: Smoke

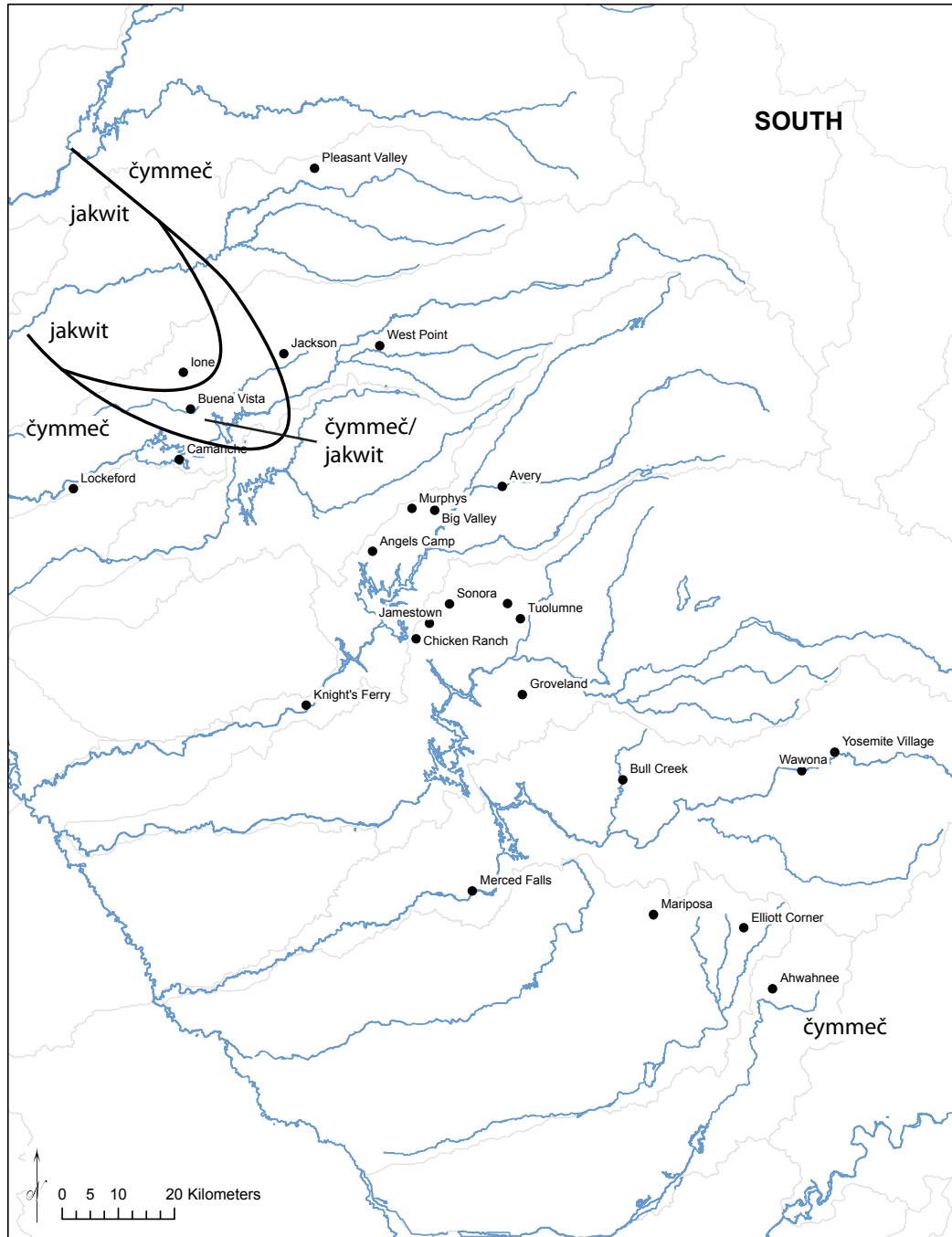


Figure 5.82: Eastern Miwok isogloss map: South

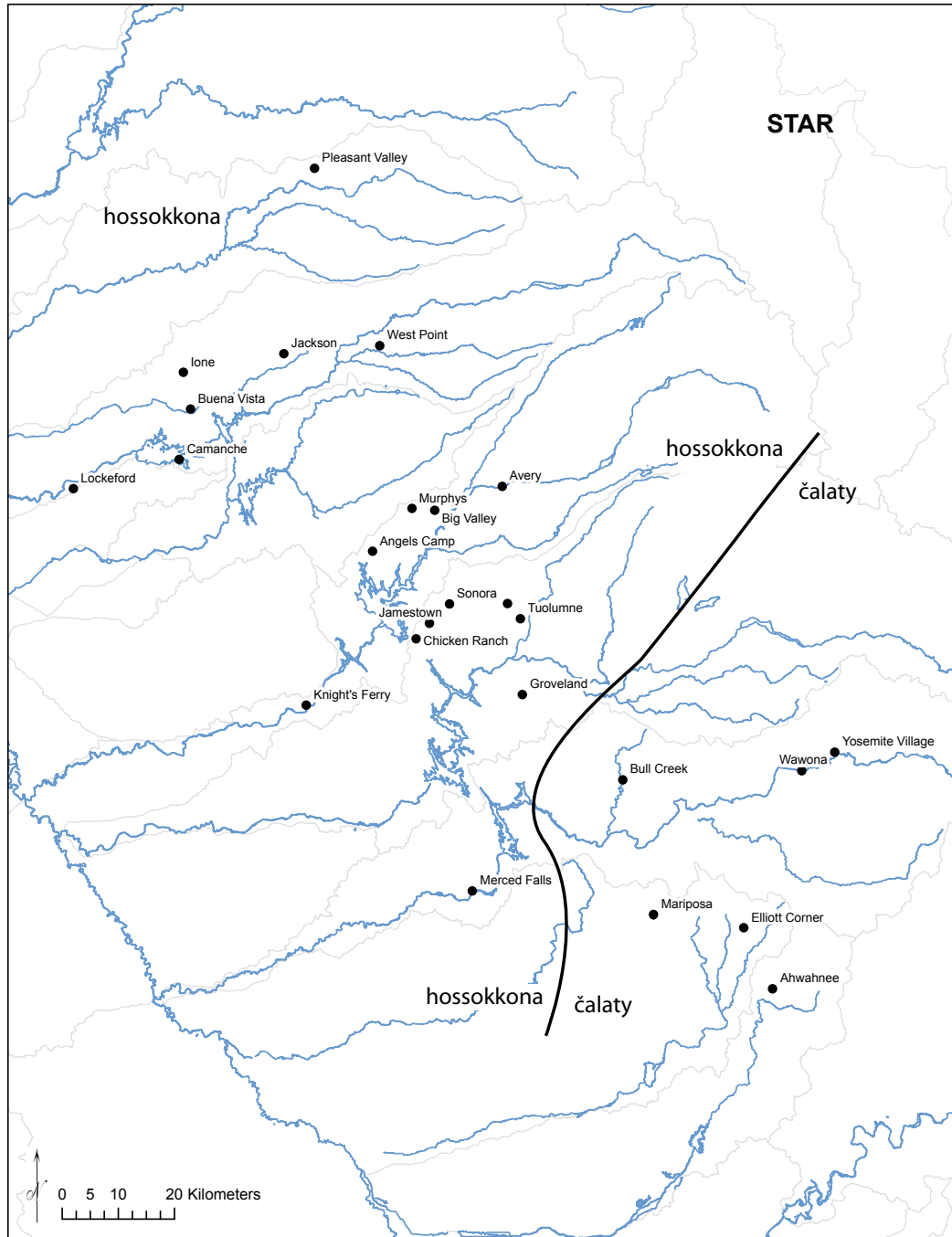


Figure 5.83: Eastern Miwok isogloss map: Star

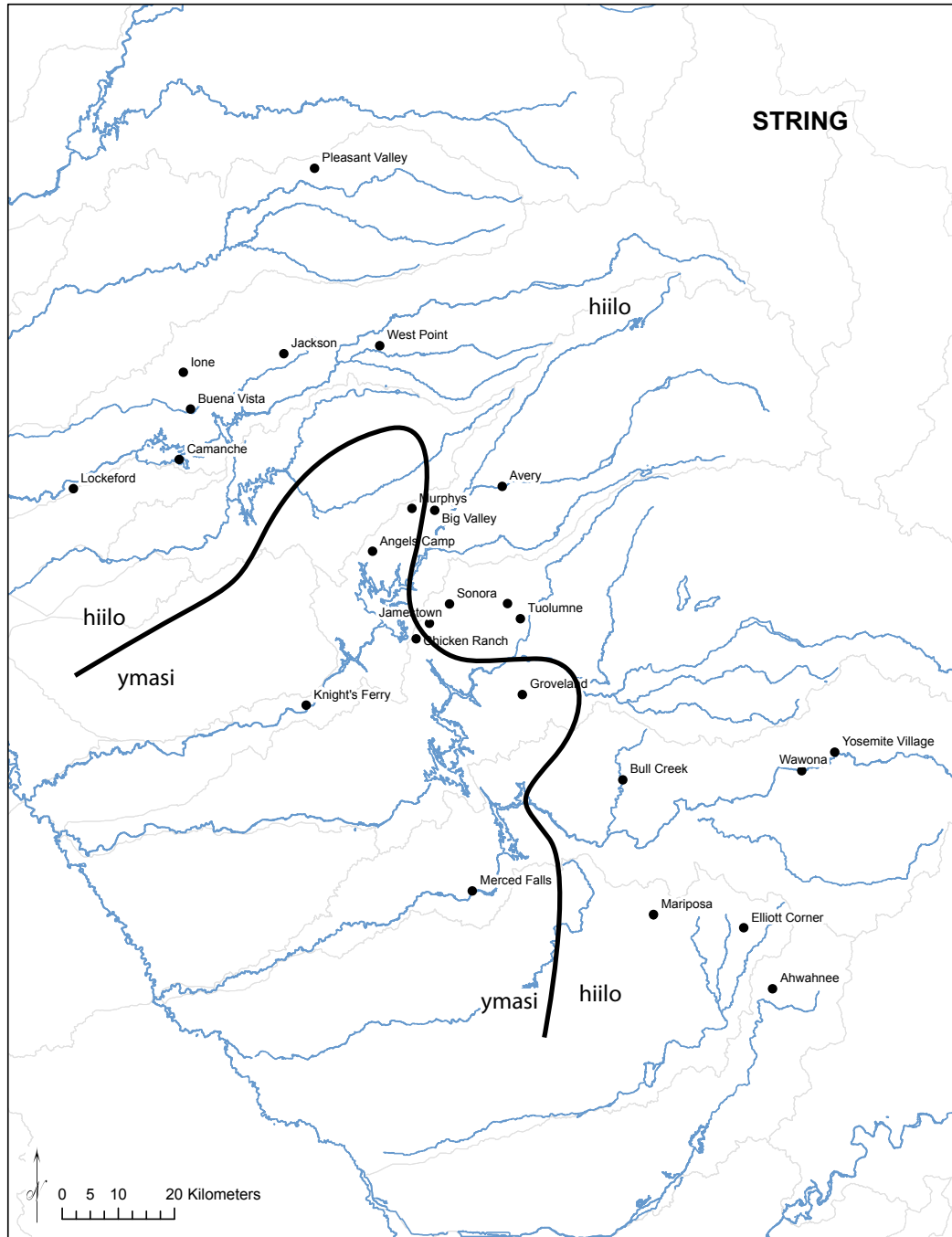


Figure 5.84: Eastern Miwok isogloss map: String

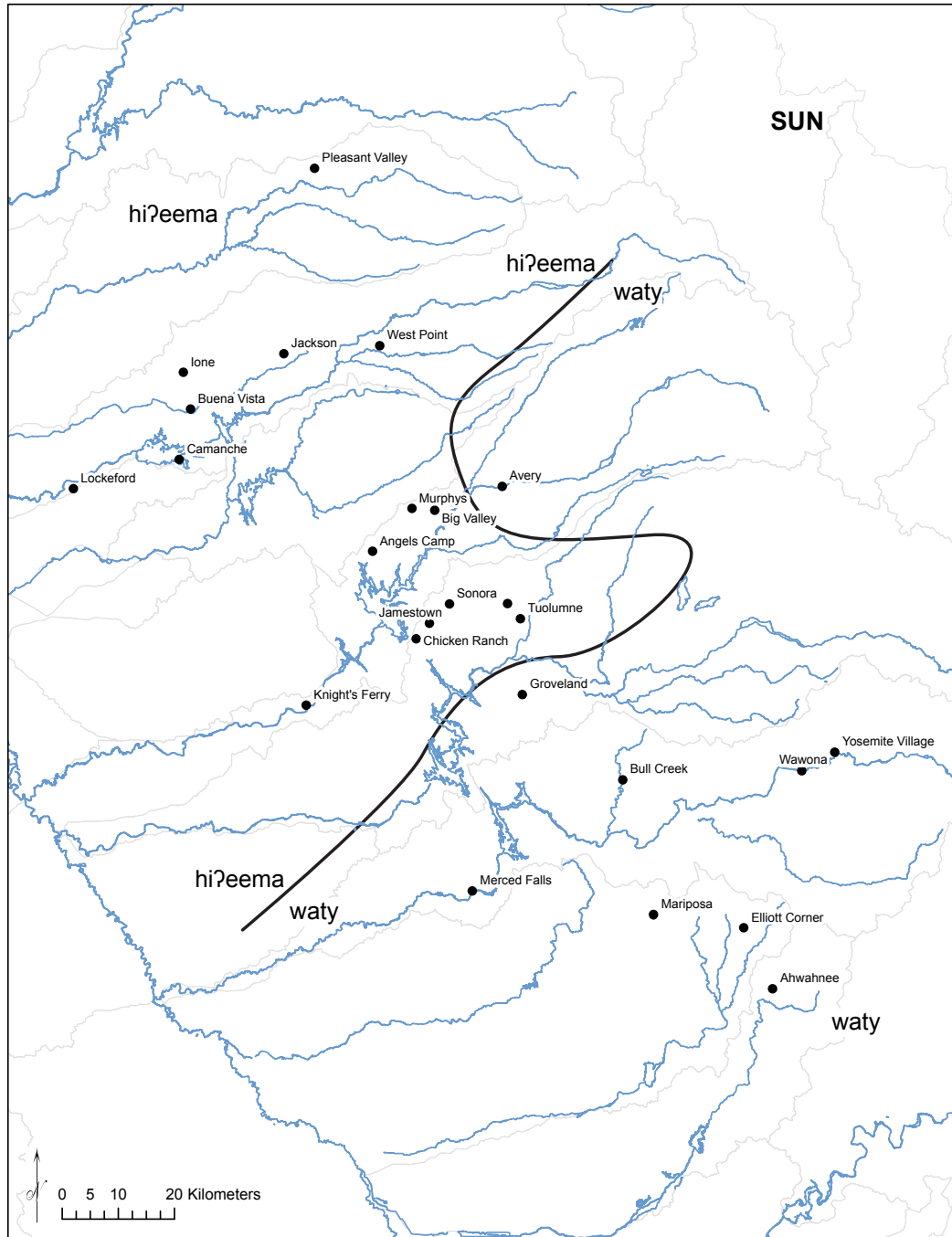


Figure 5.85: Eastern Miwok isogloss map: Sun

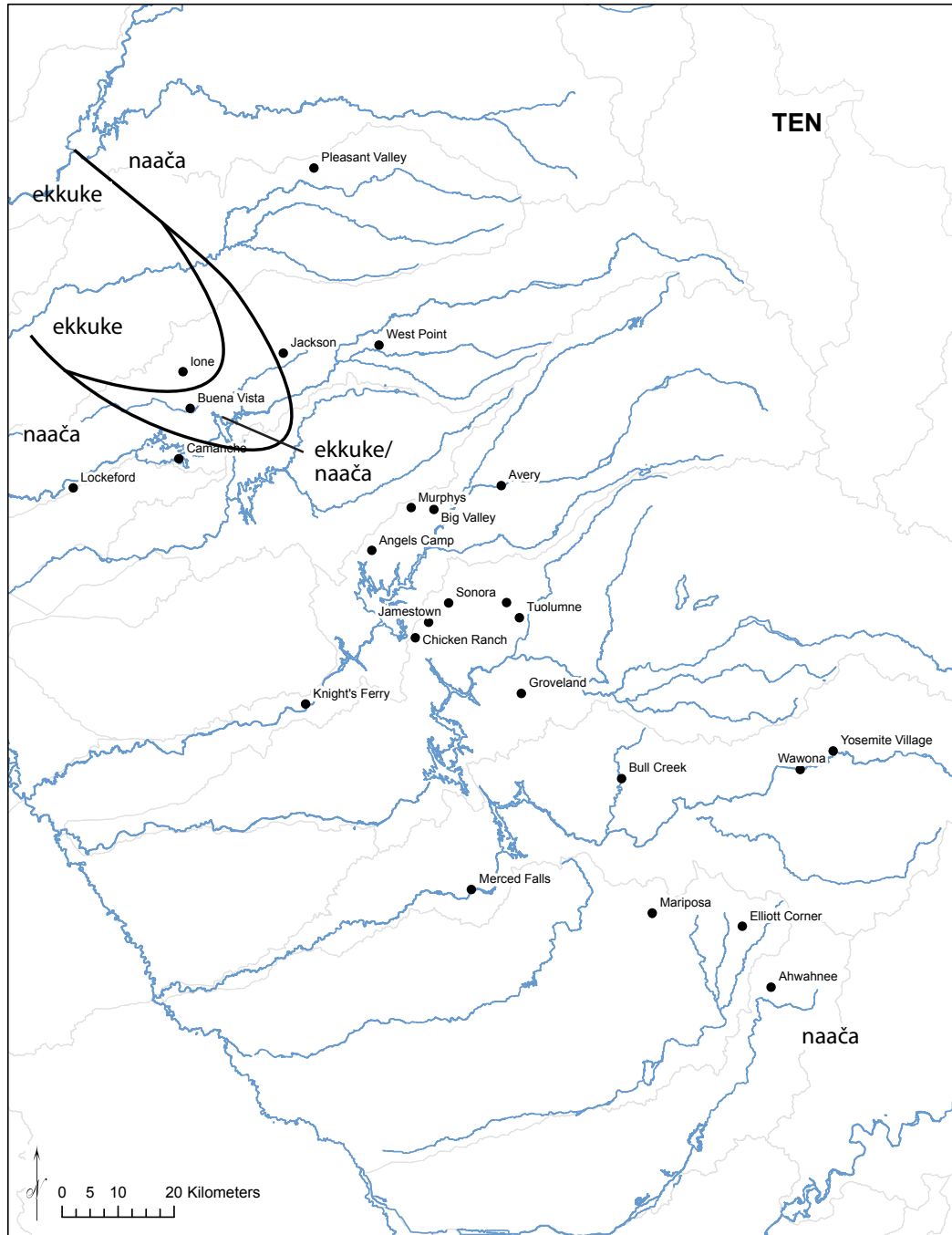


