A Treebank of the Karuk Language

by

Erik H. Maier

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 Line Mikkelsen, Chair
Professor Peter Jenks
Professor Beth Piatote

Fall 2022

A Treebank of the Karuk Language

Abstract

A Treebank of the Karuk Language

by

Erik H. Maier

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Line Mikkelsen, Chair

In this dissertation, I introduce the Karuk treebank, a collection of syntactically-parsed sentences of the Karuk language. The goals of this dissertation are, first, to describe the construction of the treebank and the rationale for its design and, second, to showcase the utility of the treebank through case studies in two domains: the order of arguments and predicates, and cases of unexpected agreement marking. The study of word order showcases the treebank's aptitude at helping us understand large-scale statistical patterns in Karuk syntax, and the latter study of agreement showcases the use of the treebank in finding rare and previously unstudied phenomena.

Chapter 1: Introduction provides the necessary background on the Karuk language and the history of its documentation by outsiders and on the treebank project itself. **Chapter 2: Annotation Guidelines** presents the annotation guidelines which were both used as a manual to guide annotators in their annotation of the language and now serve as an explanation and description of the use of every element found in Karuk treebank annotations.

Chapter 3: Argument and Predicate Order presents the first case studies utilizing the treebank, focusing on the word order of arguments and predicates. Karuk word order had often been claimed to be 'free' with every or most orders of subject, direct object, and verb attested, but their relative prevalence and the word orders of clauses with other types of argument (complements and indirect objects) or non-verbal predicates is eludicated for the first time in detail in this chapter. Methodologically, I argue in Chapter 3 for the utility of treebanks in allowing easier study of large-scale, quantitative properties of corpora than comparable, treebank-less methods. In the case studies meant to showcase this, I describe broad, word order patterns of the treebank corpus and elucidate three trends found in this data: that subordinate clauses tend to have less expressed arguments than main clauses; that subjects are less likely to be expressed in transitive clauses; and that the prevalence of pre-verbal S is driven partially by a likelier-than-expected tendency for the presence of both a subject and object to lead to pre-verbal subjects.

Chapter 4: Agreement presents the second set of case studies, focusing on two phenomena where observed agreement in the corpus does not match the agreement expected from Bright (1957)'s description of the agreement system: sentences with plural agreement where singular was expected, and vice versa. These two mismatches (and some inconsistency with one particular agreement prefix for 3pl subjects and 3pl objects) turn out to the only systematic cases that differ from Bright's description of the agreement, thus confirming his accuracy overall. Methodologically, in this chapter I argue for the utility of treebanks in locating rare phenomena in corpora that may escape notice by comparable treebank-less methods. In terms of the case studies, I elucidate in Chapter 4 a heretofore undescribed phenomenon whereby the use of plural subject agreement with a singular subject indicates subject demotion. **Chapter 5** concludes with some thoughts about the future of the treebank project.

Contents

Contents **i**

List of Figures **iv**

List of Tables **v**

Acknowledgements **vii**

1 Introduction **1**
1.1 The Karuk language 1
1.1.1 Previous research on Karuk 2
1.1.2 The corpus of the Karuk treebank 5
1.2 Design principles of the Karuk treebank 10
1.2.1 Comparing the Karuk treebank to others 12
1.3 History of the Karuk Treebank 16
1.4 Structure of the dissertation 16

2 Annotation Guidelines **18**
2.1 Types of information encoded in the treebank 18
2.2 General principles for resolving parse multiplicity 20
2.2.1 Attach High 20
2.2.2 Follow the Translation 20
2.3 Dependency graphs 21
2.4 Definitions 22
2.5 Glossing conventions 22
2.6 Syntactic statuses 23
2.6.1 VOC 23
2.6.2 PRED 24
2.6.3 FRAG 25
2.6.4 SPRED 25
2.6.5 INT 26
2.7 Syntactic relations 26

2.7.1 SBJ . . . 26
2.7.2 OBJ . . . 29
2.7.3 IO . . . 31
2.7.4 DET . . . 33
2.7.5 QUANT . . . 35
2.7.6 POSS . . . 37
2.7.7 ATR . . . 41
2.7.8 TAM . . . 44
2.7.9 NEG . . . 48
2.7.10 COMP . . . 49
2.7.11 QUOT . . . 54
2.7.12 ADV . . . 55
2.7.13 SUB . . . 56
2.7.14 APPOS . . . 58
2.7.15 COORD . . . 59
2.7.16 COP . . . 60
2.8 Person/number annotation . . . 61
2.8.1 Person/number annotation tags . . . 62
2.8.2 Annotating person/number . . . 64
2.9 Annotation of specific constructions . . . 64
2.9.1 Identifying whether discourse particles are dependents of subordinate or main clauses . . . 64
2.9.2 Multiple clauses with no coordinator or a non-*káru* coordinator . . . 65
2.9.3 Noun phrases without a lexical noun . . . 66
2.9.4 Ellipsis and other missing verbs . . . 67
2.9.5 Analyzing the third-person pronoun *uum* . . . 69
2.9.6 *vaa* used in apposition to a clause . . . 70
2.9.7 *xakaan*, *koovan*, *îin*, and *puraan* . . . 72
2.9.8 Causatives . . . 72
2.9.9 Discourse particles separated from main clause by subordinate clause 73
2.9.10 Intensifier particle *vúra* . . . 75
2.9.11 Subcategorization frames for particular verbs . . . 76
2.10 Conclusion . . . 80

3 Argument and Predicate Order **81**
3.1 General search methodology . . . 82
3.2 Previous work . . . 82
3.3 Basic properties of clauses . . . 85
3.3.1 Transitivity . . . 86
3.3.2 Subordination . . . 88
3.3.3 Verbal and non-verbal predicates . . . 90
3.3.4 Interactions . . . 92

3.4 Clauses with two or more expressed arguments . . . 94
3.4.1 Clauses with subject and direct object . . . 95
3.4.2 Clauses with subjects and complements . . . 98
3.4.3 Clauses with direct objects and complements . . . 100
3.4.4 Clauses with indirect objects and one other argument . . . 100
3.4.5 Three arguments present . . . 102
3.5 Clauses with a single expressed argument . . . 106
3.6 Interactions . . . 108
3.6.1 Subordinate Argument Suppression Tendency . . . 108
3.6.2 Transitive Subject Suppression Tendency . . . 114
3.6.3 Object-Conditioned Pre-Verbal Subject Tendency . . . 119
3.6.4 Further interactions . . . 123
3.7 Conclusion . . . 124