Figure 5.86: Eastern Miwok isogloss map: Ten

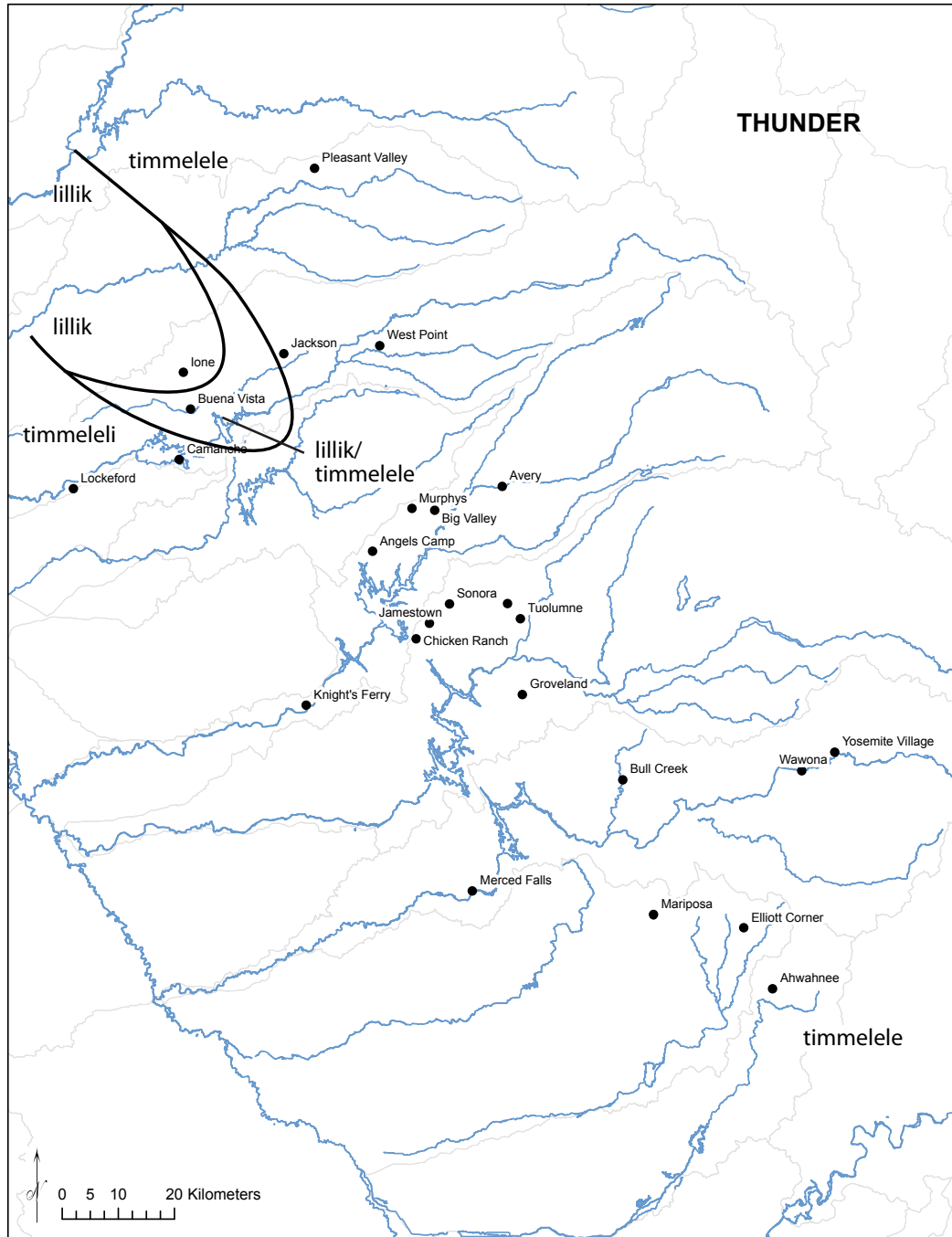


Figure 5.87: Eastern Miwok isogloss map: Thunder

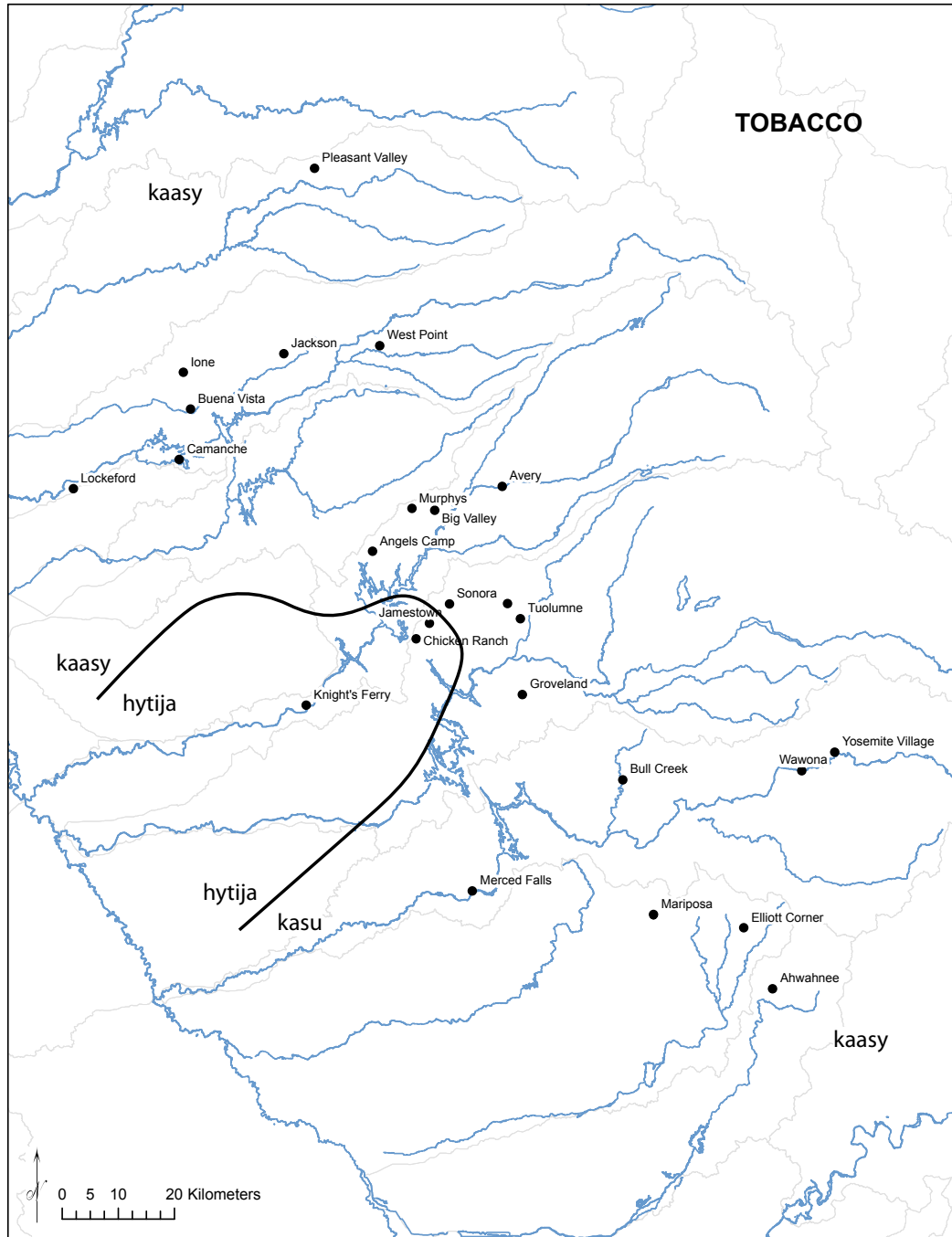


Figure 5.88: Eastern Miwok isogloss map: Tobacco

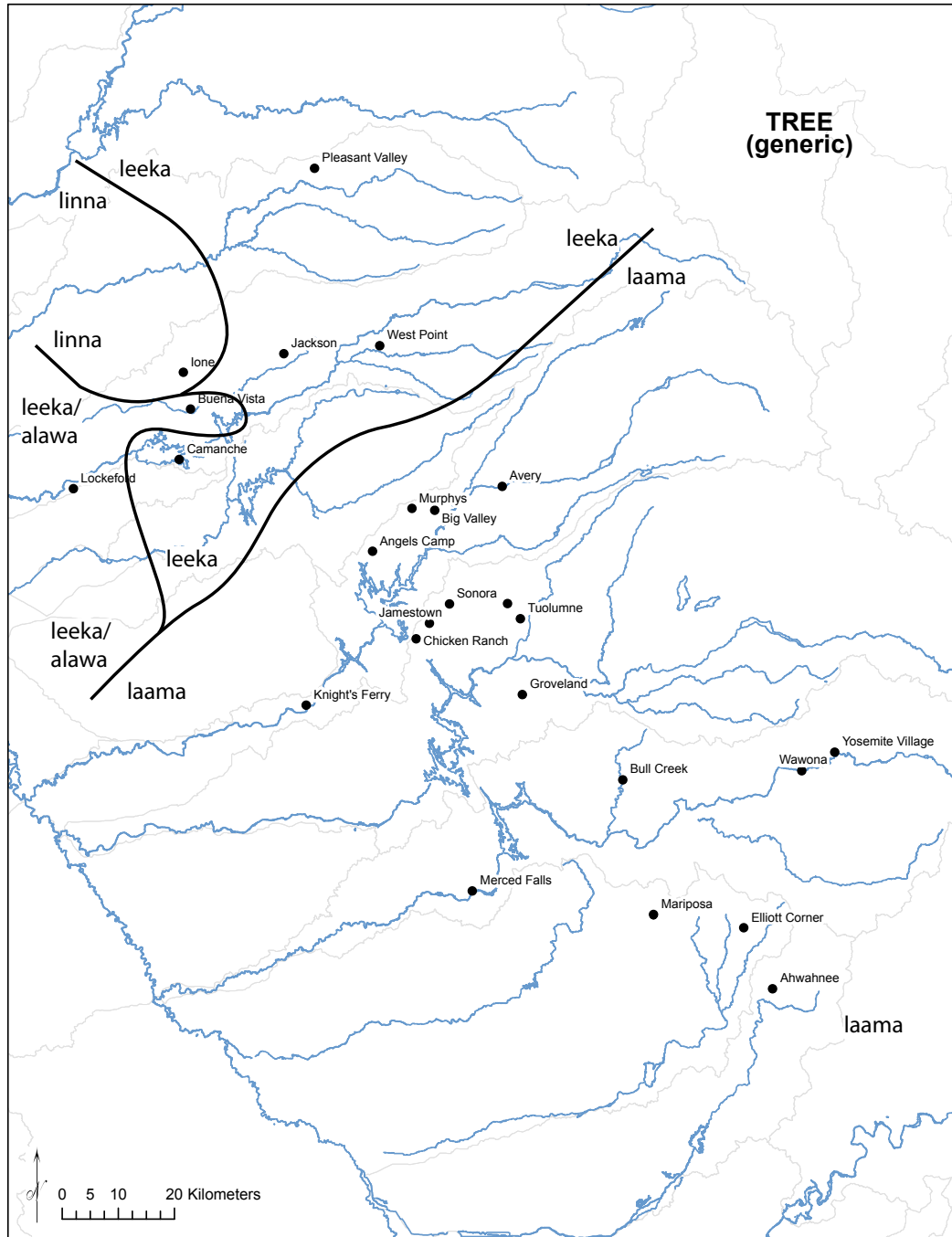


Figure 5.89: Eastern Miwok isogloss map: Tree (generic)

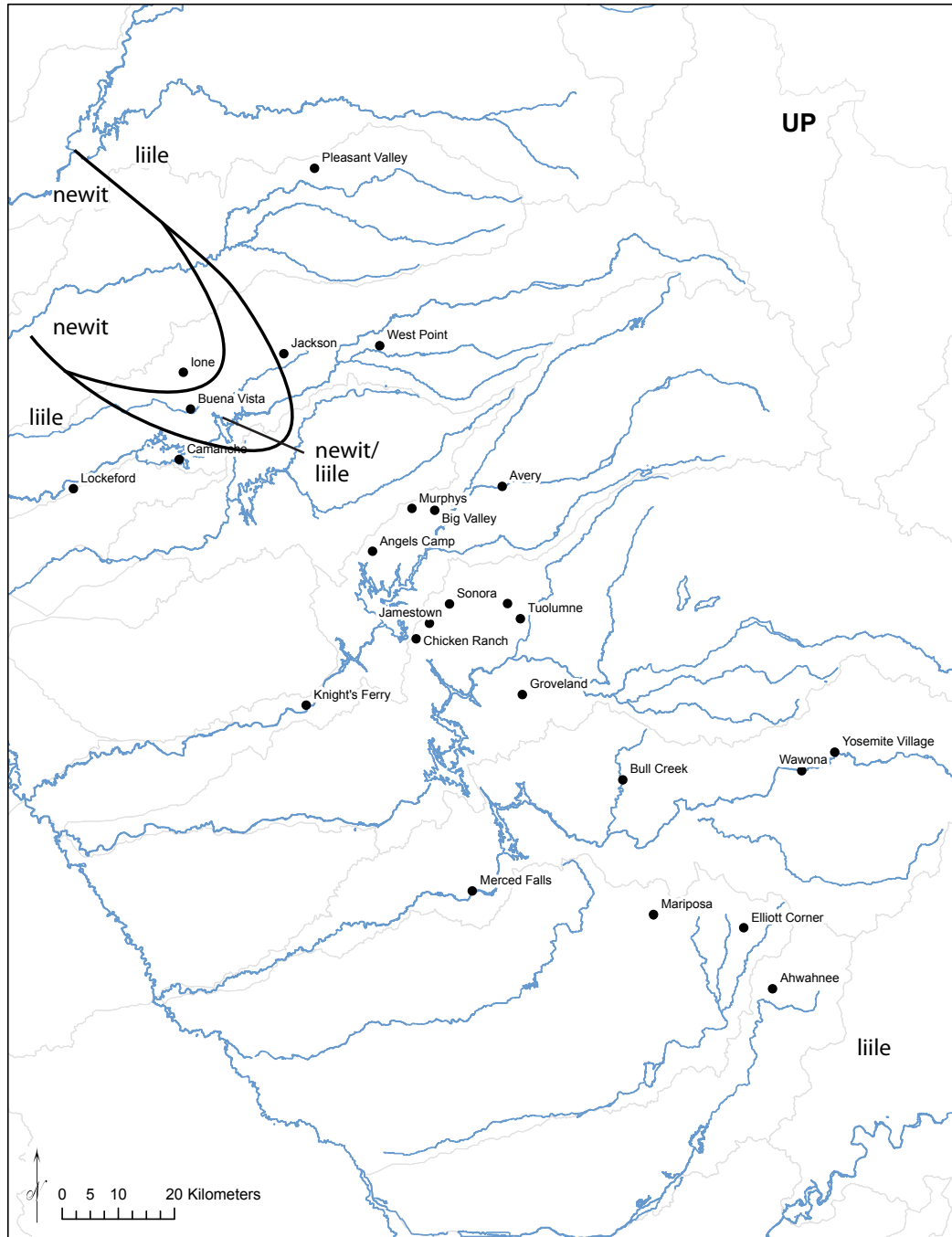
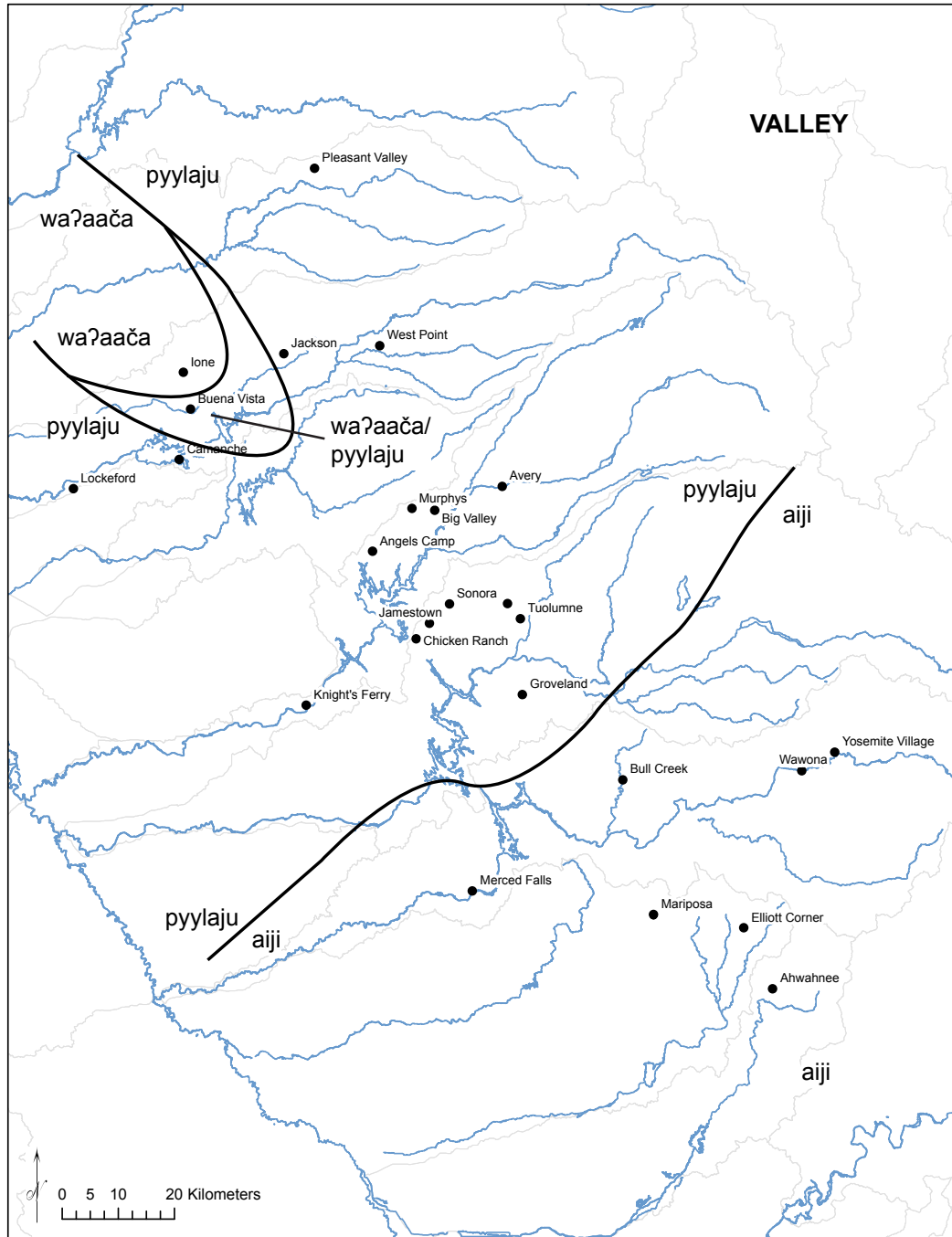


Figure 5.90: Eastern Miwok isogloss map: Up



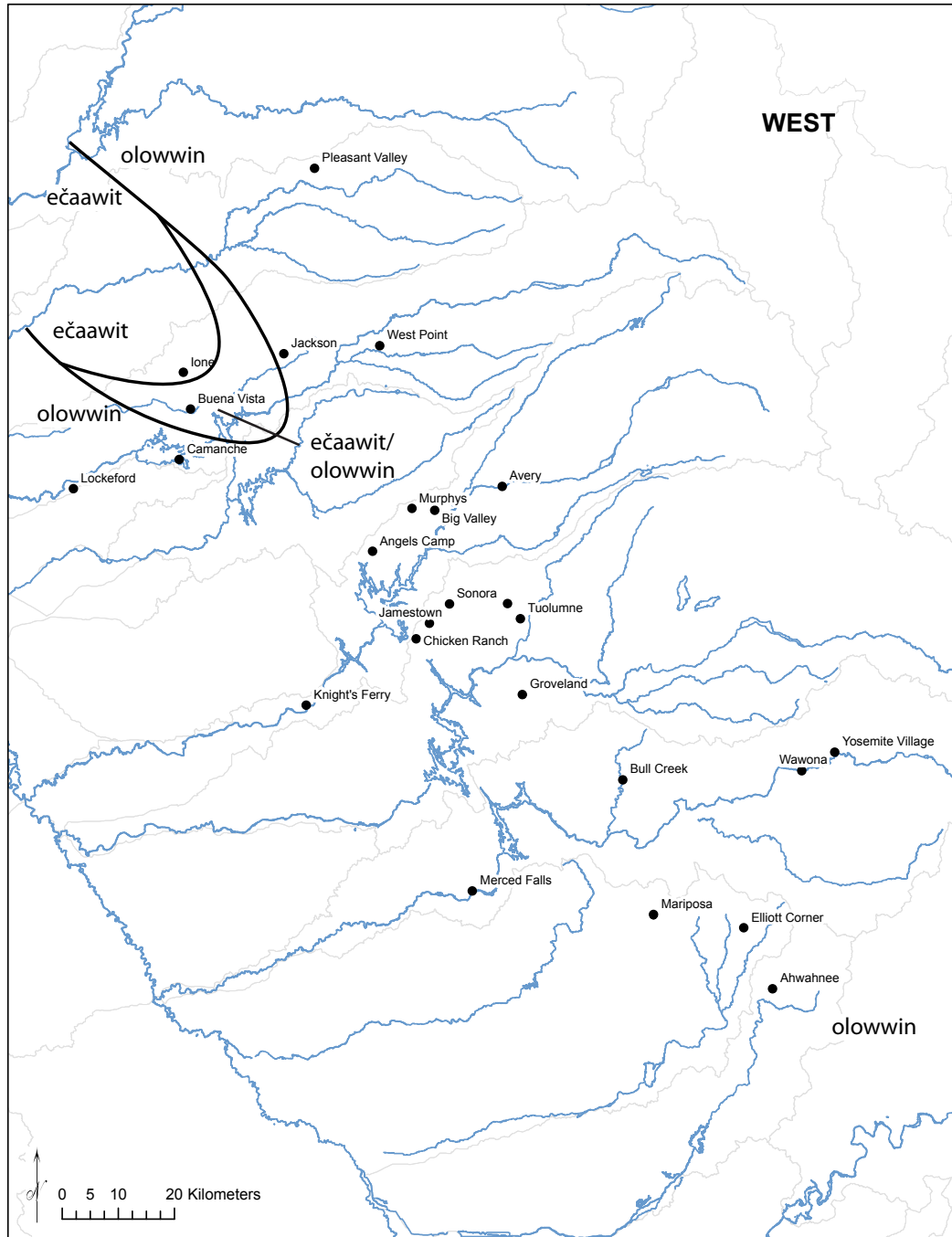


Figure 5.92: Eastern Miwok isogloss map: West

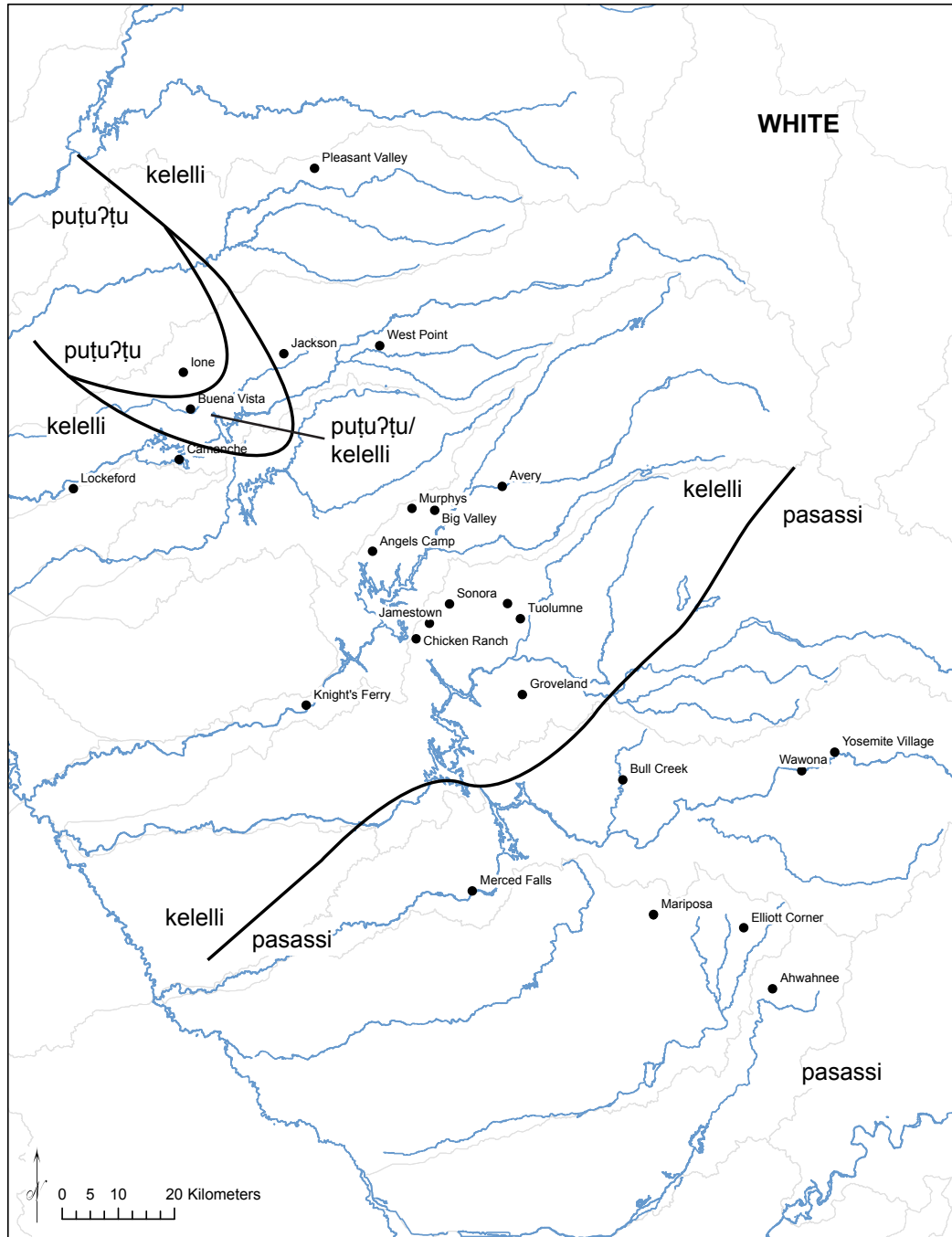


Figure 5.93: Eastern Miwok isogloss map: White

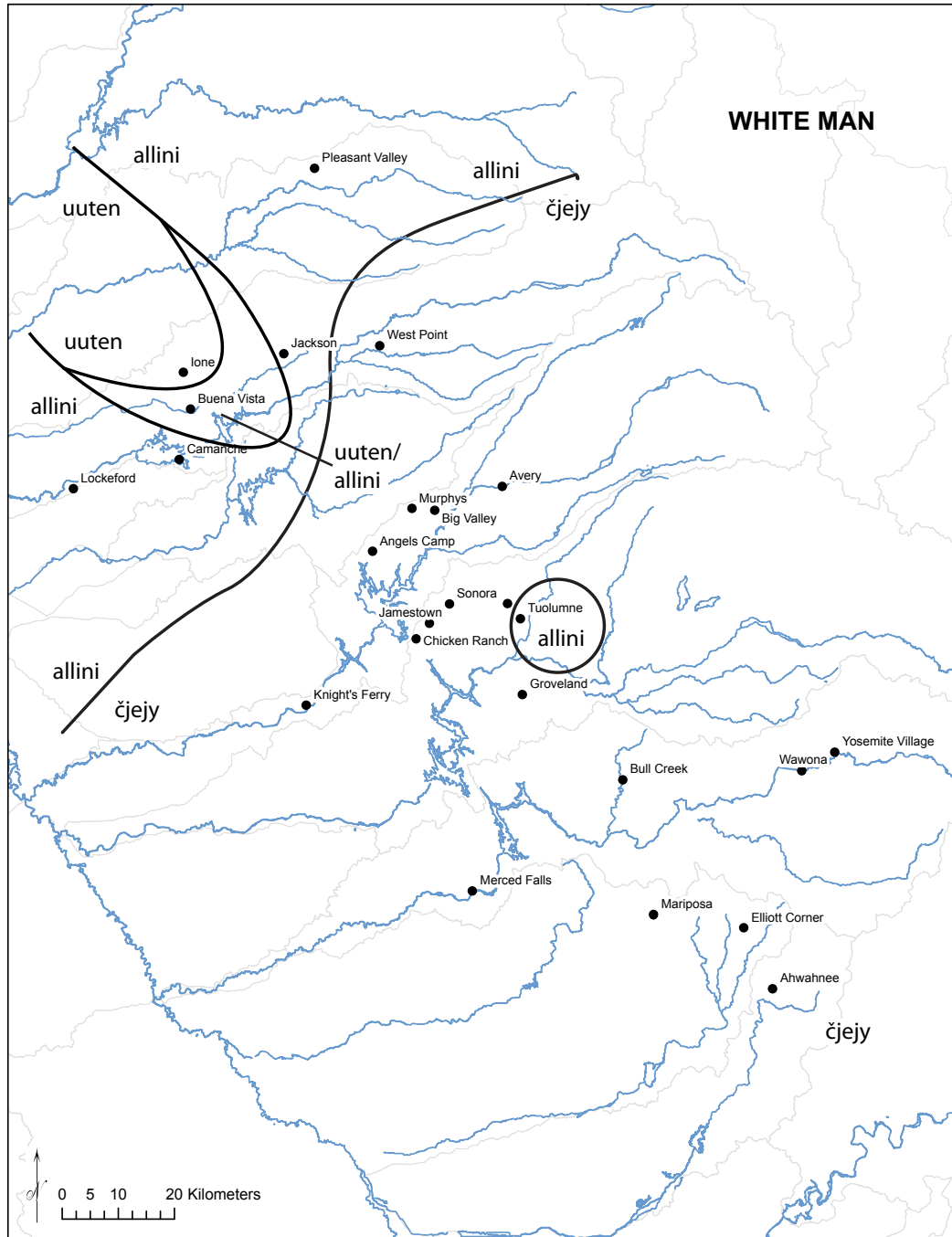


Figure 5.94: Eastern Miwok isogloss map: White Man

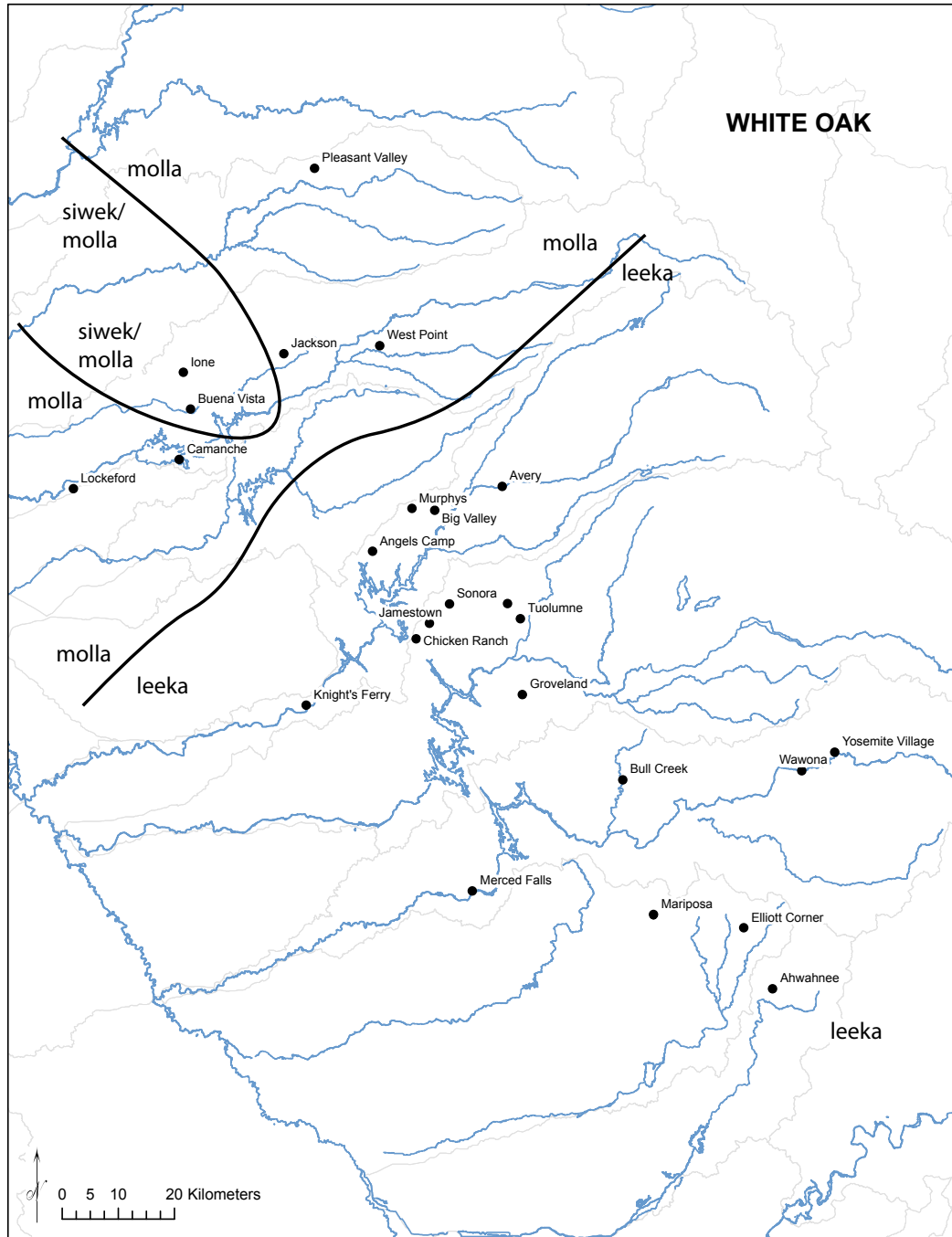


Figure 5.95: Eastern Miwok isogloss map: White Oak

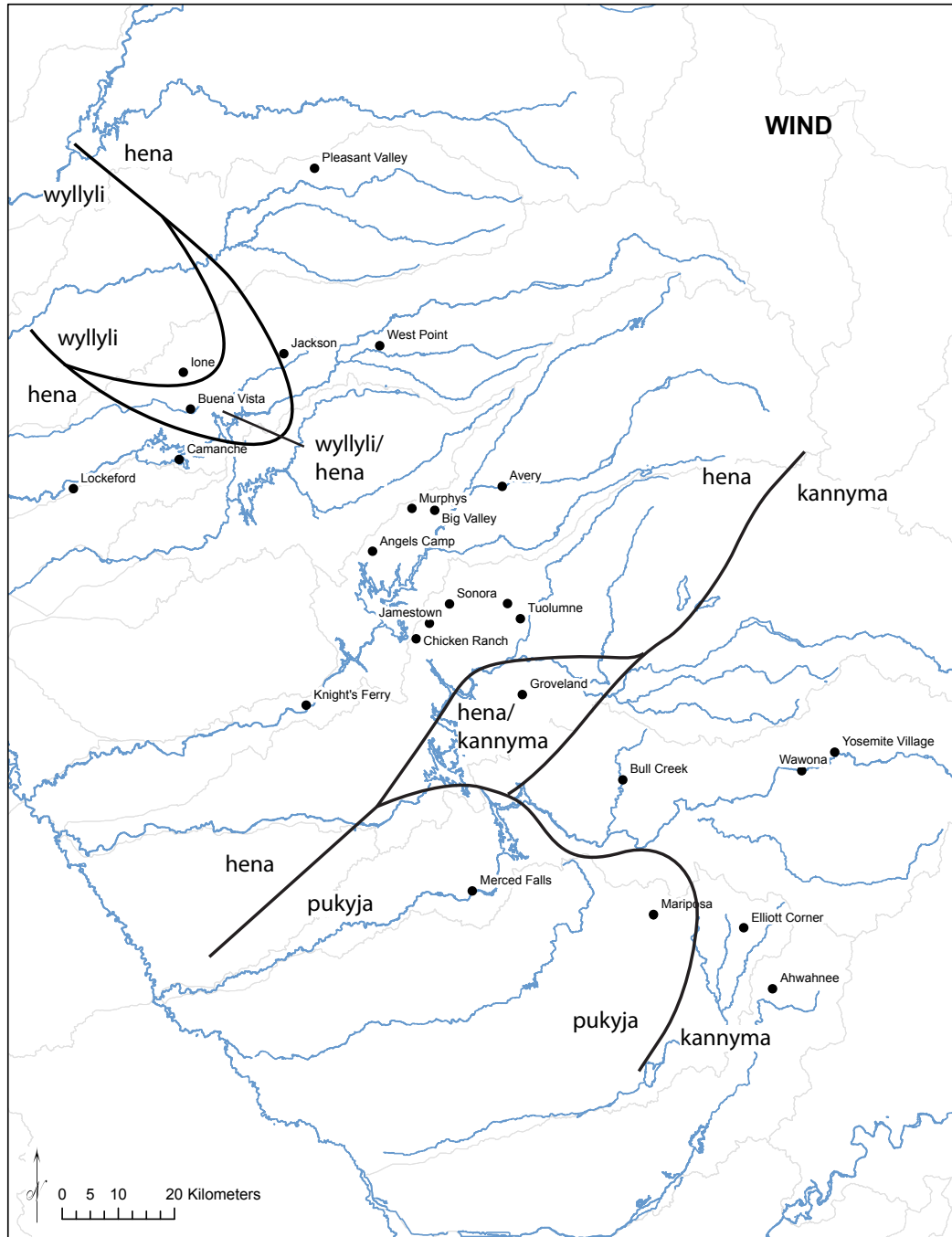


Figure 5.96: Eastern Miwok isogloss map: Wind

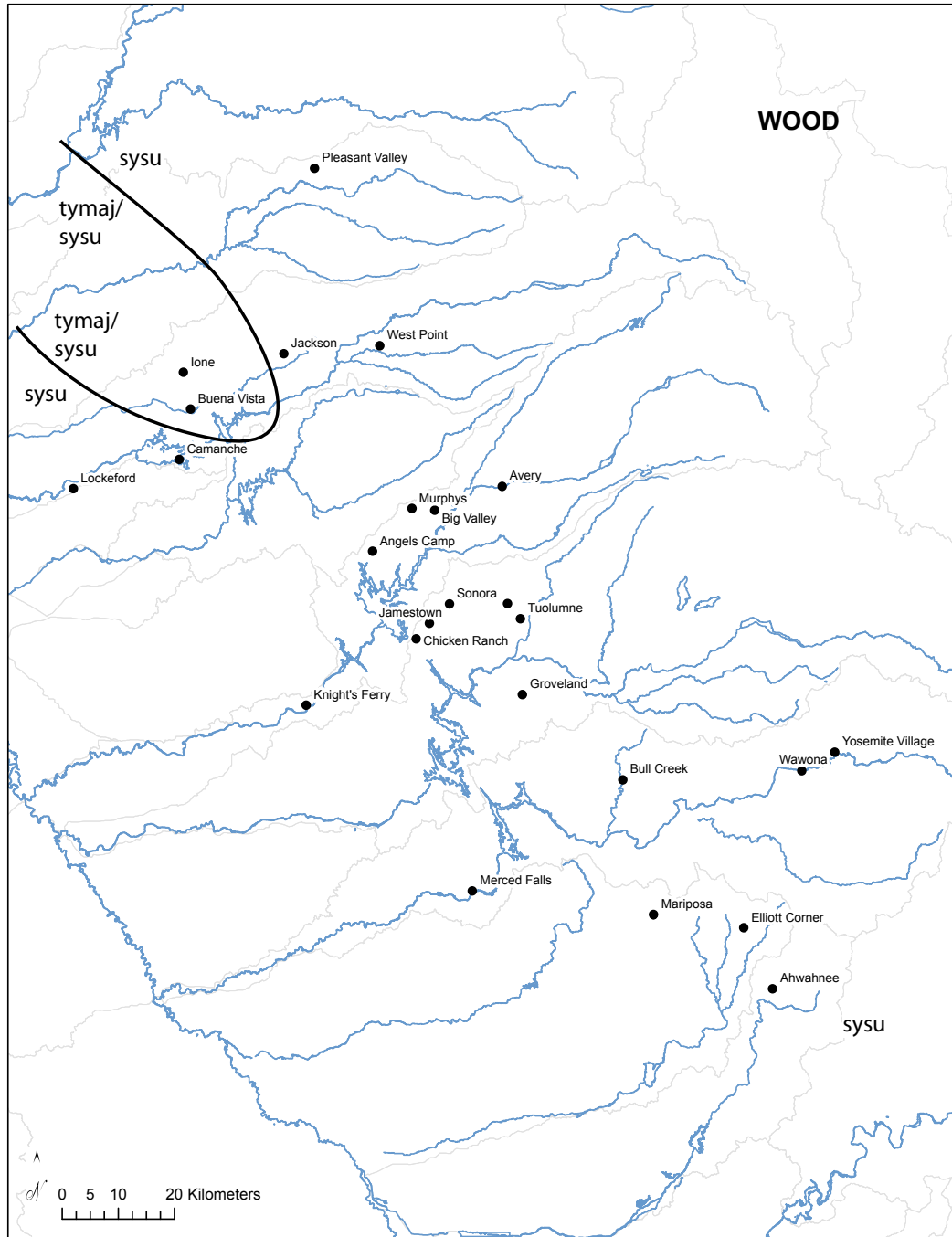


Figure 5.97: Eastern Miwok isogloss map: Wood

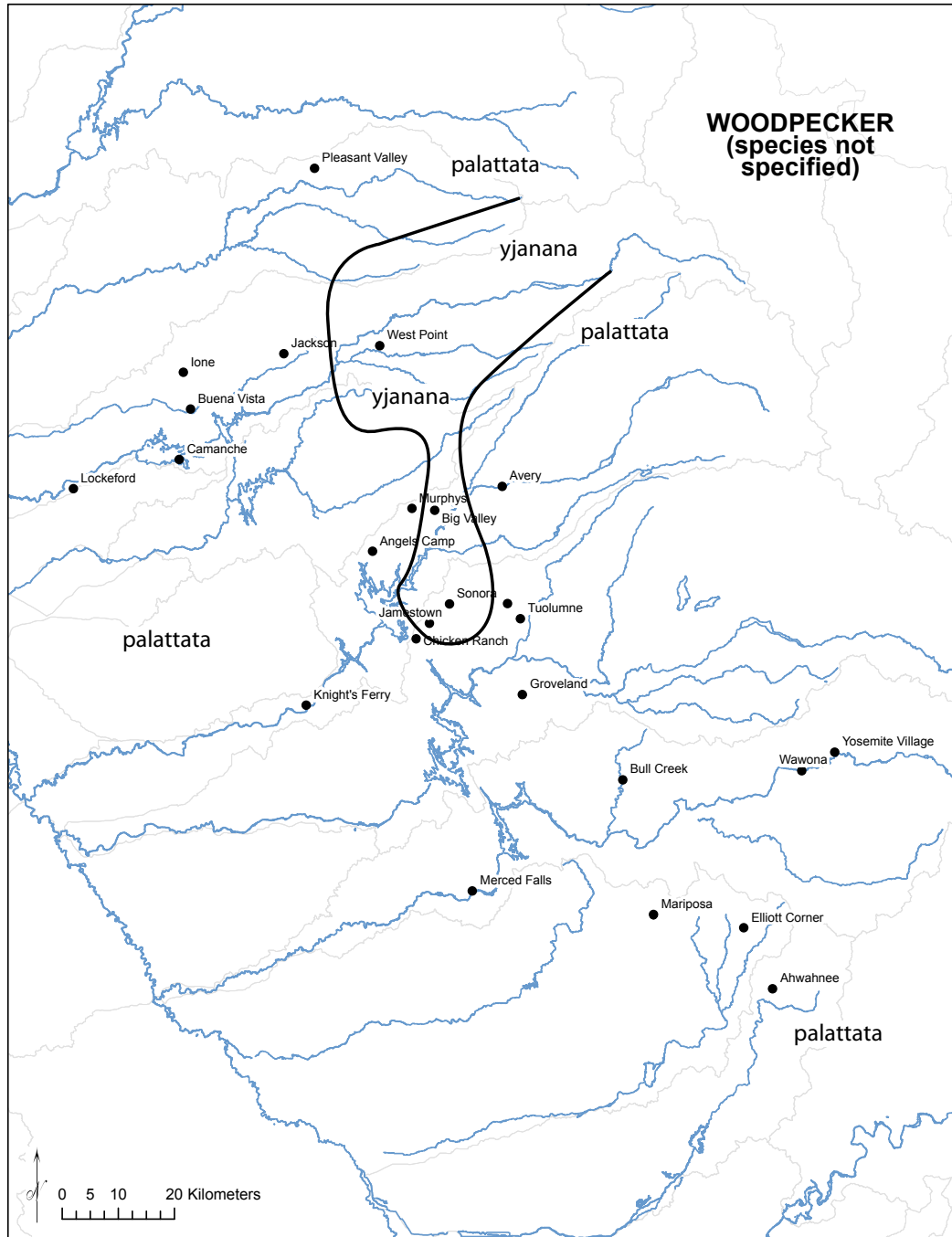


Figure 5.98: Eastern Miwok isogloss map: Woodpecker

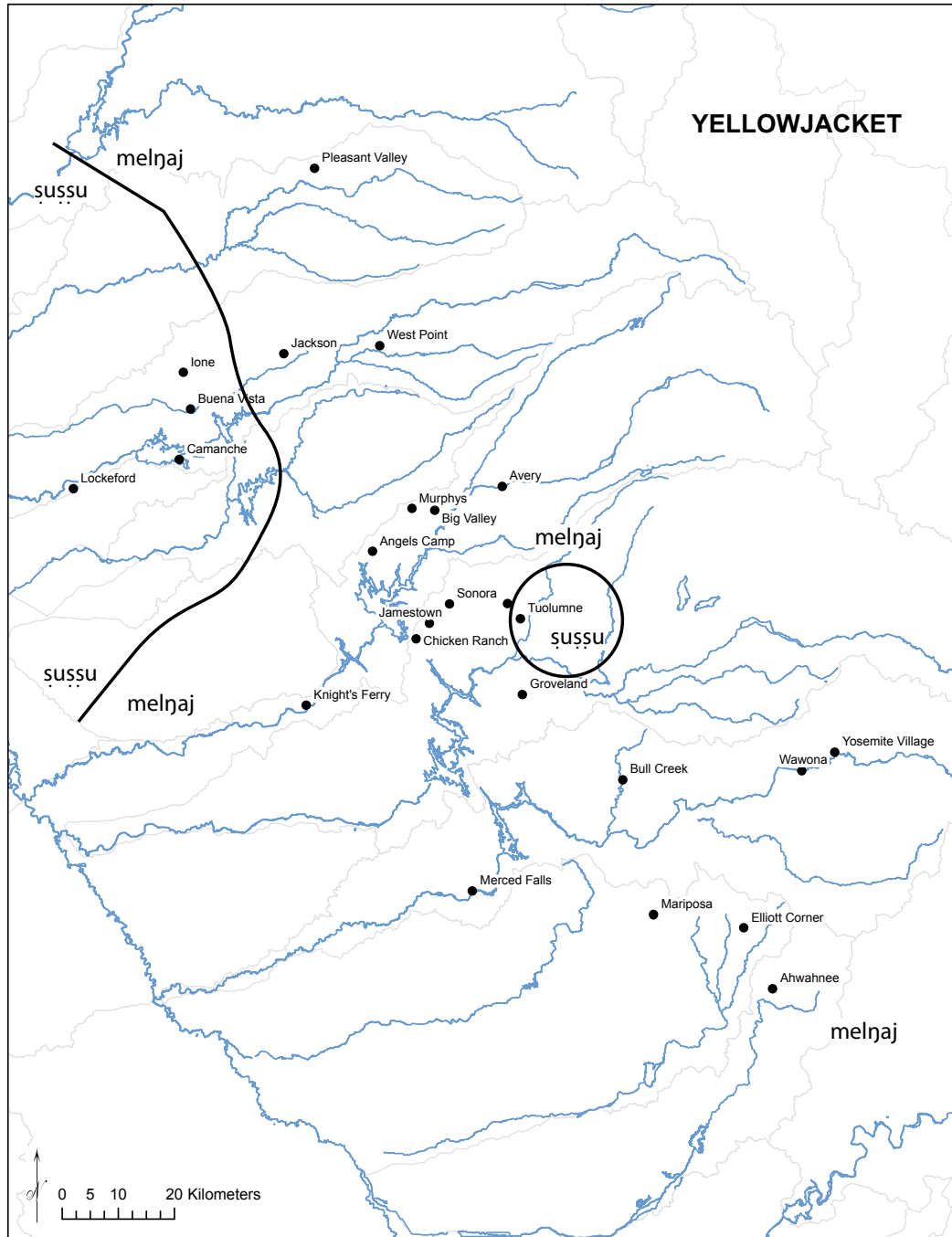


Figure 5.99: Eastern Miwok isogloss map: Yellowjacket

Appendix B

Select Mantel Correlations for Eastern Miwok Language/Geography Relationships

Table 5.1: r values for significant Mantel correlations between linguistic and geographic distances – all Eastern Miwok locations

	Euclidean	Elevation	Vegetation	Surface Water	Watershed-Constr.
Phonological	0.494	0.390	0.486	0.490	0.448
Lexical	0.706	0.603	0.699	0.698	0.662

Table 5.2: r values for significant Mantel correlations between linguistic distance and rejected models of cost distance

	Phonological	Lexical
Elevation + Water	0.375	0.592
Elevation + Vegetation	0.383	0.599
Vegetation + Water	0.485	0.699
Elevation + Water + Vegetation	0.370	0.592

Table 5.3: r values for significant Mantel correlations between phonological distance (based on data with segment length distinctions) and models of geographic distance

	Phonological (Unedited)
Euclidian	0.451
Elevation	0.356
Surface Water	0.448
Vegetation	0.444
Watershed-Constrained	0.408

Appendix C

Expanded Kessler's R^2 Results

Note: In the following tables, an asterisk (*) is used to mark values significant at the $p < 0.005$ level.

[illegible]

Table 5.4: Multilateral clustering: Round 1

	Ute	0.967	1.945	1.227	1.085	0.926	1.569	0.959	0.613	1.165	1.589	1.722	1.972	2.141	0.688	1.106	1.982	0.523	1.593	0.441	Yana
	Tümpisa Shoshone		2.952*	0.899	5.934*	1.143	2.655*	1.606*	1.370	1.103	1.271	1.336	1.234	1.414	1.531	0.949	0.934	1.215	1.014	1.205	Shasta
	Wintu			0.857	3.179*	1.131	4.486*	1.794*	1.583*	0.874	0.803	0.895	1.138	1.064	1.047	1.071	0.918	0.872	1.046	1.032	Northern Sierra Miwok
	Achumawi				0.836	1.148	0.896	1.148	1.055	1.008	0.851	1.032	1.281	1.249	1.046	1.134	0.988	0.918	1.512*	1.040	Diegueño
	Kawaiisu					1.045	3.778*	1.909*	1.259	0.879	1.086	0.990	1.042	1.381	1.141	1.004	1.398	1.167	1.169	1.051	Karuk
	Mono						1.075	0.894	1.107	0.721	0.958	1.202	0.855	0.974	0.784	0.972	0.816	0.976	1.091	1.052	Mono
	Hopi							1.314	1.362	0.684	1.004	1.122	1.041	0.919	1.044	0.934	0.938	0.738	1.006	0.681	Hopi