4 Agreement **125**
4.1 The Karuk agreement system . . . 126
4.2 Methodology . . . 130
4.3 Unexpected plural agreement . . . 131
4.3.1 Previous work . . . 132
4.3.2 Unexpected plural agreement indicates subject demotion . . . 134
4.3.3 The *-ahi* passive . . . 153
4.4 Unexpected singular agreement . . . 157
4.4.1 Explicitly plural inanimates . . . 158
4.4.2 Animates . . . 169

5 Conclusion **172**

Bibliography **174**

A Dataset: Unexpected plural agreement **179**

B Dataset: Unexpected singular agreement **184**

List of Figures

3.1 Number of clauses by transitivity 86
3.2 Number of main vs. subordinate clauses 89
3.3 Number of verbal vs. nonverbal clauses 91
3.4 Clauses by transitivity and subordination 93
3.5 Clauses by transitivity and verbal versus non-verbal status 94
3.6 Clauses with S and O 96
3.7 Clauses with S and C 99
3.8 Clauses with O and C 102
3.9 Clauses with S or O and I 103
3.10 Clauses with a single expressed argument 107

List of Tables

1.1 Summary of Karuk language documentation by outsiders 4
1.2 List of texts in the Karuk treebank 8

2.1 Tags for person/number annotation 62

3.1 Number of clauses by transitivity 86
3.2 Number of main vs subordinate clauses 88
3.3 Main and subordinate clause frequences in German, Dutch, and English (Kempen and Harbusch 2019) 89
3.4 Number of verbal vs. nonverbal clauses 91
3.5 Main/subordinate x Transitivity 92
3.6 Main/subordinate x Verbal/Non-verbal 93
3.7 Orders of clauses with S and O expressed 95
3.8 WALS Feature 81A: Order of Subject, Object, and Verb 97
3.9 Orders of clauses with S and C expressed 99
3.10 O, C, and V orders for clauses with 2+ arguments 101
3.11 O, I, and V orders for clauses with 2+ arguments 101
3.12 Sentences with three expressed arguments 103
3.13 Relative prevalence of clauses with 1 argument 106
3.14 Prevalences of clauses by number of expressed arguments 107
3.15 Interaction between argument expression and subordination 109
3.16 Interaction between presence of a single argument and subordination 110
3.17 Interaction between presence of subject as sole argument and subordination . . 110
3.18 Interaction between presence of object as sole argument and subordination . . . 111
3.19 Interaction between presence of complement as sole argument and subordination 111
3.20 Interaction between presence of S+O and subordination 112
3.21 Subordination and clause order of subject, verb, object 112
3.22 Interaction between presence of S+C and subordination 113
3.23 Subordination and clause order of subject, verb, complement 113
3.24 Interaction between presence of S+C and subordination 113
3.25 Subordination and clause order of subject, verb, complement 114
3.26 Interaction of transitivity and whether S is expressed 115
3.27 Interaction of transitivity and whether C is expressed 116

3.28 Interaction between presence of S+C and transitivity 116
3.29 Transitivity and clause order of subject, verb, complement 117
3.30 Interaction between presence of a single argument and transitivity 117
3.31 Interaction between presence of C as sole argument and transitivity 118
3.32 Interaction of pre- or post-predicate position of S and whether an object is expressed 120
3.33 Interaction of pre- or post-predicate position of O and whether S is expressed . . 121
3.34 Interaction of pre- or post-predicate position of S and whether a complement is expressed 122
3.35 Interaction of pre- or post-predicate position of O and whether a complement is expressed 122
3.36 Interaction of transitivity and S position 123
3.37 Interaction of expression of S as only argument and verbal or non-verbal predicate status 124

4.1 De Angulo and Freeland (1931)'s description of Karuk agreement 127
4.2 Karuk verbal agreement 128

Acknowledgments

This is a daunting section to write. I've been associated with UC Berkeley a long time for a graduate student, beginning in 2010 after transferring to Berkeley from community college for undergrad and going up until the completion of this dissertation: twelve years. I've worked with the Karuk language for nearly as long, beginning with a class taught by Line Mikkelsen and Andrew Garrett in Spring 2012 and then my first trip to Yreka, California to work with Karuk elders in January 2013. I've been shaped by my colleagues, peers, friends, teachers, and mentors from Berkeley, Karuk country, and elsewhere deeply over these years, and I doubt that I can accurately represent all of the gratitude I bear for them on these scant few pages. Nonetheless, I hope to record here at least some of those whose support was instrumental in bringing this dissertation into being.

I have had the great privilege and honor of working with many Karuk elders and master speakers since 2013, who taught me about the Karuk language and culture, and also about strength, about kindness, and living a good life. Though this dissertation is based on the language of an earlier generation of Karuk speakers and not on my own recordings with elders, it is only due to the teaching I received from these elders that this dissertation could possibily have been written: the late Vina Smith, the late Charlie Thom Sr., the late Charron (Sonny) Davis, the late Lucille Albers, the late Alvis (Bud) Johnson, and the late Norman Goodwin. I also owe much gratitude to the Karuk speakers, teachers, and learners who graciously accepted me, worked alongside me, spoke Karuk with me, and taught me over these years: Florrine Super, Arch Super, Crystal Richardson, Nancy Steele, Julian Lang, Tamara Alexander, LuLu Alexander, and Susan Gehr. Yôotva, nanífyiivshas. Kíri xára xákaan araráhih nuchúuphitih.

My committee members Line Mikkelsen, Peter Jenks, and Beth Piatote were incredibly flexible and made dissertation writing a smooth experience, especially given the fact that the bulk of the writing was during the COVID-19 pandemic with all its disruptions. Their comments were instrumental in making this dissertation more than just a collection of disordered observations. I owe a larger debt of gratitude to Line Mikkelsen in particular than can be fully expressed: her kind but firm advising through graduate school shepherded me through difficult times and made me into a better scholar and person, and it was largely due to her that I even became interested in the Karuk language and pursuing a linguistics PhD in the first place. In addition, Andrew Garrett provided helpful commentary on this dissertation and has been a thought-provoking interlocutor regarding the Karuk language since before I started graduate school. Andrew likewise played a large role in sparking my interest in the Karuk language and pursuing a PhD. Susanne Gahl and Amy Rose Deal both provided helpful advice on how to present quantitative results and on impersonal constructions, respectively. Belén Flores and Paula Floro made sure that all administrative issues in the department were handled smoothly, and, like many linguistic graduate students at Berkeley during their tenure, I don't think I could have finished if not for them.