	Luiseno								1.489	1.149	1.173	0.898	1.520	1.129	1.066	1.040	0.893	0.812	1.079	1.028	Luiseno
	Lake Miwok									1.101	1.403	0.905	0.989	0.683	1.219	1.064	1.092	1.072	0.992	1.027	Lake Miwok
	Cocopa										0.819	0.825	0.866	1.017	1.076	1.182	1.319	2.201*	1.112	0.981	Cocopa
	Kashaya											0.854	1.062	1.172	1.222	1.084	1.716*	1.006	1.147	0.692	Kashaya
	Northern Pomo												2.730*	1.762*	0.567	1.229	1.041	0.847	1.105	0.797	Northern Pomo
	Southeastern Pomo													2.868*	0.873	0.976	1.269	0.878	1.211	1.254	Southeastern Pomo
	Wintu														1.915	1.161	1.276	1.147	1.007	0.703	Wintu
	Atsugewi															1.150	1.257	0.973	1.466*	0.711	Atsugewi
	Diegueño																1.459	1.104	1.399*	1.006	Diegueño
	Northern Sierra Miwok																	1.031	1.324*	0.941	Northern Sierra Miwok
	Shasta																		1.116	0.651	Shasta

Table 5.5: Multilateral clustering: Round 2

[illegible]

[illegible]

[illegible]

[illegible]

Yana		0.277		0.613		0.446		1.039		1.055		1.031		1.028		0.980		0.690		0.709		1.006		0.947		0.651		1.128
Shasta		1.151		1.2.96		1.591		1.512*		1.091		1.080		0.992		1.112		1.145		1.468*		1.401*		1.321*		1.117		
Northern Sierra Miwok		1.064		0.780		0.524		0.917		0.975		0.812		1.072		2.197*		1.006		0.971		1.103		1.033				
Diegueño		0.956		1.071		2.003		0.988		0.817		0.893		1.096		1.317		1.720*		1.253		1.459						
Atsugewi		0.908		1.297		1.104		1.129		0.973		1.041		1.066		1.184		1.084		1.154								
Wintu		0.654		1.441		0.692		1.049		0.785		1.066		1.216		1.080		1.220										
Cocopa		0.888		0.871		1.581		0.850		0.958		1.170		1.041		0.819												
Lake Miwok		0.583		1.102		1.169		1.009		0.719		1.144		1.104														
Luiseno		0.948		2.310*		0.610		1.052		1.107		1.486																
Hopi		0.721		1.907		0.961		1.146		0.893																		
Karuk		0.821		0.683		0.926		1.143																				
Achumawi		1.009		0.953		1.224																						
Maidu + Nisenan		1.645		0.733																								
((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono		1.874																										
	(Northern Pomo + South- eastern Pomo) + Kashaya																											
	((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono																											
	Maidu + Nisenan																											
	Achumawi																											
	Karuk																											
	Hopi																											
	Luiseno																											
	Lake Miwok																											
	Cocopa																											
	Wintu																											
	Atsugewi																											
	Diegueño																											
	Northern Sierra Miwok																											
	Shasta																											

Table 5.10: Multilateral clustering: Round 7

[illegible]