I am thankful also for the many fellow/former graduate students at Berkeley who made for a supportive, fun, and intellectually stimulating community, including my cohort Emily

Clem, Virginia Dawson, Alice Shen, Andrew Cheng, Amalia Skilton, and Geoff Bacon, and many others, including some I have surely forgotten to name: Kenny Baclawski, Nico Baier, Madeline Bossi, Meg Cychosz, Emily Drummond, Matt Faytak, Matt Goss, Dmetri Hayes, Noah Hermalin, Edwin Ko, Raksit Lau-Preechathammarach, Tyler Lemon, Myriam Lapierre, Jack Merrill, Julia Nee, Kelsey Neely, Zachary O'Hagan, Emily Remirez, Nicholas Rolle, Hannah Sande, Clare Sandy, Martha Schwarz, Tessa Scott, Tammy Stark, Elise Stickles, and Eric Wilbanks. Some deserve a further thanks: Katie Sardinha, for all the nights at the Oakland Metro watching postmodern wrestling or Mongolian metal and all the deep discussions we've had and are yet to have; and Schuyler Laparle, for all the absurd humor, sass, board game and art nights, New Parkway movies, and illuminating conversations past and future.

I benefitted greatly from being able to work with brilliant undergraduate apprentices who helped with the annotation of the Karuk treebank: Ciara Agrella, Andrew Baker, Jessica Butcher, Sammy Keyes-Levine, Stuart Litjen, Zeynep Özselçuk, LaLa Speights-Barhatkov, Jessica Wang, Cindy Yang, and Kevin Yu. Though I cannot hope to name them all, I am also thankful to all the students I've taught in my years at Berkeley; their eagerness, excitement and inquisitiveness would often serve to rekindle my own.

Without the support and friendship of Zeynep Özselçuk, I don't think I would've finished this dissertation. Escaping to her house to play video games, draw beautiful and absurd collaborative art, and hang out with her cats or wandering through campus, browsing through the Berkeley Art Museum, Half Price, or Moe's Books – such moments were a needed counterpoint to the stress.

I owe much more than thanks to my parents, Steve and Dalene Maier, who have always supported my chosen path. And I am most grateful for the time I got to spend with my cat, Andromeda, whenever I would visit my hometown. I'll miss her forever.

Chapter 1

Introduction

In this dissertation, I introduce the Karuk treebank, a collection of syntactically-parsed sentences of the Karuk language. The goals of this dissertation are as follows: First, to describe the construction of the treebank and the rationale for choices made during its construction; and second, to showcase the utility of the treebank through case studies in two domains: the order of arguments and predicates, and cases of unexpected agreement marking. The study of word order will showcase the treebank's aptitude at helping us understand large-scale statistical patterns in Karuk syntax, and the latter study of agreement will showcase the use of the treebank in finding rare phenomena.

This introduction serves to provide a foundation for the other material in this dissertation, and as such includes the following sections meant to provide the necessary background. First, I introduce the Karuk language, the history of linguistic research on the language, and describe the corpus that the Karuk treebank consists of. I then move on to a broad overview of the principles and design of the Karuk treebank, with particular focus on how the Karuk treebank compares to other treebanks. To end, I introduce the succeeding chapters of this dissertation, and summarize the results from the case studies in Chapters 3 and 4.

1.1 The Karuk language

The Karuk language is an indigenous language of California, traditionally spoken around the middle course of the Klamath river in northern California, from downriver of Panámniik (Orleans) to upriver of Athithúfvuunupma (Happy Camp), and is used now by members of the Karuk community who still live there and across California and the world. Garrett, Gehr, et al. (in press) provides an overview of the language including grammatical, socio-cultural, and historical information. Readers are encouraged to seek there for more detailed information that this current discussion will omit.

The Karuk language is an isolate, though it has been proposed that it is part of the Hokan grouping and thus related to a large number of languages scattered throughout California, including the neighboring Shasta and Chimariko languages (Dixon and Kroeber 1913; Haas

1964; Silver 1974; Silver 1976), though there is warranted skepticism about the validity of this grouping (Poser 1995; Campbell 1997)[1] and, if the grouping is valid, it must be a very distant relationship. Though speaking unrelated languages, Karuk people have much cultural contact and affinity with neighboring Yurok and Hupa-speaking peoples; so much so that Kroeber declared the three to have an identical culture (Kroeber 1925), though further research has unsurprisingly discovered cultural variation among these groups that does not necessarily correspond to linguistic groupings (Conathan 2006).

Cook (1956) estimated the Karuk population at the time of invasion to be around 2700 individuals; by the time William Bright was documenting the language, he reports there were only around 100 individuals still speaking Karuk. In the current day, the last generation of people who learned Karuk as a first language (generally born in the first half of the 20th century) have mostly passed away, though there are several fluent second-language speakers and many people teaching and learning the language both in and out of schools. Despite the cataclysm of the invasion in the late 1840s, there has not been a time where Karuk ceased being spoken, even though intergenerational transmission of the language from parent to child has been disrupted and the language of every-day life for Karuk people has shifted to English.

In terms of linguistic features, the Karuk language has a relatively simple consonant and vowel inventory, exhibits what has been called a pitch accent system with prosodically constrained high and falling tones (see Sandy (2017) for a more thorough examination), exhibits light polysynthesis (i.e. directional suffixes with nominal meanings are common, but there is no actual incorporation of nouns into verbal stems) and typically is agglutinative though with some fusional morphology. The language is highly suffixing and head-marking. In terms of syntax, the language exhibits Hale (1983)'s three criteria for a 'non-configurational' language: free word order, free pro-drop, and free argument-splitting. More detail on the morphology and syntax will be provided in Chapter 3, regarding word order, and Chapter 4, which will discuss in detail the verbal agreement system of Karuk.

1.1.1 Previous research on Karuk

Documentation of the Karuk language by Karuk speakers working with outside researchers began shortly after the invasion of Karuk land by settlers from the United States in 1849, with the first vocabulary collected by George Gibbs in 1851-2 (Gibbs 1853; Golla 2011), and continued intermittently into the late 2010s. Taken together, Golla (2011)'s section on Karuk, Sandy (2017)'s section on Karuk documentation, and Garrett, Gehr, et al. (in press) provide a comprehensive summary of the history of documentation of the language, which

[1] As Poser states, "the reason that the Hokan family has been and remains controversial is that for many of the links that make up the overall family, evidence of genetic affliation of the sort considered probative by most historical linguists has been unacceptably thin. It is the tenuousness of the relationship of the putative Hokan languages and the lack of properly-worked out derivations from Proto-Hokan that have made it impossible to detect the shared innovations that form the basis for classical (nonlexicostatistical) subgrouping." (Poser 1995, p. 142)

the table below summarizes with the addition of identifying the Karuk speakers involved in each project. In some cases, names of the speakers were not recorded or not recorded fully, and for most cases I have sourced the names of speakers from the outside researcher's publications. In these cases, there may be speakers who are reflected only in field notes but not publications, and who are thus left out. As no comprehensive list of Karuk speakers who contributed to the documentary record exists, this will serve as a first step toward this goal.

Karuk Speaker(s)	Researcher(s)	Dates	Publications
Unknown	George Gibbs	1851-1852	(Gibbs 1853; Powers 1877)
Unknown	Stephen Powers	1872	(Powers 1877)
Unknown	Jeremiah Curtin	1889	(Curtin 1889a) (Curtin 1889b) (Curtin 1889c)
Little Ike Mary Ike Little Ike's Mother Oak-Bottom-Flat Jack Three-Dollar-Bar Billy Dick Richard's Father-in-Law Sweet William of Ishipishi Julia Bennett John Gorham Martha Horne	Alfred Kroeber	1901-1903	(Kroeber 1911) (Kroeber and Gifford 1980)
Unknown	C. Hart Merriam	1910-1921	
Mary Ike Georgia Orcutt Shan Davis Mamie Offield	E.W. Gifford		(Kroeber and Gifford 1980)
Benonie Harrie Margaret Harrie Frank Ruben Nettie Ruben	Jaime DeAngulo and L.S. Freeland		(de Angulo and Freeland 1931)
Phoebe Maddux Fritz Hansen Yaas Abner's mother Jim Pepper	J. P. Harrington	1925-1929	(Harrington 1930) (Harrington 1932b) (Harrington 1932a)

Dora Davis Johnny Pepper	Ursula McConnel	1932-1933	
Caroline Bisocin Benonie Harrie Margaret Harrie Phoebe Maddux	Hans Jørgen Uldall		
Nettie Ruben Julia Starritt Mamie Offield Chester Pepper Lottie Beck Daisy Jones Maggie Charley Emily Donahue Violet Super Grace Davis Madeline Davis	William Bright	1949-2000s	(Bright 1957) inter alia (see below)
Violet Super	Monica Macaulay	1980s	See below
Vina Smith Charron (Sonny) Davis Charlie Thom Sr. Lucille Albers Alvis Johnson Bud Smith Norman Goodwin Julian Lang Crystal Richardson	Andrew Garrett, Line Mikkelsen, Clare Sandy, Erik Hans Maier, Karie Moorman, Kayla Begay, among others	2010-present	See below

Table 1.1: Summary of Karuk language documentation by outsiders

Based on the documentation above, linguists have produced a relatively large amount of analytic work on the language. In addition to the grammar of Karuk that served as his dissertation (Bright 1957), William Bright wrote several papers about literary aspects of Karuk texts (Bright 1977; Bright 1979a; Bright 1980a; Bright 1980b; Bright 1984), effects of contact with English (Bright 1952), an ethnographic sketch (Bright 1978), onomastics (Bright 1958), maledicta (Bright 1979b), semantic similarities between Karuk, Yurok, and Hupa, with Jane Bright (Bright and Bright 1965), and a dictionary with Karuk archivist and linguist Susan Gehr (Bright and Gehr 2004), which is the source of the online dictionary

Ararahih'urípih.[2]

Monica Macaulay has written extensively about a wide variety of aspects of Karuk grammar, including the suffixal status of morphemes Bright analyzed as clitics (Macaulay 1989), pitch accent (Macaulay 1990), inverse marking and verbal agreement (Macaulay 1992), reduplication (Macaulay 1993), obviative marking (Macaulay 2000), directional suffixes (Macaulay 2004) and, with Claudia Brugman, evidentiality (Brugman and Macaulay 2015) and the function of the discourse particle *káruma* (Brugman and Macaulay 2009).

Recent efforts growing out of the Berkeley Karuk research project have produced research into a variety of topics, including: reduplication (Conathan and Wood 2003), linguistic contact effects with Yurok and Hupa (Conathan 2004; Conathan 2006), prosody (Sandy 2017), pragmatic influences on word order (Yu 2021), canonicality of agreement (White 2015), the polyfunctionality of the instrumental suffix *-ara* (Moorman 2015), directional suffixes (Maier 2016a), prospective aspect (Carpenter 2014), and resumptive pronouns (Davis et al. 2020). The Berkeley Karuk research project also led to the creation of *Ararahih'urípih*, 'Karuk language net,' the online Karuk dictionary and corpus, and, of course, the Karuk treebank that relies on *Ararahih'urípih*.

The complex verbal agreement system, which will be discussed in more depth in Chapter 4, has also attracted work in theoretical syntax in addition to the work by Macaulay cited above (Campbell 2012; Béjar 2003; Béjar and Rezac 2009; Kumaran 2018).

1.1.2 The corpus of the Karuk treebank

The sentences comprising the treebank are taken from *Ararahih'urípih*, 'Karuk language net' the online dictionary and text corpus of the Karuk language, which currently includes around 7000 morphologically-annotated sentences, recorded at various points by various researchers and Karuk speakers in the time period from 1903 to 2016. The Karuk treebank consists of a subcorpus of the online text corpus, focused on one major source of *Ararahih'uripih*'s texts: the texts published in Bright (1957), the landmark grammar, lexicon, and text collection of Karuk. These are texts recorded in the late 1940s period of William Bright's documentation. Table 1.2 below lists the texts in this corpus by their identifier in the online corpus, their title, the speaker of each text, the length in sentences[3], and the genre. In total, there are 93 texts comprising, according to Bright's count of sentences, 3776 sentences[4].