Yana		0.000		0.274		0.448		1.040		1.053		0.975		0.695		0.714		1.008		0.944		0.647		1.127
Shasta		0.992		1.153		1.587		1.514*		1.090		1.112		1.150		1.466*		1.400*		1.322*		1.115		
Northern Sierra Miwok		1.820		1.066		0.529		0.915		0.973		2.203*		1.006		0.972		1.099		1.030				
Diegueño		0.365		0.955		1.999		0.989		0.817		1.320		1.718*		1.248		1.455						
Atsugewi		1.009		0.915		1.100		1.130		0.975		1.183		1.083		1.152								
Wintu		1.149		0.655		0.691		1.045		0.785		1.074		1.223										
Cocopa		1.118		0.884		1.595		0.848		0.958		0.817												
Lake Miwok		0.947		0.582		1.160		1.007		0.717														
Karuk		0.778		0.826		0.930		1.144																
Achumawi		1.519		1.011		1.227																		
Maidu + Nisenan		0.000		1.616																				
(Northern Pomo + Southeastern Pomo) + Kashaya		2.278																						
	(((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono) + Hopi + Luiseno																							
	(Northern Pomo + South-eastern Pomo) + Kashaya																							
	Maidu + Nisenan																							
	Achumawi																							
	Karuk																							
	Lake Miwok																							
	Cocopa																							
	Wintu																							
	Atsugewi																							
	Diegueño																							
	Northern Sierra Miwok																							
	Shasta																							

Table 5.12: Multilateral clustering: Round 9

Yana	0.582	0.000	0.275	0.443	1.040	1.052	0.695	0.712	1.005	0.939	1.128
Shasta	2.224	0.997	1.150	1.594	1.513*	1.090	1.146	1.471*	1.400*	1.322*	
Diegueño	1.185	0.370	0.954	1.997	0.988	0.817	1.716*	1.251	1.456		
Atsugewi	2.258	1.010	0.919	1.103	1.131	0.976	1.087	1.151			
Wintu	0.914	1.146	0.663	0.695	1.046	0.785	1.221				
Cocopa	0.824	1.109	0.884	1.589	0.848	0.960					
Karuk	0.597	0.775	0.819	0.924	1.144						
Achumawi	0.726	1.514	1.017	1.228							
Maidu + Nisenan	0.891	0.000	1.615								
(Northern Pomo + Southeastern Pomo) + Kashaya	0.000	2.288									
((((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono) + Luiseño) + Hopi	1.746										
Lake Miwok + Northern Sierra Miwok											
((((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono) + Luiseño) + Hopi											
(Northern Pomo + South- eastern Pomo) + Kashaya											
Maidu + Nisenan											
Achumawi											
Karuk											
Cocopa											
Wintu											
Atsugewi											
Diegueño											
Shasta											

Table 5.13: Multilateral clustering: Round 10

Yana		0.632	0.582	0.000	0.276	0.447	1.040	1.051	0.717	1.003	1.126
Shasta		2.292	2.214	0.992	1.150	1.589	1.511*	1.092	1.461*	1.400*	
Atsugewi		1.108	2.242	1.002	0.918	1.116	1.129	0.972	1.149		
Wintu		1.917	0.920	1.189	0.651	0.684	1.046	0.785			
Karuk		1.420	0.598	0.773	0.823	0.929	1.146				
Achumawi		1.287	0.728	1.511	1.012	1.226					
Maidu + Nisenan		1.644	0.891	0.000	1.642						
(Northern Pomo + Southeastern Pomo) + Kashaya		0.000	0.000	2.258							
(((((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono) + Luiseño) + Hopi		1.395	1.754								
Lake Miwok + Northern Sierra Miwok		0.000									
	Cocopa + Diegueño										
	Lake Miwok + Northern Sierra Miwok										
	(((((Ute + Kawaiisu) + Tümpisa Shoshone) + Mono) + Luiseño) + Hopi										
	(Northern Pomo + South- eastern Pomo) + Kashaya										
	Maidu + Nisenan										
	Achumawi										
	Karuk										
	Wintu										
	Atsugewi										
	Shasta										

Table 5.14: Multilateral clustering: Round 11

Appendix D

California Word List

all: *Achumawi* tōlo; *Atsugewi* móqja; *Cocopa* kwin^y; *Diegueño* nyaamat; *Hopi* soosoy; *Karuk* koo*; *Kashaya* t̥'i; *Kawaiisu* mono-yo; *Lake Miwok* múʔe; *Luiſeño* čó·ʔun; *Maidu* ʔypék'anbe; *Mono* nasymy*; *Nisenan* ma·no; *Northern Pomo* diley; *Northern Sierra Miwok* ʔoksapa; *Shasta* ʔi·wawaw; *Southeastern Pomo* ʔuxkat; *Tümpisa Shoshone* oyoontü; *Ute* manúni; *Wintu* ko; *Yana* buiʒa..gu..ʔ

back: *Achumawi* ilqōs; *Atsugewi* jʔuʔqit; *Cocopa* ʔimák; *Diegueño* hetat; *Hopi* hoota; *Karuk* vásih; *Kashaya* bac'o; *Kawaiisu* howaa-vi; *Lake Miwok* lúuma; *Luiſeño* má·ča-t; *Maidu* k̥iwí; *Mono* pih; *Nisenan* kiw; *Northern Pomo* c'idi; *Northern Sierra Miwok* kycy·j-y; *Shasta* tú·xu; *Southeastern Pomo* kat; *Tümpisa Shoshone* kwaimpi; *Ute* ʔóáʒ-vi; *Wintu* yay; *Yana* mak'i

bad: *Achumawi* hōqwaī̄; *Atsugewi* sawoqáʔawi; *Cocopa* xcaq; *Diegueño* wellich; *Hopi* nukus; *Karuk* k̥aarim; *Kashaya* pišusu; *Kawaiisu* ʔataa; *Lake Miwok* ʔóbu; *Luiſeño* ʔaláxwi; *Maidu* wasá; *Mono* syhta; *Nisenan* ʔos; *Northern Pomo* tiyiš; *Northern Sierra Miwok* saj·e; *Shasta* k'arik^wícciʔ; *Southeastern Pomo* c'inak'oʔay; *Tümpisa Shoshone* nakütsasa; *Ute* ʔyuvúy-py-ni; *Wintu* c'Ep; *Yana* mallap'a

big: *Achumawi* waʔwā; *Atsugewi* jupáw; *Cocopa* ptay; *Diegueño* 'iikuu; *Hopi* wuko; *Karuk* k̥eetch; *Kashaya* baht̥^he; *Kawaiisu* ʔi-vee; *Lake Miwok* ʔáde; *Luiſeño* muká-t; *Maidu* teté; *Mono* papa; *Nisenan* muk'; *Northern Pomo* maʔo; *Northern Sierra Miwok* ʔeʔ·a; *Shasta* k'impíʔ; *Southeastern Pomo* b̥tenik; *Tümpisa Shoshone* pia; *Ute* ʔavá-t̥j; *Wintu* qom; *Yana* baʒal

bird: *Achumawi* jena; *Atsugewi* jené·stiká; *Cocopa* ša; *Diegueño* 'aashaa; *Hopi* masas'iʔytaqa; *Karuk* achviiv; *Kashaya* s'ihta; *Kawaiisu* wižiki-ži; *Lake Miwok* méle; *Luiſeño* ʔihéj-ma-l; *Maidu* k'útt'yt'ỳ; *Mono* cihipa'; *Nisenan* t'yt'ym*; *Northern Pomo* c'it; *Northern Sierra Miwok* mice-ma; *Shasta* c'i·c'a·x; *Southeastern Pomo* c'ʔa*; *Tümpisa Shoshone* kasattsikantün; *Ute* wici·ci; *Wintu* c'il; *Yana* ʒudʔawi

black: *Achumawi* hakē·j̄i; *Atsugewi* joʔjitáʔwi; *Cocopa* n̥i·t̥^y; *Diegueño* nyilly; *Hopi* qōmvi; *Karuk* ikxáramkunish; *Kashaya* k'ili; *Kawaiisu* tuhu-ki; *Lake Miwok* mulúumulu; *Luiſeño* yuvá-ta; *Maidu* síwsìw; *Mono* tuhmu; *Nisenan* kylm; *Northern Pomo* xac'e; *Northern Sierra Miwok* kulul-i; *Shasta* ʔe·pxuʔtáraxiʔ; *Southeastern Pomo* lq'olq'okin; *Tümpisa Shoshone* tuppapi; *Ute* túu-kwa-r̄j; *Wintu* culu-l; *Yana* pal

blood: *Achumawi* ahti; *Atsugewi* isri; *Cocopa* n̥x^wat; *Diegueño* 'ehwatt; *Hopi* ungwa; *Karuk* áax; *Kashaya* balay; *Kawaiisu* pii-pi; *Lake Miwok* kícaw; *Luiſeño* ʔów-la; *Maidu*

sedé; *Mono* paah; *Nisenan* sedej; *Northern Pomo* balay; *Northern Sierra Miwok* kic·aw-y; *Shasta* ʔáxta; *Southeastern Pomo* blay; *Tümpisa Shoshone* paoʔ; *Ute* páa-pu; *Wintu* sa·q; *Yana* wadʔduwi

bone: *Achumawi* jōje; *Atsugewi* juqji; *Cocopa* ya·k; *Diegueño* aq; *Hopi* öoqa; *Karuk* ípih; *Kashaya* ʔihya; *Kawaiisu* ʔoho-vi; *Lake Miwok* kúlum; *Luiŕeño* kulá-wu-t; *Maidu* býmí; *Mono* ʔoho; *Nisenan* bym; *Northern Pomo* ya; *Northern Sierra Miwok* kyc·yc-y; *Shasta* ʔá·kʔ; *Southeastern Pomo* ya; *Tümpisa Shoshone* tsuhmippüh; *Ute* ʔqó-vu; *Wintu* paq; *Yana* ʔixdalʔ

breathe: *Achumawi* ināht; *Atsugewi* ussəts; *Cocopa* ya·ş; *Diegueño* yaas; *Hopi* hiikwista; *Karuk* pimyaáhva; *Kashaya* haʔtʰo; *Kawaiisu* soo-ki; *Lake Miwok* hén; *Luiŕeño* hakwís; *Maidu* hónwèj; *Mono* suwah-ka; *Nisenan* honpin; *Northern Pomo* šixa:čin*; *Northern Sierra Miwok* hen·a; *Shasta* esur*; *Southeastern Pomo* myokit*; *Tümpisa Shoshone* sumakkain; *Ute* sqá-qay; *Wintu* pʰu-r; *Yana* exdurubʔa

cold: *Achumawi* asjē; *Atsugewi* eskap; *Cocopa* xsu-r; *Diegueño* ʔekwiiy; *Hopi* suusungwa; *Karuk* áthiik; *Kashaya* qahsil; *Kawaiisu* šituʔi-di; *Lake Miwok* típmuʔi; *Luiŕeño* pitú-ptu-ş; *Maidu* dúpe; *Mono* ʔycyʔy; *Nisenan* bəkʔəs; *Northern Pomo* kasiliʔ*; *Northern Sierra Miwok* tuʔ·uʔu; *Shasta* ʔissíkʔ; *Southeastern Pomo* ksilit; *Tümpisa Shoshone* üitsüʔin; *Ute* supú-rʔay; *Wintu* tEm; *Yana* xaatʔai

come: *Achumawi* ūnn; *Atsugewi* wiw; *Cocopa* yi·; *Diegueño* weyiw; *Hopi* pew; *Karuk* ;naa *Kashaya* ad*; *Kawaiisu* wa-ki; *Lake Miwok* ʔóni; *Luiŕeño* haqwá·çi; *Maidu* ʔyjé; *Mono* kiʔ; *Nisenan* daw; *Northern Pomo* bilhum*; *Northern Sierra Miwok* ʔy·ny; *Shasta* atuk*; *Southeastern Pomo* xolo; *Tümpisa Shoshone* kimma; *Ute* cawi; *Wintu* hEn; *Yana* mxʒa

cut: *Achumawi* à·kā-t; *Atsugewi* yokʔət; *Cocopa* cukát; *Diegueño* aakatt; *Hopi*uku; *Karuk* ikvit; *Kashaya* ca**; *Kawaiisu* kavi; *Lake Miwok* cʔíkaʔi; *Luiŕeño* wóki; *Maidu* wykʔýs; *Mono* qopa; *Nisenan* wacʔa-t; *Northern Pomo* čaha; *Northern Sierra Miwok* te·py; *Shasta* icut*; *Southeastern Pomo* ʔkʔecʔi; *Tümpisa Shoshone* kaʔah; *Ute* kырáy; *Wintu* dOp; *Yana* mi

die: *Achumawi* iumm; *Atsugewi* pweʔiniú; *Cocopa* mşpa; *Diegueño* melay; *Hopi* mooki; *Karuk* iv; *Kashaya* hcʰoyic*; *Kawaiisu* yuweʔe-kwee; *Lake Miwok* jóok; *Luiŕeño* taqwáya; *Maidu* wóno; *Mono* tyja-wi; *Nisenan* wo·n; *Northern Pomo* kʔala; *Northern Sierra Miwok* camsy; *Shasta* irik*; *Southeastern Pomo* qʔlali; *Tümpisa Shoshone* tiyaih; *Ute* yaʔáy; *Wintu* min; *Yana* micʔsaa

dig: *Achumawi* ʔsay; *Atsugewi* ehxy; *Cocopa* ʔu·pún; *Diegueño* wehwall; *Hopi* hangwa; *Karuk* írip; *Kashaya* hal**; *Kawaiisu* horo; *Lake Miwok* hém; *Luiŕeño* héya; *Maidu* bá; *Mono* tyhoi; *Nisenan* syk; *Northern Pomo* hele; *Northern Sierra Miwok* ʔol·u; *Shasta* icmai*; *Southeastern Pomo* ʔdukʔi; *Tümpisa Shoshone* hotaʔ; *Ute* ʔqráy; *Wintu* hap; *Yana* ʔaa

dog: *Achumawi* āʔlāʔmoki; *Atsugewi* hoʔma; *Cocopa* xaʔ; *Diegueño* ʔehatt; *Hopi* pooko; *Karuk* chishiih; *Kashaya* hayu; *Kawaiisu* pugu; *Lake Miwok* háju; *Luiŕeño* ʔawá·l; *Maidu* sý; *Mono* puhku; *Nisenan* sukku; *Northern Pomo* hayu; *Northern Sierra Miwok* cuku; *Shasta* ʔá·psu; *Southeastern Pomo* hayu; *Tümpisa Shoshone* isapungku; *Ute* sarí·ci; *Wintu* suk; *Yana* suusu

drink: *Achumawi* iss; *Atsugewi* i·jy; *Cocopa* şi; *Diegueño* wesii; *Hopi* hiiko; *Karuk* ish; *Kashaya* ʔqʔo*; *Kawaiisu* hivi; *Lake Miwok* ʔuşşu; *Luiŕeño* pá·ʔi; *Maidu* mo; *Mono* hipi; *Nisenan* mo·; *Northern Pomo* kʔocim; *Northern Sierra Miwok* ʔuhu; *Shasta* ic*; *Southeastern*

Pomo bc'ak; *Tümpisa Shoshone* hipi"; *Ute* ʔiví; *Wintu* bOl; *Yana* sii

dry: *Achumawi* wi-nā-suʔsi; *Atsugewi* mé-qwé; *Cocopa* ʃʔar; *Diegueño* saay; *Hopi* lakpu; *Karuk* iváxrah; *Kashaya* s'uwac*; *Kawaiisu* tavasi-kwee-pi; *Lake Miwok* káaj; *Luiसेño* ʔawáxa; *Maidu* p'ík'ál; *Mono* pa-hsa-hky'; *Nisenan* k'u; *Northern Pomo* ʔ'aboʔ; *Northern Sierra Miwok* hewe; *Shasta* k'iccur*; *Southeastern Pomo* k'bokin; *Tümpisa Shoshone* pasa"; *Ute* tavási; *Wintu* xOn; *Yana* kul

ear: *Achumawi* īsat; *Atsugewi* asmak; *Cocopa* šma-l; *Diegueño* hemall; *Hopi* naqvu; *Karuk* tīiv; *Kashaya* šima; *Kawaiisu* naga-vi-vi; *Lake Miwok* ʔálok; *Luiसेño* náq-la; *Maidu* bonó; *Mono* nahqa; *Nisenan* bono·; *Northern Pomo* šima; *Northern Sierra Miwok* ʔo-kos-u; *Shasta* ʔisa-k'; *Southeastern Pomo* xmanca; *Tümpisa Shoshone* nangki; *Ute* nʔká-vi; *Wintu* ma-t; *Yana* malʔgu

earth: *Achumawi* tī-qāʔtē; *Atsugewi* tayh'əq; *Cocopa* maʔ; *Diegueño* 'emat; *Hopi* tut-skwa; *Karuk* ithívthaaneen; *Kashaya* ʔama; *Kawaiisu* tii-pi; *Lake Miwok* jówa; *Luiसेño* ʔéx-la; *Maidu* k'awí; *Mono* ty-pih; *Nisenan* k'aw*; *Northern Pomo* ma; *Northern Sierra Miwok* wal-i; *Shasta* t'arak; *Southeastern Pomo* xqo; *Tümpisa Shoshone* sokopin; *Ute* tʔvú-pu; *Wintu* pom; *Yana* biiwi

eat: *Achumawi* āmm; *Atsugewi* yəmm; *Cocopa* n'awí; *Diegueño* wesaaw; *Hopi* nōösa; *Karuk* av; *Kashaya* maʔa*; *Kawaiisu* kaʔa; *Lake Miwok* jólum; *Luiसेño* kwáʔ; *Maidu* pe; *Mono* tyhka; *Nisenan* pa; *Northern Pomo* maʔan; *Northern Sierra Miwok* tuj-u; *Shasta* icku*; *Southeastern Pomo* qwala; *Tümpisa Shoshone* tükka; *Ute* tʔkáy; *Wintu* b; *Yana* ma

eye: *Achumawi* asā; *Atsugewi* úyi; *Cocopa* ʔiyú; *Diegueño* eyiiw; *Hopi* poosi; *Karuk* yúup; *Kashaya* huʔuy; *Kawaiisu* puʔi-vi; *Lake Miwok* šút; *Luiसेño* pús-la; *Maidu* hini; *Mono* puh; *Nisenan* hi-n; *Northern Pomo* ʔuy; *Northern Sierra Miwok* syt-y; *Shasta* ʔwí; *Southeastern Pomo* ʔuy; *Tümpisa Shoshone* pui; *Ute* puʔi-vu; *Wintu* tu-m; *Yana* cu

fall: *Achumawi* anajqāl; *Atsugewi* u-t; *Cocopa* pam; *Diegueño* wenall; *Hopi* poosi; *Karuk* ikváayrin; *Kashaya* hc^ha**"; *Kawaiisu* wiʔi-ku; *Lake Miwok* ʔút; *Luiसेño* hulúka; *Maidu* wodá; *Mono* wyʔi; *Nisenan* bo; *Northern Pomo* lok'; *Northern Sierra Miwok* ly-mehny; *Shasta* uc*; *Southeastern Pomo* člaki; *Tümpisa Shoshone* hapikkwan; *Ute* wíʔh; *Wintu* dil; *Yanamid*ʔ

far: *Achumawi* pōjyoī; *Atsugewi* haquʔohó·; *Cocopa* mātʔxmí-rʔy; *Diegueño* 'ekur; *Hopi* haq; *Karuk* yiiv; *Kashaya* bahcil; *Kawaiisu* miho; *Lake Miwok* ʔedáak; *Luiसेño* wá-m; *Maidu* hadá; *Mono* qwenaʔa; *Nisenan* lamdi; *Northern Pomo* čadil; *Northern Sierra Miwok* tot-oʔ; *Shasta* ʔuk^wák; *Southeastern Pomo* maʔbcil; *Tümpisa Shoshone* manakwappüh; *Ute* míyh; *Wintu* kel; *Yana* pau

fight: *Achumawi* ahti-wa; *Atsugewi* wo-wí-w; *Cocopa* nʔup; *Diegueño* menyuup; *Hopi* naayawi; *Karuk* vathiv; *Kashaya* qamuʔ***; *Kawaiisu* na-paka; *Lake Miwok* wóocakʂumʔi; *Luiसेño* néqpi; *Maidu* hóm̐pajto; *Mono* pihty hky'; *Nisenan* wodoj; *Northern Pomo* dibemu*; *Northern Sierra Miwok* myl-a; *Shasta* akahe*; *Southeastern Pomo* c'imk'ol; *Tümpisa Shoshone* napitüngkūn; *Ute* naǵó-qoʔáy; *Wintu* ʔik; *Yana* ʔiwau

fire: *Achumawi* mällis; *Atsugewi* wo-jyají; *Cocopa* ʔaʔá; *Diegueño* 'aaw; *Hopi* uuwingw; *Karuk* áah; *Kashaya* ʔoho; *Kawaiisu* kuna; *Lake Miwok* wíki; *Luiसेño* kú-t; *Maidu* sá; *Mono* qoso* so; *Nisenan* sa; *Northern Pomo* ho; *Northern Sierra Miwok* wyke; *Shasta* ʔimmá; *Southeastern Pomo* xo; *Tümpisa Shoshone* wayantün; *Ute* naʔáy-tʔ; *Wintu* p^ho-h; *Yana* ʔau

flow: *Achumawi* ajum; *Atsugewi* juwo-mí; *Cocopa* kup; *Diegueño* wenuw; *Hopi* muuna;