[2] *Ararahih'urípih* can be accessed online at `https://linguistics.berkeley.edu/~karuk/`

[3] The length given here is based off of sentence breaks that Bright provided in his transcription, and does not conform to the number of sentences in the treebank. Each Bright sentence can potentially be many independent syntactic sentences in the treebank annotation; Bright utilized a prosodic heuristic for determining sentence breaks, while the treebank defines a sentence by grouping together all dependents of an independent, main predicate.

[4] This comes to 5303 sentences according to how the treebank defines sentences, as will be clarified in Chapter 3.

Text ID	Title	Speaker	Length	Genre
WB_KL-0	Swearing	Julia Starritt	14	Ethnological
WB_KL-01	Coyote's Journey	Nettie Ruben	137	Pikvah
WB_KL-02	Coyote's Homecoming	Nettie Ruben	73	Pikvah
WB_KL-02a	Coyote's Homecoming	Nettie Ruben	60	Pikvah
WB_KL-03	Coyote's Journey	Chester Pepper	177	Pikvah
WB_KL-04	Coyote's Journey	Julia Starritt	178	Pikvah
WB_KL-05	Coyote's Journey	Mamie Offield	145	Pikvah
WB_KL-06	Coyote Goes to a War Dance	Julia Starritt	76	Pikvah
WB_KL-07	Coyote Trades Songs	Nettie Ruben	61	Pikvah
WB_KL-08	Coyote Goes to the Sky	Julia Starritt	62	Pikvah
WB_KL-09	Coyote Trades Songs and Goes to the Sky	Mamie Offield	58	Pikvah
WB_KL-10	Coyote Steals Fire	Mamie Offield	63	Pikvah
WB_KL-11	Coyote as Doctor	Nettie Ruben	51	Pikvah
WB_KL-12	Coyote Tries to Reach the Sun	Chester Pepper	7	Pikvah
WB_KL-13	Coyote's Gambling Song	Chester Pepper	20	Pikvah
WB_KL-14	Coyote Eats His Own Excrement	Nettie Ruben	34	Pikvah
WB_KL-15	Coyote as Lawmaker	Nettie Ruben	17	Pikvah
WB_KL-16	Coyote Marries r His Own Daughter	Julia Starritt	50	Pikvah
WB_KL-17	Coyote Gives Salmon and Acorns to Mankind	Mamie Offield	60	Pikvah
WB_KL-18	The Perils of Weasel	Lottie Beck	74	Pikvah
WB_KL-19	The Perils of Weasel	Mamie Offield	30	Pikvah
WB_KL-20	The Perils of Weasel	Daisy Jones	70	Pikvah
WB_KL-21	The Hair in the Soup	Julia Starritt	42	Pikvah
WB_KL-22	Old Man Turtle Dances	Lottie Beck	28	Pikvah
WB_KL-23	The Greedy Father	Lottie Beck	72	Pikvah
WB_KL-24	The Greedy Father	Mamie Offield	47	Pikvah
WB_KL-25	Duck Hawk and His Wife	Lottie Beck	26	Pikvah
WB_KL-26	Duck Hawk and His Wife	Chester Pepper	18	Pikvah
WB_KL-27	Duck Hawk and His Wife	Mamie Offield	32	Pikvah
WB_KL-28	Blue Jay as Doctor	Mamie Offield	22	Pikvah
WB_KL-29	Blue Jay as Doctor	Nettie Ruben	24	Pikvah
WB_KL-30	The Story of Tan Oak Acorn	Lottie Beck	25	Pikvah
WB_KL-31	The Story of Tan Oak Acorn	Nettie Ruben	12	Pikvah
WB_KL-32	The Bear and the Deer	Julia Starritt	99	Pikvah
WB_KL-33	How Deer Meat was Lost and Regained	Mamie Offield	90	Pikvah

WB_KL-34	Lizard and Grizzly Bear	Nettie Ruben	73	Pikvah
WB_KL-35	The Story of Madrone	Lottie Beck	53	Pikvah
WB_KL-36	Why Towhee Has Red Eyes	Nettie Ruben	8	Pikvah
WB_KL-37	Eel and Sucker	Mamie Offield	7	Pikvah
WB_KL-38	The Story of Slug	Mamie Offield	18	Pikvah
WB_KL-39	The Adulterers Discovered	Nettie Ruben	25	Pikvah
WB_KL-40	The Story of Bear	Nettie Ruben	22	Pikvah
WB_KL-41	The Creation of Eels	Nettie Ruben	27	Pikvah
WB_KL-42	The Story of Crane	Nettie Ruben	11	Pikvah
WB_KL-43	Mocking Bird and Swamp Robin	Nettie Ruben	16	Pikvah
WB_KL-44	Why Lightning Strikes Trees	Nettie Ruben	16	Pikvah
WB_KL-45	The Victory Over Fire	Mamie Offield	23	Pikvah
WB_KL-46	The Story of Skunk	Nettie Ruben	25	Pikvah
WB_KL-47	The Origin of the Pikiawish	Chester Pepper	18	Pikvah
WB_KL-48	The Origin of the Pikiawish	Nettie Ruben	24	Pikvah
WB_KL-49	Medicine for the Return of a Sweetheart	Nettie Ruben	35	Ánav
WB_KL-50	Medicine to Get a Husband	Nettie Ruben	25	Ánav
WB_KL-51	Love Medicine	Chester Pepper	57	Ánav
WB_KL-52	Medicine for the Return of Wives	Chester Pepper	86	Ánav
WB_KL-53	Deer-hunting Medicine	Chester Pepper	28	Ánav
WB_KL-54	Shinny Game Medicine	Mamie Offield	40	Ánav
WB_KL-55	Wrestling Medicine	Mamie Offield	27	Ánav
WB_KL-56	The Flood	Mamie Offield	11	Uhyanathêepar
WB_KL-57	The Boy from Itúkuk	Nettie Ruben	152	Uhyanathêepar
WB_KL-58	A Trip to the Land of the Dead	Mamie Offield	57	Uhyanathêepar
WB_KL-59	The Pool in Big Rock	Nettie Ruben	40	Uhyanathêepar
WB_KL-60	The Snake People	Daisy Jones	32	Uhyanathêepar
WB_KL-61	The Kidnapped Child	Lottie Beck	48	Uhyanathêepar
WB_KL-62	The Devil Discovered	Mamie Offield	25	Uhyanathêepar
WB_KL-63	The Devil Who Died Laughing	Mamie Offield	16	Uhyanathêepar
WB_KL-64	The Devil and the Girl	Mamie Offield	39	Uhyanathêepar
WB_KL-65	The White Man's Gifts	Nettie Ruben	23	Historical
WB_KL-66	How the Rube Family was Named	Julia Starritt	22	Historical
WB_KL-67	A Quack Doctor	Julia Starritt	12	Historical
WB_KL-68	Indian Food	Maggie Charley	26	Ethnological