Karuk thúufhi; *Kashaya* hwoq**; *Kawaiisu* nukwi; *Lake Miwok* híccuw; *Luiŕeño* yá?; *Maidu* hínk'oj; *Mono* huu; *Nisenan* jo-no; *Northern Pomo* banema; *Northern Sierra Miwok* hu·ju; *Shasta* ek'ira**; *Southeastern Pomo* cali; *Tümpisa Shoshone* okwe"; *Ute* tųvwá-ğay; *Wintu* c'Oh; *Yana* ba

fly: *Achumawi* ata·; *Atsugewi* pijilímmu; *Cocopa* man; *Diegueño* weman; *Hopi* puualti; *Karuk* ikxip; *Kashaya* ca**; *Kawaiisu* yozi; *Lake Miwok* lílu; *Luiŕeño* wí·la; *Maidu* káj; *Mono* johci; *Nisenan* hu·; *Northern Pomo* p^hit'am; *Northern Sierra Miwok* syle·t·y; *Shasta* uh*; *Southeastern Pomo* flali; *Tümpisa Shoshone* yücü"; *Ute* yucí; *Wintu* t^hEw; *Yana* bahbil

foot: *Achumawi* jikoh; *Atsugewi* jujki; *Cocopa* ?imí·; *Diegueño* emily; *Hopi* kuku; *Karuk* fíthih; *Kashaya* q^hama; *Kawaiisu* nabi-vi; *Lake Miwok* kólo; *Luiŕeño* ?é-t; *Maidu* pàyi; *Mono* kyhky; *Nisenan* paj; *Northern Pomo* ɣama; *Northern Sierra Miwok* kolo; *Shasta* ?ák'us; *Southeastern Pomo* xman; *Tümpisa Shoshone* nampe; *Ute* nápa; *Wintu* ma; *Yana* lal

four: *Achumawi* ha?tāmē; *Atsugewi* ha?qaw; *Cocopa* spap; *Diegueño* chepap; *Hopi* naalöyöm; *Karuk* piith; *Kashaya* mihca; *Kawaiisu* wa-cuu-yu; *Lake Miwok* ?otótta; *Luiŕeño* wasá?; *Maidu* c'ýji; *Mono* wacy; *Nisenan* c'yj; *Northern Pomo* ɬak'; *Northern Sierra Miwok* ?oj-is-a; *Shasta* ?iraha·ya; *Southeastern Pomo* daq'o; *Tümpisa Shoshone* wattsüwi; *Ute* whcúwiini; *Wintu* l'aw; *Yana* daumi

give: *Achumawi* aw; *Atsugewi* se?ray; *Cocopa* i·; *Diegueño* winy; *Hopi* maqa; *Karuk* êeh; *Kashaya* dihqa**; *Kawaiisu* maga; *Lake Miwok* hína; *Luiŕeño* ?óvi; *Maidu* méj; *Mono* kija; *Nisenan* bo; *Northern Pomo* dikan; *Northern Sierra Miwok* ?am·y; *Shasta* icu*; *Southeastern Pomo* ?ka; *Tümpisa Shoshone* uttuh; *Ute* tugwáy; *Wintu* dOy; *Yana* halɣamaa

good: *Achumawi* tūsŕi; *Atsugewi* wussai; *Cocopa* px^way; *Diegueño* 'ehan; *Hopi* lolma; *Karuk* yav; *Kashaya* q'o?di; *Kawaiisu* pišaa; *Lake Miwok* ?eméene; *Luiŕeño* ló·vi; *Maidu* jahá; *Mono* cawu; *Nisenan* wen; *Northern Pomo* k'idi; *Northern Sierra Miwok* kuci; *Shasta* k'arisá?; *Southeastern Pomo* c'ma; *Tümpisa Shoshone* tsao; *Ute* ?á-tj; *Wintu* cal; *Yana* c'ub?

green: *Achumawi* mäsūqatī; *Atsugewi* psu?qəth; *Cocopa* xpsiw; *Diegueño* hepeshiw; *Hopi* mokingpu; *Karuk* píriškunish; *Kashaya* s'aq^hala; *Kawaiisu* puhi-gi; *Lake Miwok* šiwišiw; *Luiŕeño* kunó; *Maidu* títtit; *Mono* puhi; *Nisenan* koc'is*; *Northern Pomo* c'axat*; *Northern Sierra Miwok* co-ki; *Shasta* ?e-ricumpaxi?**; *Southeastern Pomo* c'yac'yakin; *Tümpisa Shoshone* sakwaapi; *Ute* sağwá-ğa-rj; *Wintu* c'ar; *Yana* c'gai

hand: *Achumawi* il; *Atsugewi* éy?ti; *Cocopa* ?išá·I^y; *Diegueño* esally; *Hopi* maa; *Karuk* tiiik; *Kashaya* cohšo; *Kawaiisu* mo?o-vi; *Lake Miwok* ?úkku; *Luiŕeño* má-t; *Maidu* má; *Mono* mah; *Nisenan* ma·; *Northern Pomo* ɬ^hana; *Northern Sierra Miwok* ?uk-us-u; *Shasta* ?ápka; *Southeastern Pomo* ?tan; *Tümpisa Shoshone* mo'o; *Ute* mō?ǵ-vi; *Wintu* se; *Yana* dal

head: *Achumawi* lāh; *Atsugewi* náha; *Cocopa* mkur; *Diegueño* hellytaa; *Hopi* qötö; *Karuk* axvāah; *Kashaya* ho?t'o; *Kawaiisu* toci-vi; *Lake Miwok* cáanna; *Luiŕeño* yú·la; *Maidu* ?onó; *Mono* wooh; *Nisenan* c'ol; *Northern Pomo* šina; *Northern Sierra Miwok* han·a; *Shasta* ca·ráw; *Southeastern Pomo* xya; *Tümpisa Shoshone* pampi; *Ute* tųcí-vi; *Wintu* p^hoyoq; *Yana* t'a?lai

hear: *Achumawi* i?tēw; *Atsugewi* kuka; *Cocopa* í-p?i·?; *Diegueño* yip; *Hopi* navota; *Karuk* thitiv; *Kashaya* šoc**; *Kawaiisu* naa-kee; *Lake Miwok* ?álu; *Luiŕeño* náqma; *Maidu* pín; *Mono* nahqa; *Nisenan* pin; *Northern Pomo* šoči; *Northern Sierra Miwok* ɬolo-j-u; *Shasta* ikhyawik**; *Southeastern Pomo* xko; *Tümpisa Shoshone* nangkah; *Ute* nykáy; *Wintu* mut; *Yana* gaa

heart: *Achumawi* haraʔji; *Atsugewi* yopoʔáyti; *Cocopa* ʔiyá-y; *Diegueño* iichih; *Hopi* unangwa; *Karuk* imyah; *Kashaya* s'uhkul; *Kawaiisu* pihyi-pi; *Lake Miwok* c'íddidik; *Lui-seño* şún-la; *Maidu* hòní; *Mono* piwy; *Nisenan* hon; *Northern Pomo* xam; *Northern Sierra Miwok* wyski; *Shasta* ʔiwa-súrʔ; *Southeastern Pomo* ʔ'ukut*; *Tümpisa Shoshone* pihwün; *Ute* mugúa-vi; *Wintu* p^hu-r; *Yana* zuguc'i

hit: *Achumawi* atqāl; *Atsugewi* túp'əq; *Cocopa* ştu; *Diegueño* chetoo; *Hopi* wuvaata; *Karuk* iik; *Kashaya* p^hak'uom**; *Kawaiisu* tono; *Lake Miwok* múla; *Lui-seño* şápa; *Maidu* wódato; *Mono* toh-pakita; *Nisenan* wə; *Northern Pomo* phabačin; *Northern Sierra Miwok* ʔokla; *Shasta* ahiʔ; *Southeastern Pomo* nema; *Tümpisa Shoshone* tükwan; *Ute* ʔoʔnáy; *Wintu* hEp; *Yana* hal

hold: *Achumawi* ijasāntūw; *Atsugewi* wawasw; *Cocopa* şawín; *Diegueño* weyuuw; *Hopi* yawta; *Karuk* áhariv; *Kashaya* c^ha:**; *Kawaiisu* yawi; *Lake Miwok* hílaw; *Lui-seño* yá-w; *Maidu* ménu; *Mono* cee; *Nisenan* sido-do; *Northern Pomo* ; *Northern Sierra Miwok* ʔeteksy; *Shasta* aʔame-; *Southeastern Pomo* celi; *Tümpisa Shoshone* tsaikkan; *Ute* caʔáy; *Wintu* c'ak; *Yana* auwi

hunt: *Achumawi* ò-s; *Atsugewi* mahjiw; *Cocopa* n^yay; *Diegueño* wenyay; *Hopi* maqto; *Karuk* ákunva; *Kashaya* šiʔbi***; *Kawaiisu* tiyahe; *Lake Miwok* ʔónah; *Lui-seño* nó-ti; *Maidu* múhùn; *Mono* qoih nywi; *Nisenan* hun; *Northern Pomo* buʔun*; *Northern Sierra Miwok* wynty; *Shasta* imm*; *Southeastern Pomo* bo; *Tümpisa Shoshone* wasüwükki; *Ute* ʔynáay; *Wintu* k'Od; *Yana* hi

ice: *Achumawi* lāʔqaj; *Atsugewi* p'əʔturi; *Cocopa* smis; *Diegueño* shemetii; *Hopi* patusngwa; *Karuk* ikákrih; *Kashaya* ʔihyu; *Kawaiisu* pa-raʔasi-pi; *Lake Miwok* k'íiwil; *Lui-seño* tó-yi-t; *Maidu* ʔejý; *Mono* tyh-ʔahsyh; *Nisenan* killi-t; *Northern Pomo* yu; *Northern Sierra Miwok* sus-a; *Shasta* ʔiyúʔaʔ; *Southeastern Pomo* mam; *Tümpisa Shoshone* patüasüppüh; *Ute* paráʔsi-py; *Wintu* kik; *Yana* daarik'u

kill: *Achumawi* ā-tw; *Atsugewi* pwəhn; *Cocopa* nak; *Diegueño* aamuuch; *Hopi* niina; *Karuk* iykar; *Kashaya* p^hak'um**; *Kawaiisu* paka; *Lake Miwok* kátt; *Lui-seño* móknu; *Maidu* wónoti; *Mono* pahca; *Nisenan* wo-nti; *Northern Pomo* čaban; *Northern Sierra Miwok* jin-a; *Shasta* ic*; *Southeastern Pomo* mdoka; *Tümpisa Shoshone* pakkah; *Ute* paqǵáy; *Wintu* l'o-m; *Yana* waʔic'ap'a

knee: *Achumawi* qohwāy; *Atsugewi* pulluj; *Cocopa* ʔimí-; *Diegueño* hemetun; *Hopi* tamö; *Karuk* páthak; *Kashaya* moq^ho; *Kawaiisu* tana-vi; *Lake Miwok* t^hok'óllu; *Lui-seño* qáx-may; *Maidu* pok'ósi; *Mono* tah nopoto; *Nisenan* podok*; *Northern Pomo* yasis; *Northern Sierra Miwok* hoŋ-oj-u; *Shasta* ʔíc'ipka; *Southeastern Pomo* k'da; *Tümpisa Shoshone* tangappüh; *Ute* táa-vy; *Wintu* p^huyeq; *Yana* puudik'i

know: *Achumawi* i-masʔāt; *Atsugewi* pé-matikas; *Cocopa* ʔu-yá-; *Diegueño* nur; *Hopi* navotiʔyta; *Karuk* áapunma; *Kashaya* ʔdu:ci**; *Kawaiisu* pucugu; *Lake Miwok* nénut; *Lui-seño* ʔayáli; *Maidu* jákkít; *Mono* suh-ta-pyha-~i; *Nisenan* ʔesak*; *Northern Pomo* šan; *Northern Sierra Miwok* hyjʔyksy; *Shasta* iweyka-ʔ**; *Southeastern Pomo* fdika; *Tümpisa Shoshone* sumpanai; *Ute* pucúcugway; *Wintu* t'ip; *Yana* diŋiba

laugh: *Achumawi* aliyum; *Atsugewi* russit; *Cocopa* ʔu-şáy; *Diegueño* uusay; *Hopi* tayati; *Karuk* ikshah; *Kashaya* c^huway***; *Kawaiisu* kiya; *Lake Miwok* jómu; *Lui-seño* tó-yu; *Maidu* núk; *Mono* jawi; *Nisenan* də-; *Northern Pomo* mic'e:liʔin*; *Northern Sierra Miwok* hyja-k-y; *Shasta* arikiʔ*; *Southeastern Pomo* keykeyli; *Tümpisa Shoshone* yahiʔ; *Ute* kiyǵ-ʔni; *Wintu*

nOy; Yana jal

leg: *Achumawi* saʔyā; *Atsugewi* wistí; *Cocopa* ʔimí; *Diegueño* emily; *Hopi* hokya; *Karuk* ápsiih; *Kashaya* šaku; *Kawaiisu* yuʔu-vi; *Lake Miwok* lóolo; *Luiſeño* ʔé-t; *Maidu* tòli; *Mono* huhka; *Nisenan* lu-l; *Northern Pomo* sino*; *Northern Sierra Miwok* kaw·ali; *Shasta* ʔarawéʔ; *Southeastern Pomo* q'a; *Tümpisa Shoshone* nungkwappüh; *Ute* yuʔúa-vi; *Wintu* lurur; *Yana* gaadu

lie: *Achumawi* ajpūmaj; *Atsugewi* witáq; *Cocopa* paʔ; *Diegueño* welyak; *Hopi* waʔök; *Karuk* ásish; *Kashaya* miʔi*; *Kawaiisu* havi; *Lake Miwok* cáte; *Luiſeño* hówa; *Maidu* t'ýj; *Mono* hapi; *Nisenan* wo; *Northern Pomo* miʔi; *Northern Sierra Miwok* jaŋ·ac·y; *Shasta* au*; *Southeastern Pomo* mʔi; *Tümpisa Shoshone* hapiʔ; *Ute* ʔavi; *Wintu* bEy; *Yana* mil

liver: *Achumawi* owē; *Atsugewi* úpsi; *Cocopa* cpušú; *Diegueño* chepesii; *Hopi* nuuma; *Karuk* váfish; *Kashaya* cahla; *Kawaiisu* niwi-bi; *Lake Miwok* kúlla; *Luiſeño* nó·ma; *Maidu* kýlla; *Mono* nywy*; *Nisenan* kylla; *Northern Pomo* šala; *Northern Sierra Miwok* kyl·a; *Shasta* ʔé·psiʔ; *Southeastern Pomo* šlal*; *Tümpisa Shoshone* nümü; *Ute* núu·pu; *Wintu* kila; *Yana* ima

long: *Achumawi* wātaxjūji; *Atsugewi* ijki; *Cocopa* kuʔ; *Diegueño* ʔequll; *Hopi* wuupa; *Karuk* vāaram; *Kashaya* ʔahqol; *Kawaiisu* paʔa-togo; *Lake Miwok* ʔadáak; *Luiſeño* tavú·lvu·š; *Maidu* lámpe; *Mono* ʔyty; *Nisenan* lamdi; *Northern Pomo* kol; *Northern Sierra Miwok* hylaw·a; *Shasta* ʔukk^waxawhiʔ*; *Southeastern Pomo* bcilin; *Tümpisa Shoshone* küpütappi; *Ute* paʔá-toŋwá-tj; *Wintu* kel; *Yana* ʒul

man: *Achumawi* yālyū; *Atsugewi* kəswi; *Cocopa* ʔapá; *Diegueño* ʔiikwich; *Hopi* taaqa; *Karuk* ávansa; *Kashaya* hiʔbaya; *Kawaiisu* taʔni-pizi; *Lake Miwok* tájh; *Luiſeño* yaʔá·š; *Maidu* jèpí; *Mono* nana; *Nisenan* jepʔ; *Northern Pomo* ba; *Northern Sierra Miwok* ŋaŋ·a; *Shasta* ʔawatikhwa·*; *Southeastern Pomo* cawi; *Tümpisa Shoshone* tangummü; *Ute* taʔwá·ci; *Wintu* wi; *Yana* hisi

many: *Achumawi* qām; *Atsugewi* jəkwúwo; *Cocopa* ʔʔay; *Diegueño* hemiiy; *Hopi* wuuhaq; *Karuk* táay; *Kashaya* baʔ^he; *Kawaiisu* ʔawa-vi; *Lake Miwok* cáne; *Luiſeño* hihé; *Maidu* pí; *Mono* ʔewa; *Nisenan* lokʔ; *Northern Pomo* baʔ^he; *Northern Sierra Miwok* waŋa; *Shasta* ʔukk^waríʔ; *Southeastern Pomo* bʔeqat; *Tümpisa Shoshone* soontün; *Ute* ʔaváʔna-tj; *Wintu* boh; *Yana* han

mountain: *Achumawi* aqo; *Atsugewi* ehew; *Cocopa* wi; *Diegueño* matetay; *Hopi* tuukwi; *Karuk* túuyship; *Kashaya* dono; *Kawaiisu* kee-vi; *Lake Miwok* páwih; *Luiſeño* qawí·ča; *Maidu* jamáni; *Mono* toja; *Nisenan* jaman; *Northern Pomo* dano; *Northern Sierra Miwok* lu·paj·y; *Shasta* wákk^wey; *Southeastern Pomo* kno; *Tümpisa Shoshone* toyapi; *Ute* káa-vi; *Wintu* buli; *Yana* ʒiigal

mouth: *Achumawi* āp; *Atsugewi* á·ppu; *Cocopa* ʔiyá; *Diegueño* aa; *Hopi* moʔa; *Karuk* apmaan; *Kashaya* ʔaha; *Kawaiisu* tibi-vi; *Lake Miwok* lúppe; *Luiſeño* tamá·t; *Maidu* simí; *Mono* tyhpe; *Nisenan* sim; *Northern Pomo* ha; *Northern Sierra Miwok* ʔaw·o; *Shasta* ʔaw; *Southeastern Pomo* xasto; *Tümpisa Shoshone* tümpe; *Ute* ʔupá-vi; *Wintu* qol; *Yana* bal

name: *Achumawi* tilīqa·tʔke; *Atsugewi* r'əksehí; *Cocopa* muʔ; *Diegueño* chehich; *Hopi* tungwni; *Karuk* ithvuy; *Kashaya* šihcima*; *Kawaiisu* niyaa-vi; *Lake Miwok* lákte; *Luiſeño* túŋ·la; *Maidu* já; *Mono* na-niya·hna; *Nisenan* ja; *Northern Pomo* ši; *Northern Sierra Miwok* ʔoja·se; *Shasta* kawcik; *Southeastern Pomo* xin; *Tümpisa Shoshone* niha; *Ute* níaa·n; *Wintu* yet; *Yana* ʔoozaiyau

neck: *Achumawi* ḥállōq; *Atsugewi* úpki; *Cocopa* mpuk; *Diegueño* eyepuk; *Hopi* kwaapi; *Karuk* vúup; *Kashaya* miha; *Kawaiisu* kura-vi; *Lake Miwok* heléeki; *Luiſeño* qelá-t; *Maidu* k'ùjì; *Mono* kuhta; *Nisenan* k'ujsoḱ'; *Northern Pomo* k^hu; *Northern Sierra Miwok* tok·olo·la; *Shasta* ʔík'uk; *Southeastern Pomo* q'oy; *Tümpisa Shoshone* kutan; *Ute* kurá-vi; *Wintu* dok; *Yana* dul

night: *Achumawi* maxēkʔjə; *Atsugewi* ape·na; *Cocopa* cn^yam; *Diegueño* hunn; *Hopi* took-ila; *Karuk* ikxáram; *Kashaya* duwe; *Kawaiisu* tugwa-nu; *Lake Miwok* káwul; *Luiſeño* tú·ku·mi·t; *Maidu* pó; *Mono* tokawano*; *Nisenan* po·; *Northern Pomo* diwe; *Northern Sierra Miwok* kawy·l·y; *Shasta* ʔapxá·; *Southeastern Pomo* ʔwe; *Tümpisa Shoshone* tukwanni; *Ute* tugwá-na; *Wintu* cipi; *Yana* basiiyau

nose: *Achumawi* yamē·; *Atsugewi* iwʔti; *Cocopa* ʔixú; *Diegueño* ehui; *Hopi* yaqa; *Karuk* yúfiv; *Kashaya* ʔila; *Kawaiisu* mu·vi·toʔo·bi; *Lake Miwok* húk; *Luiſeño* mú·vi·l; *Maidu* híky; *Mono* mu·pih; *Nisenan* ko·l; *Northern Pomo* la; *Northern Sierra Miwok* huk·u; *Shasta* ʔé·r; *Southeastern Pomo* la; *Tümpisa Shoshone* mupin; *Ute* moḡótʔq·pu; *Wintu* sO; *Yana* liilimau

one: *Achumawi* hamīs; *Atsugewi* jíw; *Cocopa* šit; *Diegueño* 'ehink; *Hopi* suukya; *Karuk* yítha; *Kashaya* hku; *Kawaiisu* suu·yu; *Lake Miwok* kénne; *Luiſeño* supúl; *Maidu* sýtti; *Mono* symy; *Nisenan* witte·; *Northern Pomo* č'a; *Northern Sierra Miwok* lut·i; *Shasta* c'á·ʔaʔ; *Southeastern Pomo* dan; *Tümpisa Shoshone* sümüttün; *Ute* súuyis; *Wintu* k'et; *Yana* bai

person: *Achumawi* is; *Atsugewi* awtʔé; *Cocopa* capáy; *Diegueño* 'iipay; *Hopi* sino; *Karuk* áraar; *Kashaya* ʔacaʔ; *Kawaiisu* niwi; *Lake Miwok* kóoca; *Luiſeño* ʔa·tá·x; *Maidu* májdy; *Mono* nyhmy; *Nisenan* majdy*; *Northern Pomo* čaʔ; *Northern Sierra Miwok* miw·y; *Shasta* ʔis; *Southeastern Pomo* mfo; *Tümpisa Shoshone* nüümü; *Ute* núu·ci; *Wintu* wi; *Yana* bin

play: *Achumawi* awaqōm; *Atsugewi* kwe·yúʔ; *Cocopa* lʔuy; *Diegueño* suunaarr; *Hopi* naataplawu; *Karuk* ixtíivha; *Kashaya* hquam***; *Kawaiisu* kiya; *Lake Miwok* ʔéela; *Luiſeño* múx; *Maidu* téte; *Mono* tybiiha*; *Nisenan* k'a·to; *Northern Pomo* t^ha·ʔan; *Northern Sierra Miwok* ʔaw·i; *Shasta* iruhu*; *Southeastern Pomo* c'ac'awlic'; *Tümpisa Shoshone* nui"; *Ute* kiyáy; *Wintu* laq; *Yana* yaa

pull: *Achumawi* ilū; *Atsugewi* lúllasw; *Cocopa* cxaʔ^y; *Diegueño* aalul; *Hopi* langakniʔyta; *Karuk* ishrúun; *Kashaya* šu***; *Kawaiisu* ca·nikwi; *Lake Miwok* híwa; *Luiſeño* hó·ti; *Maidu* wihúl; *Mono* cah; *Nisenan* wi·; *Northern Pomo* ši; *Northern Sierra Miwok* kup·a; *Shasta* a·xukaʔ; *Southeastern Pomo* muki*; *Tümpisa Shoshone* pisoto"; *Ute* piyó·ġwa·ʔay; *Wintu* qin; *Yana* heewa

rain: *Achumawi* aʔjīm̄j; *Atsugewi* kwayit; *Cocopa* pʔay; *Diegueño* 'ekwi; *Hopi* yokva; *Karuk* pathrih; *Kashaya* ʔihc^he; *Kawaiisu* ʔuwa·ri; *Lake Miwok* ʔúupa; *Luiſeño* xíla; *Maidu* k'adík; *Mono* tyyhkwa; *Nisenan* ba·j; *Northern Pomo* bišema; *Northern Sierra Miwok* nyk·a; *Shasta* ʔistuk'***; *Southeastern Pomo* sen; *Tümpisa Shoshone* ümatün; *Ute* ʔuwáy; *Wintu* luh; *Yana* barik'u

red: *Achumawi* taxta·qī; *Atsugewi* isríta·ʔwi; *Cocopa* x^waʔ; *Diegueño* 'ehwatt; *Hopi* paalangpu; *Karuk* áaxhich; *Kashaya* k'is; *Kawaiisu* ʔaga·ki; *Lake Miwok* ʔáwa; *Luiſeño* qwayá·qwyá·š; *Maidu* láklàkpe; *Mono* ʔahqa pono; *Nisenan* lak; *Northern Pomo* t'ac'; *Northern Sierra Miwok* weṭeṭ·i; *Shasta* ʔé·xtiʔ; *Southeastern Pomo* t'anṭ'anṭ·ankin; *Tümpisa Shoshone* angkapi; *Ute* ʔaká·ġa·rj; *Wintu* ted; *Yana* wadʔ

river: *Achumawi* ajūmā·; *Atsugewi* jumé·; *Cocopa* xawíʔ^y; *Diegueño* 'ehaa kwenuw; *Hopi* paayu; *Karuk* ishkēesh; *Kashaya* biʔda; *Kawaiisu* ʔawa·ti poʔo; *Lake Miwok* wúwwe; *Luiſeño*

waní-š; *Maidu* sèwí; *Mono* pa-huu-tyh; *Nisenan* sew; *Northern Pomo* bida:ʔ*; *Northern Sierra Miwok* waka-l-y; *Shasta* ʔatitáywiʔ; *Southeastern Pomo* bda; *Tümpisa Shoshone* paa okwetün; *Ute* nukwí-tj; *Wintu* mem; *Yana* daaxa

root: *Achumawi* waʔtu; *Atsugewi* waʔtu; *Cocopa* šma; *Diegueño* pehemaa; *Hopi* nga; *Karuk* épuum; *Kashaya* ʔihq^hoʔ; *Kawaiisu* tina; *Lake Miwok* šúuli; *Luiſeño* kwí·namus; *Maidu* pìwí; *Mono* tytyna; *Nisenan* papak*; *Northern Pomo* yem; *Northern Sierra Miwok* jo·meca; *Shasta* ʔará·ʔiʔ; *Southeastern Pomo* šuma*; *Tümpisa Shoshone* po'opi; *Ute* tynáa-vi; *Wintu* c'araw; *Yana* sidpayu

rope: *Achumawi* ā·sli; *Atsugewi* nu·ki; *Cocopa* yuʔár; *Diegueño* chekoot; *Hopi* wik-pangwa; *Karuk* áan; *Kashaya* sulemaʔ; *Kawaiisu* tuʔuru; *Lake Miwok* cikóote; *Luiſeño* wí·ča-t; *Maidu* k'ùk'ú; *Mono* tyqahpo; *Nisenan* k'uk'; *Northern Pomo* kaši; *Northern Sierra Miwok* ciko-te; *Shasta* ʔássa; *Southeastern Pomo* c'ikoʔ*; *Tümpisa Shoshone* tümuhun; *Ute* sávũ; *Wintu* c'e·k; *Yana* orogki

rotten: *Achumawi* wiʔtūʔpī; *Atsugewi* tú·pisi; *Cocopa* šax; *Diegueño* weshah; *Hopi* aavu; *Karuk* xâat; *Kashaya* ht^hoʔ*; *Kawaiisu* ʔataa-kwee-pi; *Lake Miwok* kúuh; *Luiſeño* ʔapá·kwaya; *Maidu* holókpe; *Mono* pihi; *Nisenan* holwok*; *Northern Pomo* ʔ^hoʔ; *Northern Sierra Miwok* jepe; *Shasta* kíkk^wat'iʔ; *Southeastern Pomo* šoʔit; *Tümpisa Shoshone* pisi"; *Ute* pikí; *Wintu* yulel; *Yana* yuʔla

salt: *Achumawi* tī·s; *Atsugewi* ʔí·wti; *Cocopa* šʔir; *Diegueño* 'esily; *Hopi* öonga; *Karuk* yúfish; *Kashaya* ʔaʔq'o; *Kawaiisu* ʔowa-vi; *Lake Miwok* kójjo; *Luiſeño* ʔéŋ-la; *Maidu* bá; *Mono* ʔohma; *Nisenan* ba; *Northern Pomo* šeʔe; *Northern Sierra Miwok* koj-o; *Shasta* ʔá·tax; *Southeastern Pomo* sya; *Tümpisa Shoshone* ongwapi; *Ute* ʔqǫ·vī; *Wintu* we·λ; *Yana* wiicu

sand: *Achumawi* tā·s; *Atsugewi* tʔis; *Cocopa* šʔa; *Diegueño* mes-haraay; *Hopi* tuuwa; *Karuk* yúxnaam; *Kashaya* miʔaʔ; *Kawaiisu* sihwa-bi; *Lake Miwok* šúkuj; *Luiſeño* ʔéxva-l; *Maidu* bymýky; *Mono* pa-hsiwah; *Nisenan* ʔa·hil; *Northern Pomo* mičaʔ; *Northern Sierra Miwok* wiskala; *Shasta* t'áccu; *Southeastern Pomo* hoʔo*; *Tümpisa Shoshone* pasingompin; *Ute* siwá·pũ; *Wintu* c'er; *Yana* kuiʔyau

say: *Achumawi* iss; *Atsugewi* iss; *Cocopa* ʔi; *Diegueño* wii; *Hopi* pangqawu; *Karuk* piip; *Kashaya* mic*; *Kawaiisu* mee; *Lake Miwok* háaju; *Luiſeño* yá; *Maidu* ʔa; *Mono* ʔine; *Nisenan* ha; *Northern Pomo* heʔ; *Northern Sierra Miwok* kac·y; *Shasta* imm*; *Southeastern Pomo* nu; *Tümpisa Shoshone* yúkwi"; *Ute* máy-kh; *Wintu* leweq; *Yana* tii

scratch: *Achumawi* ajexyāk; *Atsugewi* twojoq; *Cocopa* šx^wan^y; *Diegueño* hesak; *Hopi* haari; *Karuk* akxárap; *Kashaya* k'is***; *Kawaiisu* ca-weʔe; *Lake Miwok* šéeka; *Luiſeño* šó·ki; *Maidu* c'ýwak; *Mono* cah~woʔi; *Nisenan* c'y; *Northern Pomo* k'a:ʔin*; *Northern Sierra Miwok* ʔisy·j-y; *Shasta* a·cxiʔi·k; *Southeastern Pomo* ʔexaʔkit*; *Tümpisa Shoshone* kito'a; *Ute* cáa-ʔnəy; *Wintu* k'ar; *Yana* c'u

see: *Achumawi* īnimā·; *Atsugewi* pí·ma; *Cocopa* wi·; *Diegueño* ewuuw; *Hopi* postala; *Karuk* mah; *Kashaya* ca*; *Kawaiisu* pi-kee; *Lake Miwok* ʔúte; *Luiſeño* tów; *Maidu* c'e; *Mono* puh ni; *Nisenan* wə·; *Northern Pomo* čadin; *Northern Sierra Miwok* heʔe·j-y; *Shasta* eʔ*; *Southeastern Pomo* mko; *Tümpisa Shoshone* puni"; *Ute* pūnī·kya; *Wintu* win; *Yana* diwai

sew: *Achumawi* īla·m; *Atsugewi* raqp; *Cocopa* šuk^wíʔ^y; *Diegueño* peyuur; *Hopi* tuuʔiha; *Karuk* ikrup; *Kashaya* bi*; *Kawaiisu* ca-pugwiʔi; *Lake Miwok* šáme; *Luiſeño* ʔuláʔ·qi; *Maidu* pyjé; *Mono* atsuna*; *Nisenan* puje*; *Northern Pomo* bit'elen; *Northern Sierra Miwok*

pakly; *Shasta* utwik; *Southeastern Pomo* bšuti; *Tümpisa Shoshone* tattsokwiih; *Ute* caġáy; *Wintu* hur; *Yana* wakui

short: *Achumawi* jokjəʔaxjūji; *Atsugewi* iwintka; *Cocopa* xlʷʔuʔ; *Diegueño* meputtk; *Hopi* hiisava; *Karuk* ipshûunkinach; *Kashaya* biʔt'e; *Kawaiisu* toveʔe-pii-či; *Lake Miwok* c'otóoj; *Luiseno* kapá-kpa-ma-l; *Maidu* t'és; *Mono* tapotsitsiʔini-tu*; *Nisenan* t'un; *Northern Pomo* bikuʔ*; *Northern Sierra Miwok* cap-ut-i; *Shasta* ʔihíkku; *Southeastern Pomo* q'otki*; *Tümpisa Shoshone* toppotts; *Ute* ka-paʔá-toġó-wa-tj; *Wintu* worot; *Yana* bagangu

sing: *Achumawi* ē-s; *Atsugewi* ie-j; *Cocopa* sya; *Diegueño* 'echweyuuw; *Hopi* tawta; *Karuk* pákurih; *Kashaya* cahno*; *Kawaiisu* kaa; *Lake Miwok* kója; *Luiseno* hé-la; *Maidu* sól; *Mono* hupija-tu; *Nisenan* sol; *Northern Pomo* bedi; *Northern Sierra Miwok* myl-i; *Shasta* ecn*; *Southeastern Pomo* xenol; *Tümpisa Shoshone* hupianaih; *Ute* káay; *Wintu* c'a-w; *Yana* hee

sit: *Achumawi* uskīm; *Atsugewi* juʔkup; *Cocopa* wa; *Diegueño* wenak; *Hopi* qatu; *Karuk* kûur; *Kashaya* ca*; *Kawaiisu* kari; *Lake Miwok* hówo; *Luiseno* táwa; *Maidu* wólkinu; *Mono* qahty; *Nisenan* bo-; *Northern Pomo* čima; *Northern Sierra Miwok* huŋ-e; *Shasta* iss*; *Southeastern Pomo* sa; *Tümpisa Shoshone* katü"; *Ute* karí; *Wintu* la; *Yana* wa

skin: *Achumawi* toqo-lūji; *Atsugewi* íwi; *Cocopa* nʷkʷʔaʔ; *Diegueño* ekwally; *Hopi* puukya; *Karuk* máan; *Kashaya* s'iʔda; *Kawaiisu* ʔasiʔa; *Lake Miwok* ʃúluk; *Luiseno* tá.va-š; *Maidu* posála; *Mono* sihpóʔo; *Nisenan* po; *Northern Pomo* siyan; *Northern Sierra Miwok* pyc-eʔa; *Shasta* ʔicka; *Southeastern Pomo* k'otal; *Tümpisa Shoshone* nümüa; *Ute* síʔaa-vi; *Wintu* la-s; *Yana* miwi