WB_KL-69	Salmon Fishing	Julia Starritt	24	Ethnological
WB_KL-70	Deer Hunting	Nettie Ruben	22	Ethnological
WB_KL-71	Bear Hunting	Nettie Ruben	26	Ethnological
WB_KL-72	Elk Hunting	Nettie Ruben	12	Ethnological
WB_KL-73	Making Acorn Soup	Julia Starritt	32	Ethnological
WB_KL-74	Making Acorn Soup	Nettie Ruben	23	Ethnological
WB_KL-75	Soaking Acorns	Julia Starritt	13	Ethnological
WB_KL-76	The Sweathouse	Julia Starritt	16	Ethnological
WB_KL-77	The Living-house	Julia Starritt	20	Ethnological
WB_KL-78	The Shinny Game	Julia Starritt	29	Ethnological
WB_KL-79	Work Contests	Nettie Ruben	10	Ethnological
WB_KL-80	The Sucking Doctor	Julia Starritt	17	Ethnological
WB_KL-81	The Sweating Doctor	Julia Starritt	12	Ethnological
WB_KL-82	The Pikiawish at Katimin	Emily Donahue	24	Ethnological
WB_KL-83	The Pikiawish at Katimin	Nettie Ruben	50	Ethnological
WB_KL-84	The Pikiawish at Clear Creek	Maggie Charley	37	Ethnological
WB_KL-85	Preparing Basket Materials	Emily Donahue	38	Ethnological
WB_KL-86	Indian Clothes	Julia Starritt	13	Ethnological
WB_KL-87	Tattoos	Julia Starritt	8	Ethnological
WB_KL-88	Professor Gifford's Visit	Emily Donahue	17	Anecdote
WB_KL-89	A Birthday Party	Julia Starritt	6	Anecdote
WB_KL-90	Smoke	Julia Starritt	3	Anecdote
WB_KL-91	A Blow-out	Julia Starritt	5	Anecdote
WB_KL-92	Responses to Pictures	Julia Starritt	98	Elicitation

Table 1.2: List of texts in the Karuk treebank

A word is in order about the genre labels provided here. The assignment of texts to particular genres follows the grouping set out in (Bright 1957), but I have opted to rename these groups for ease of reference, since Bright's labels are lengthy as shall be seen, and to use Karuk terms where possible. Bright labels the text groupings based on the era that the events of the story take place in. The largest genre group in terms of texts are pikvahs, with 48 of the 93 texts. Bright labels these "The Era of the Ikxaréeyav[5]" and subdivides the texts into subgroups based on their protagonists, i.e. Coyote, Weasel, etc. The Karuk word for stories of this type is *pikvah*, both as a noun referring to a story of this type and a verb meaning 'to tell stories.' In the introduction to his book of translated Karuk stories *Ararapíkva*, Karuk speaker, activist and artist Julian Lang describes pikvahs thusly:

[5]The Karuk word *ikxaréeyav* is translated as 'spirit person' or 'god' and refers to the race of beings that in Karuk/Northern Californian mythology inhabited the world before the arrival of humans. These *ikxaréeyav* and their actions are understood to provide the blueprint for the lives and behavior of animals and people today.

> Creation stories (**pikva**) chronicle our creation; the origin of the Karuk people and the Karuk World is found in the **pikva**. Some of the stories are told in the exact words of the **Ikxaréeyav** Spirit People. [...] The Ikxaréeyavs already knew how Human's future was going to be. They left their creation stories behind so we would be able to learn from their wisdom, to learn from their trials. (Lang 1994, pp. 27-28)

Another set of texts is a subtype of pikvah: Ánav, which is the Karuk word for 'medicine' and refers both to this type of story and to physical substances used to cure or alleviate sickness. Bright labeled these texts "Medicine stories." Ánav are pikvahs recited as part of a ritual, often to help the reciter accomplish a task that the story's ikxaréeyav protagonist accomplished in the story. These stories often end with a formulaic exhortation to this ikxaréeyav, as in (1) taken from Chester Pepper's *Love Medicine*:

(1) víri naa kích îin tá nu-'apunmîik. víri naa káru vaa
so 1SG.PRO only OBV PER 1SG>2SG-learn.from so 1SG.PRO also thus
ni-kuph-eesh.
1SG-do-PROSP
'I alone have learned it from you. I will do that way also." (Chester Pepper, "Love Medicine," WB_KL-51:56-56)

The texts which are labelled here Uhyanathêepar are stories about legendary human beings, Karuk people who existed in the time before the invasion of settlers and before the direct memories of living people. Bright labelled these as from "The Era of the Indian," but I follow Lang (1994) in the use of the term uhyanathêepar, which he uses to refer to this era. Lang defines it as such: "They are stories of the **Yaas'ára**, the first Human beings, our ancestors. We call the human-story era Uhyanathêepar" (p. 28)[6]

The texts which I have labelled 'historical' are likewise stories about human beings, but told about the time period after the United States invasion. Bright identified these as stories about "The Era of the White Man." These stories typically lack the supernatural elements found in the genres described above, and in the three stories of this type in Bright's corpus, they detail aspects of Karuk people's interactions with white settlers and among themselves in the years following the invasion.

The texts labelled 'ethnological' are all descriptions of traditional (pre-invasion) Karuk living and culture; Bright labelled these "Ethnological Descriptions" and divided them into these categories: Food, buildings, recreation, shamans, world renewal, and miscellaneous. These texts typically speak in generic terms, as opposed to telling a story about specific individuals as in the other genres here.

The texts I have labelled 'anecdotes' are what Bright called "personal accounts" and they are just that: stories directly from the lives of the Karuk people who told the stories.