sleep: *Achumawi* ūʔma-t; *Atsugewi* wij; *Cocopa* šma; *Diegueño* hema; *Hopi* puuwi; *Karuk* ikvūt-ha; *Kashaya* sima; *Kawaiisu* ʔipii; *Lake Miwok* ʔéec; *Luiseno* kúp; *Maidu* túj; *Mono* ʔywi; *Nisenan* tuj; *Northern Pomo* sima; *Northern Sierra Miwok* jaŋe-m-y; *Shasta* icmae*; *Southeastern Pomo* knak'a; *Tümpisa Shoshone* üppüih; *Ute* puí; *Wintu* xi-n; *Yana* sam

small: *Achumawi* jō-kjə; *Atsugewi* iyiwíŋka; *Cocopa* ʔcaš; *Diegueño* 'estik; *Hopi* hiisay; *Karuk* nīnamich; *Kashaya* qawi; *Kawaiisu* ʔaakanuu-či; *Lake Miwok* cíceh; *Luiseno* kihá-t; *Maidu* tibi; *Mono* tsitsi'i-tu*ʔino; *Nisenan* la-j; *Northern Pomo* bičuʔ; *Northern Sierra Miwok* ʔic-ipit-i; *Shasta* ʔat'ukʷ**"; *Southeastern Pomo* k'učin; *Tümpisa Shoshone* tütiittsittsi; *Ute* míi-py-cj; *Wintu* ku-t'e; *Yana* t'inii

smell: *Achumawi* īmaslīmā; *Atsugewi* wripsa; *Cocopa* xʷi; *Diegueño* wehwii; *Hopi* navota; *Karuk* imshákar; *Kashaya* mihše*; *Kawaiisu* ponoho-ri; *Lake Miwok* húkʃu; *Luiseno* húʃi; *Maidu* hí; *Mono* ʔekwi; *Nisenan* t'yn; *Northern Pomo* miše; *Northern Sierra Miwok* huka; *Shasta* ummaik; *Southeastern Pomo* mxecit*; *Tümpisa Shoshone* kwana"; *Ute* ʔugwí; *Wintu* sub; *Yana* hul

spit: *Achumawi* aʔeʔtus; *Atsugewi* staw; *Cocopa* sxʔu; *Diegueño* chehutt; *Hopi* tōha; *Karuk* yuh; *Kashaya* pʰuʔ***; *Kawaiisu* čičiiʔi; *Lake Miwok* tukéetj; *Luiseno* čúxi; *Maidu* t'úp; *Mono* tuhi; *Nisenan* do-; *Northern Pomo* kaliʔ'agan; *Northern Sierra Miwok* tukja; *Shasta* epxuti-k*; *Southeastern Pomo* c'daʔkiqat*; *Tümpisa Shoshone* tusi"; *Ute* kycíʔnaŋ; *Wintu* l'uqa; *Yana* poʔla

stand: *Achumawi* aʔjā-wa; *Atsugewi* aqá-s; *Cocopa* pʔa-; *Diegueño* pehekwii; *Hopi* wunu; *Karuk* ihyárih; *Kashaya* ʔ'eʔ*; *Kawaiisu* wini; *Lake Miwok* tálah; *Luiseno* wí-ta; *Maidu* týskinu; *Mono* wynyh; *Nisenan* dəkə; *Northern Pomo* ʔoman; *Northern Sierra Miwok*

hac-icyksy; *Shasta* ak*; *Southeastern Pomo* ktoma; *Tümpisa Shoshone* wünü”; *Ute* wuní; *Wintu* hEk; *Yana* waak’i

star: *Achumawi* jā-mīq; *Atsugewi* náywiǰənó-qa; *Cocopa* k^wɬ^yap; *Diegueño* kwenmesaap; *Hopi* soohu; *Karuk* atáyraam; *Kashaya* q^hamos’; *Kawaiisu* puucii-vi; *Lake Miwok* tóʔle; *Luiſeño* šúʔ-la; *Maidu* lylý; *Mono* ta-cinuh; *Nisenan* mo-lo; *Northern Pomo* ʔ^hoʔ^hol; *Northern Sierra Miwok* hos-ok-ona; *Shasta* xakk^wé-sur; *Southeastern Pomo* sexayk’uyk’uy; *Tümpisa Shoshone* tatsiumpi; *Ute* púucii-vi; *Wintu* λuyu-q; *Yana* iʔlala

stone: *Achumawi* ālisti; *Atsugewi* nú-yehe; *Cocopa* xucdotr; *Diegueño* ’ewily; *Hopi* owa; *Karuk* as; *Kashaya* q^haʔbe; *Kawaiisu* ti-bi; *Lake Miwok* hóka; *Luiſeño* tó-ta; *Maidu* ʔó; *Mono* tyh-pih; *Nisenan* ʔo-; *Northern Pomo* xabe; *Northern Sierra Miwok* sawa; *Shasta* ʔiccaʔ; *Southeastern Pomo* χbe; *Tümpisa Shoshone* tümpin; *Ute* ʔupúy-ci; *Wintu* soh; *Yana* k’ai

suck: *Achumawi* ājij; *Atsugewi* wiʔjik; *Cocopa* sucúk; *Diegueño* wepis; *Hopi* tsootsona; *Karuk* vir; *Kashaya* ʔt’o**; *Kawaiisu* čohmi; *Lake Miwok* c’úm; *Luiſeño* hípi; *Maidu* dúť; *Mono* pihci; *Nisenan* mic’i; *Northern Pomo* ʔ’on; *Northern Sierra Miwok* cujmu; *Shasta* impa*; *Southeastern Pomo* ʔ’okit*; *Tümpisa Shoshone* ; *Ute* picí; *Wintu* bOl; *Yana* pu

sun: *Achumawi* jōl; *Atsugewi* jineʔwó-; *Cocopa* n’a; *Diegueño* ’enyaa; *Hopi* taawa; *Karuk* kúusrah; *Kashaya* haʔda; *Kawaiisu* ta-vi; *Lake Miwok* híi; *Luiſeño* timé-t; *Maidu* ʔékdám pòk’ó; *Mono* ta-pe; *Nisenan* ʔok; *Northern Pomo* mit’a; *Northern Sierra Miwok* hiʔe-ma; *Shasta* c’úwar; *Southeastern Pomo* da; *Tümpisa Shoshone* tape; *Ute* tavá-ci; *Wintu* sas; *Yana* t’ui

swim: *Achumawi* eh̥pā-; *Atsugewi* apsjaw; *Cocopa* xaɬ^y nup; *Diegueño* aapull; *Hopi* mo-mortima; *Karuk* ikpuh; *Kashaya* ʔahq^haciw**; *Kawaiisu* paa-ge-nukwi; *Lake Miwok* tímmuh; *Luiſeño* wá-ya; *Maidu* pyjéto; *Mono* pa-hapi; *Nisenan* pi; *Northern Pomo* k’oʔaman; *Northern Sierra Miwok* ʔypsy; *Shasta* ip*; *Southeastern Pomo* χaykam; *Tümpisa Shoshone* nokoit-soih; *Ute* ʔaví-vqrí; *Wintu* dum; *Yana* puu

that: *Achumawi* qē-; *Atsugewi* kuʔja; *Cocopa* pu-; *Diegueño* puu; *Hopi* pam; *Karuk* pay; *Kashaya* haʔu*; *Kawaiisu* ʔu-na; *Lake Miwok* máa; *Luiſeño* ʔó-nu; *Maidu* ʔàní; *Mono* ʔu; *Nisenan* my-; *Northern Pomo* mul; *Northern Sierra Miwok* no-; *Shasta* ʔinnáʔ; *Southeastern Pomo* ʔui; *Tümpisa Shoshone* sa; *Ute* ʔú; *Wintu* peyeh; *Yana* dairi

thick: *Achumawi* waniʔyaymī-; *Atsugewi* upéʔwo; *Cocopa* šmuck; *Diegueño* menuttk; *Hopi* pööngala; *Karuk* ítpuum; *Kashaya* q^hoʔt’o; *Kawaiisu* mucu-gwi; *Lake Miwok* t’údduju; *Luiſeño* tupú-tpu-š; *Maidu* kówwil; *Mono* woqo; *Nisenan* di-l; *Northern Pomo* lay; *Northern Sierra Miwok* me; *Shasta* ʔitikissi**; *Southeastern Pomo* kmudikit*; *Tümpisa Shoshone* tuhuntappütün; *Ute* tumúuta-rj; *Wintu* cutup; *Yana* bac’ilt’ai

think: *Achumawi* ipsyū; *Atsugewi* ʔmí-p; *Cocopa* ɬ^yyu-m; *Diegueño* iichaa; *Hopi* wuuwa; *Karuk* xus; *Kashaya* ʔt’a**; *Kawaiisu* muguwa-ri; *Lake Miwok* háli; *Luiſeño* wóyʔa; *Maidu* húwěje; *Mono* suh-mija; *Nisenan* pis; *Northern Pomo* dine; *Northern Sierra Miwok* ʔon-i; *Shasta* iyacac*; *Southeastern Pomo* nant’ac’i; *Tümpisa Shoshone* mukuatu; *Ute* máy-kh-ni; *Wintu* λ’am; *Yana* gima

three: *Achumawi* jāsti; *Atsugewi* qiski; *Cocopa* xmuk; *Diegueño* hemuk; *Hopi* paayom; *Karuk* kuyraak; *Kashaya* sibo; *Kawaiisu* pehe-yu; *Lake Miwok* deléeka; *Luiſeño* pá-hay; *Maidu* sáp’y; *Mono* pahi; *Nisenan* sapwij; *Northern Pomo* subu; *Northern Sierra Miwok* tolo-kos-u; *Shasta* xácki; *Southeastern Pomo* hoqat; *Tümpisa Shoshone* pahi; *Ute* páy-ni;

Wintu panuḷ; *Yana* bul

throw: *Achumawi* āpp; *Atsugewi* yappaw; *Cocopa* cap; *Diegueño* wetom; *Hopi* tuuva; *Karuk* path; *Kashaya* ko**; *Kawaiisu* tavi; *Lake Miwok* lét'a; *Luiſeño* péŋa; *Maidu* wydým; *Mono* wyna-ʔi; *Nisenan* wi-; *Northern Pomo* čatkan*; *Northern Sierra Miwok* ʔakwa; *Shasta* appi-ma; *Southeastern Pomo* bdeka; *Tümpisa Shoshone* kuna"; *Ute* tʔrávi; *Wintu* λEy; *Yana* oowa

tie: *Achumawi* īnē-mī-; *Atsugewi* runar; *Cocopa* xir; *Diegueño* tuunnak; *Hopi* soma; *Karuk* inhi; *Kashaya* ʔbe**; *Kawaiisu* kagi; *Lake Miwok* cáni; *Luiſeño* tú-ča; *Maidu* jodót; *Mono* wyh-tawa; *Nisenan* du-; *Northern Pomo* p^habe; *Northern Sierra Miwok* hyspa; *Shasta* irut*; *Southeastern Pomo* naked; *Tümpisa Shoshone* tsokwah; *Ute* tapícʔay; *Wintu* cOc; *Yana* hedʔ

tongue: *Achumawi* iplē-; *Atsugewi* aphli; *Cocopa* mpaʔ; *Diegueño* enepall; *Hopi* lengi; *Karuk* ápriih; *Kashaya* haba; *Kawaiisu* ʔegu-bi; *Lake Miwok* léttip; *Luiſeño* wé-yi; *Maidu* ʔèní; *Mono* ʔeqo; *Nisenan* ʔal; *Northern Pomo* haba; *Northern Sierra Miwok* nep-it-y; *Shasta* ʔéhena; *Southeastern Pomo* bal; *Tümpisa Shoshone* okon; *Ute* ʔagó-pi; *Wintu* tahal; *Yanabawa* ʔla

tooth: *Achumawi* ijā-; *Atsugewi* iʔjaw; *Cocopa* ʔiyá; *Diegueño* eyaaw; *Hopi* tama; *Karuk* vuh; *Kashaya* hoʔo; *Kawaiisu* tawa-bi; *Lake Miwok* kút; *Luiſeño* tamá-t; *Maidu* c'ík'í; *Mono* tawa; *Nisenan* c'awa*; *Northern Pomo* ʔo; *Northern Sierra Miwok* kyt-y; *Shasta* ʔíc'aw; *Southeastern Pomo* ʔo; *Tümpisa Shoshone* taman; *Ute* tawá-pu; *Wintu* si-; *Yana* kic'au

two: *Achumawi* haʔaq; *Atsugewi* hoqi; *Cocopa* xwak; *Diegueño* hewak; *Hopi* lööyö; *Karuk* áxak; *Kashaya* q^ho; *Kawaiisu* waha-yu; *Lake Miwok* ʔóʔta; *Luiſeño* wéh; *Maidu* péne; *Mono* waha; *Nisenan* pe-n; *Northern Pomo* xo; *Northern Sierra Miwok* ʔoti-ko; *Shasta* xúkk^waʔ; *Southeastern Pomo* ʔos; *Tümpisa Shoshone* waha; *Ute* wáy-ni; *Wintu* pa-l; *Yana* ux

walk: *Achumawi* oqe; *Atsugewi* iʔw; *Cocopa* pʔaw; *Diegueño* wamp; *Hopi* waynuma; *Karuk* áhoo; *Kashaya* wa*; *Kawaiisu* miya; *Lake Miwok* wícaj; *Luiſeño* wukála; *Maidu* ʔynó; *Mono* nywi; *Nisenan* he; *Northern Pomo* waden; *Northern Sierra Miwok* wy-ny; *Shasta* att*; *Southeastern Pomo* wali; *Tümpisa Shoshone* tattükinnaih; *Ute* paǵáy-ʔway; *Wintu* q'ay; *Yana* nix

wash: *Achumawi* ajopsaj; *Atsugewi* jwiʔtór; *Cocopa* ʂʔul; *Diegueño* allehwas; *Hopi* tuu-vahoma; *Karuk* pithxah; *Kashaya* dahsex**; *Kawaiisu* pa-zagi; *Lake Miwok* cólla; *Luiſeño* páci; *Maidu* c'uk'út; *Mono* pa-hcaqa; *Nisenan* juc'u-; *Northern Pomo* dase; *Northern Sierra Miwok* hek-a; *Shasta* iccuk^wik; *Southeastern Pomo* xasoka; *Tümpisa Shoshone* koitsoih; *Ute* caq̣xóyi; *Wintu* yOq; *Yana* k'al-di

water: *Achumawi* as; *Atsugewi* áji; *Cocopa* xa; *Diegueño* 'ehaa; *Hopi* paahu; *Karuk* áas; *Kashaya* ʔahq^ha; *Kawaiisu* poʔo; *Lake Miwok* kíik; *Luiſeño* pá-la; *Maidu* mò mí; *Mono* pa; *Nisenan* mom; *Northern Pomo* ʔa; *Northern Sierra Miwok* kik-y; *Shasta* ʔacca; *Southeastern Pomo* ʔa; *Tümpisa Shoshone* paa; *Ute* páa; *Wintu* mem; *Yana* xa

white: *Achumawi* tī-wī-jī-; *Atsugewi* pehkuri; *Cocopa* xma-ʔ; *Diegueño* nemeshap; *Hopi* qöötsa; *Karuk* taahkúnish; *Kashaya* qahle; *Kawaiisu* see-gi; *Lake Miwok* ceʔáaw; *Luiſeño* xwáya; *Maidu* dáldāl; *Mono* tohci nu; *Nisenan* kow; *Northern Pomo* kale; *Northern Sierra Miwok* kelel-i; *Shasta* ʔít'a-yuʔ; *Southeastern Pomo* t'ot'okin; *Tümpisa Shoshone* tosapi; *Ute* sá-ǵa-rj; *Wintu* xa-y; *Yana* dalabʔsa

wind: *Achumawi* ʔo-m; *Atsugewi* wissu-í; *Cocopa* cxa; *Diegueño* yaayp; *Hopi* huukyangw;

Karuk ikréemyah; *Kashaya* ?ihya; *Kawaiisu* nee-di; *Lake Miwok* héna; *Luiseno* húŋ-la; *Maidu* býwo; *Mono* hyhkwah; *Nisenan* by·je; *Northern Pomo* ya; *Northern Sierra Miwok* he·nisa; *Shasta* ʔá·ska; *Southeastern Pomo* ya; *Tümpisa Shoshone* nüetün; *Ute* núa-rj; *Wintu* k'ah; *Yana* ʒuk'al

woman: *Achumawi* amitē·wjən; *Atsugewi* minəríja; *Cocopa* sʔak; *Diegueño* siny; *Hopi* wuuti; *Karuk* asiktávaan; *Kashaya* ʔimaʔa; *Kawaiisu* momoʔo; *Lake Miwok* pócci; *Luiseno* ʂuǵá-l; *Maidu* kylé; *Mono* hyyhpi'; *Nisenan* kyle; *Northern Pomo* maʔa; *Northern Sierra Miwok* ʔos·a; *Shasta* taríc'iʔ; *Southeastern Pomo* bted; *Tümpisa Shoshone* wa'ippü; *Ute* mamá-ci; *Wintu* p'Oq; *Yana* p'udʔ

Achumawi data from Olmsted (1966)

Atsugewi data from Olmsted (1984)

Cocopa data from Crawford (1989)

Diegueño data from Couro and Hutcheson (1973)

Hopi data from Hill (1998)

Karuk data from Bright and Gehr (2008), except * from Bright (1957)

Kashaya data from Oswalt (1957), except * from Oswalt.002.039, ** from Oswalt.001.041, ***from Kashaya dictionary draft, all in Oswalt (n.d., 1967)

Kawaiisu data from Zigmund et al. (1990)

Lake Miwok data from Callaghan (1965)

Luiseno data from Bright (1968)

Maidu data from Shipley (1963)

Mono data from Lamb (1958), except * from Bethel et al. (1993)

Nisenan data from Uldall and Shipley (1966), except * from Sawyer.097 in Sawyer (n.d.)

Northern Pomo data from Oswalt.001.055 and Oswalt.001.056 in Oswalt (n.d., 1967), except * from Vihman.002.003 in Vihman (1970)

Northern Sierra Miwok data from Callaghan (1987)

Shasta data from Olmsted and Bright (1959), except * from Angulo.007 in de Angulo and Freeland (n.d.) and Silver.002 in Silver (1961)

Southeastern Pomo data from Haynie and Kelsey (2007), except * from Moshinsky (1965)

Tümpisa Shoshone data from Daley (1989)

Ute data from The Southern Ute Tribe (1979)

Wintu data from Pitkin (1985)

Yana data from Sapir and Swadesh (1960)

Transcriptions in this list are based upon original sources. Word lists for the test were transcribed using a standardized ASCII system, discarding non-phonemic information.