[6]Note that the term uhyanathêepar is a term referring to the human era, not necessarily to refer to stories of this era in the way that pikvah refers directly to the stories themselves. The etymology of uhyanathêepar is unclear to me, though the segment uhyana could be a prepound form of the word *uuhyan* meaning 'word.'

Typically, they consist of short, slice-of-life episodes such as the preparations for a birthday party happening the day of the story's telling (WB_KL-89: A Birthday Party, told by Julia Starritt).

The final genre here I have labelled 'elicitation.' Bright grouped the single text with this designation in with his 'personal accounts,' but I have split it from them because the nature of this text (WB_KL-92: Responses to pictures) is quite different from the others in that grouping. In text 92, Julia Starritt describes a set of drawings which Bright shows her. As such, the text is composed of many short sequences of sentences describing pictures, typically focused on the spatial orientation of figures in the drawing. A characteristic sentence from this text is given below:

(2) yeeripáxvu u-hyárih, ú-ksuupku-ti ípaha, káan u-'íihya
girl 3SG-stand 3SG-point.at-DUR tree, there 3SG-stand
A girl is standing, she is pointing at a tree, it is standing there. (Julia Starritt, WB_KL-92:32)

There is evidence of other emic Karuk genres of storytelling that are not found within Bright's corpus. In his introduction to *Karok Myths and Formulas*, E.W. Gifford writes that the "threefold segregation of stories in [his] paper is on the basis of native classification: myths (pikuava) [pikvah], formulas (anava) [ánav], and confessions (bigishtu'u) [pikishkoo[7]]." (Kroeber and Gifford 1980, p.107). Pikvah and ánav have been earlier described, but no stories in Bright's corpus have been specifically identified as pikishkoo or confessions. Gifford writes:

> The confessions recount the doings of earlier members of living families and are believed to have supernatural potency, like the medicine formulas. In other words, they are recounted to achieve results today. Thus, they really constitute recent additions to the stock of medicine formulas, but deal with known human beings and not with the immortals. They are used especially to cure illness in infants and children. Wrongdoing of adults causes illness in infants; confession has potentialities for curing. Even the confessed misdeeds of ancestors are potent for cures today. (p.107)

The ánav stories found in Bright's corpus all feature ikxaréeyav protagonists, and such none of them could fit into the genre of pikishkoo.

1.2 Design principles of the Karuk treebank

The foremost goal of the Karuk Treebank project was to create a research tool which can enable syntactic research on the Karuk language that is based on accurate knowledge of the

[7]The Karuk word *pikishkoo* is a verb meaning 'to absolve of a crime.' Gifford apparently mistook the *k* in this word for a *t*.

corpus as a whole. Ideally, the treebank will allow researchers to come to generalizations which reflect the true nature of the corpus and, assuming the corpus has adequate coverage, the language. More specifically, the nature of this research tool is that, based on the systematic syntactic annotation it is comprised of, it allows researchers to isolate sets of sentences that share some syntactic property or properties. It should be able to isolate all and only the sentences which are relevant to whatever syntactic question the researcher has in mind. If a researcher is studying the variable positions of the subject, they can use the treebank to find those sentences where the subject is postverbal, for example, and if they are studying the construction of *hôoy íf* rhetorical questions being used to imply negation, they will be able to isolate only those sentences which include that construction. Most importantly, the researcher is able to isolate the relevant set of sentences without needing to comb through the entire corpus searching for examples themselves. In creating the treebank, that combing has already been completed, and the nature of the annotations in the treebank ensures that a wide variety of syntactic properties can be investigated in the future without the time-consuming step of "manual" searching through the corpus.

Given this goal, the treebank does not necessarily encode in its annotations syntactic structures which are substantively motivated empirically. Rather, the syntactic annotations are meant to produce the type of subgrouping described in the previous paragraph. There are several constructions which remain syntactically mysterious but are given naive annotations for the purpose of being able to isolate all and only the examples of those constructions. Of course, in many cases the treebank's annotations will reflect well-motivated ideas about the true syntactic structure (at least as much as can be in the particular types of information that the treebank encodes). It is generally in the analysis of particular constructions where isomorphism between the actual syntactic structure and treebank annotation is not maintained.

An important feature of the treebank is that it is based on the already robust lexical and morphological annotation present in *Ararahih'urípih*. This annotation already allows one to find all instances of a particular word or morpheme in the corpus. This enables the treebank to be streamlined in not having to encode morphological or lexical information alongside syntactic information.[8] The syntactic annotation adds a new, abstract layer to the annotation, such that one can search not just for morphemes but also syntactic structures. The subgrouping of sentences which the treebank seeks to enable is a result of shared labor between syntactic and morphological information, with the eventual goal being the ability to find subgroups of sentences based on simultaneous lexical, morphological, and syntactic criteria.

As of the writing of this dissertation, the treebank is not publically accessible. All searching of the treebank's annotations are done using Python scripts that search through the xml files which contain the treebank's data. This, of course, is a strong barrier to other

[8]This streamlining is not trivial. A major hurdle in the production of treebanks like the Penn English Treebank is the development of tools to do part-of-speech tagging of a corpus (Marcus, Santorini, and Marcinkiewicz 1993), which is already present in the Karuk corpus.

researchers within and without the Karuk community being able to use the treebank for its intended purpose. As such, future efforts must be directed at integrating the treebank into the online Karuk dictionary and corpus, which already has wide usage in the community. The complexity of the annotations is another barrier to usage, as the large variety of syntactic relations and other annotation categories (as will be described in Chapter 2) will require effort to learn fully. However, there are likely other ways to integrate information gleaned from the treebank into the online dictionary that will not require as much time investment: having a list of all the attested objects that a verb takes automatically added to that verb's dictionary entry, for example. That sort of information is encoded in the treebank and a script could quickly enumerate all the objects of every verb. Then, users of the online dictionary would have some more information to be able to see the difference between words than otherwise have very similar English translations: for example, *ifik* and *imkaanva* are both routinely translated as 'to gather' in texts, but differ in what sorts of things are licit objects for each one, with *ifik* being used primarily for acorns (and other, non-food items that one picks up from the ground) and *imkaanva* being used for food more generally (but not for acorns). The exact shape of how the treebank will be integrated into the website remains to be worked out in the future, with consultation from the Karuk community about how best to make the treebank useful to their language reclamation efforts.

The particular framework employed in the Karuk Treebank is a type of dependency grammar, where syntactic structure is represented as binary relations between a head and a dependent.

1.2.1 Comparing the Karuk treebank to others

There are a few major axes upon which treebanks differ. Two are of particular relevance to situating the Karuk treebank in the realm of treebanks. One axis is related to the purpose which the treebank is designed for: is the treebank designed for use by linguists to utilize corpus data to better understand language or investigate linguistic questions, or is the corpus for natural language processing researchers to utilize for evaluating the performance of parsers? Treebanks need not take a hard stand on this question, and can be useful for both (and other purposes besides). Both purposes are in fact mentioned in Marcus, Santorini, and Marcinkiewicz (1993)'s description of the design of perhaps the most well-known treebank, the Penn English Treebank: "Such corpora [treebanks] are beginning to serve as important research tools for investigators in natural language processing, speech recognition, and integrated spoken language systems, as well as in theoretical linguistics." (p. 1). However, aspects of the design of any given treebank may reveal a focus toward one or the other purpose. Two quotes make this distinction clear. Burga, Öktem, and Wanner (2017), introducing a set of revisions they made to the METU-Sabancı Turkish treebank, state:

> Dependency treebanks are crucial for the development of statistical NLP applications, including sentence parsing and generation. To obtain good performance,

> well-defined and coherent treebank annotation schemes are needed. To provide an outcome that is good not only in quantitative but also in qualitative terms in the sense that it is well-suited for various down-stream applications, the annotation scheme must be **equally rigorous from the linguistic viewpoint**. (1) (Emphasis added.)

Compare this to a quote from Taylor (2020), describing the design of treebanks used for investigating diachronic syntax:

> Although annotation schemes are necessarily informed by linguistic theory, the annotation is intended to be practical rather than to make theoretical claims about the structure of the language. Its purpose is only to facilitate the extraction of the data, leaving it to the researcher to do the analysis. (7)

From the NLP perspective represented in the Burga, Öktem, and Wanner (2017) quote, the linguistic accuracy of the treebank is important, whereas the theoretical linguistic approach represented by the Taylor (2020) quote is clearly less concerned with accuracy, and more with producing a tool helpful for linguistic analysis.[9] As discussed in the previous section above, the primary goal of the Karuk treebank is akin to the goal described by Taylor (2020): to provide a tool for better understanding of the Karuk language. Because the study of Karuk syntax is not as well-developed as the study of the syntax of languages like English or Turkish, it is difficult to know if annotations can be accurate without first having the tools to better study the language.

A further axis of variation is the particular choice of framework, with the two most common being dependency grammar and constituency grammar. The Penn English Treebank, already mentioned, is based on constituency grammar, with representations that encode intermediate nodes that do not correspond to any particular word in the sentence (e.g. phrasal nodes such as NP). Dependency grammar treebanks are common,[10] and generally involve syntactic relations only between words on the surface, with no intermediate nodes that don't correspond to an actual word in the sentence. The Karuk treebank utilizes a dependency grammar framework, for similar reasons as discussed for the focus on being a tool for linguistic research: dependency grammar, by focusing on relations between words, forces

[9] The similarities with historical treebanks and the Karuk treebank are deeper than just a focus on assisting linguistic research. Historical treebanks, as corpora of historical languages no longer spoken, are annotated by people who are not native speakers of the language being annotated. Likewise, the Karuk treebank is annotated by people who are not native speakers of Karuk. The intended goal for the Karuk treebank is also to have syntactic annotation for texts ranging over the entire period of Karuk language documentation, which would make it also a diachronic treebank.

[10] A few examples of these: there is the Prague Dependency Treebank of Czech (Hajic and Hladka 2001), the Ancient Greek and Latin dependency treebank (Bamman and Crane 2011), and the over 100 treebanks using the Universal Dependency framework `https://universaldependencies.org/`

no assumptions of intermediate structure, for which one may not be sure of the empirical validity.[11]

Another important fact about the Karuk treebank relevant for comparison to other treebanks is the nature of the Karuk language itself. The Karuk language is polysynthetic and has relatively free word order,[12] and is highly endangered, with a relatively small corpus. Languages with a similar profile are rarely found to have treebanks, though the Universal Dependencies project includes treebanks of several similar languages, and as such comparisons with the Universal Dependencies framework will be discussed in more depth.

Comparison with Universal Dependencies

An important recent development in treebanking is the creation of the Universal Dependencies framework, which purports to provide a universal dependency grammar-based annotation scheme to create treebanks for any language and thus allow for more direct comparability between treebanks (Nivre, Marneffe, et al. 2016). The framework includes a universal set of part-of-speech tags and dependency relations, as well as a data format for including morphological annotation in the form of features marked on individual words (rather than, as in the Karuk corpus, splitting words into separate morphemes). Included in the over 100 languages with UD treebanks are a few that are comparable to Karuk typologically or in terms of level of endangerment, which are listed below:

- Warlpiri, 55 sentences, 314 words
- Skolt Sami, 10 sentences, 43 words
- North Sami, 3,122 sentences, 26,485 words
- Mbya Guarani, 1,114 sentences, 13,089 words

It is beyond the scope of this section to describe all of the UD dependency relations, but some general differences from the Karuk treebank are discussed here. Readers are encouraged to check the Universal Dependencies annotation guidelines, specifically the list of dependency relations found at `https://universaldependencies.org/u/dep/index.html`. In general, the Karuk treebank uses fewer dependency relations. This is due to the UD system having multiple relations for a particular type of dependency, based on the part-of-speech or other

[11] One should note, however, that the constituency grammar representations of, say, the Penn English Treebank, are not necessarily as abstract or complex as state-of-the-art Minimalist syntactic representations. Penn English Treebank representations are quite "flat" compared to those used in theoretical syntax, and lack many of the functional projections (vP, TP, etc.) expected in such representations. As such, the choice between a constituency grammar and dependency grammar approach is not necessarily as stark as it may seem.

[12] Languages with these linguistic features - morphological complexity and free word order - are reported to do worse in terms of parsing accuracy compared to languages without these features (Nivre, Hall, et al. 2007), though see (Soegaard 2017) for a perspective questioning of this interpretation.

syntactic information about the dependent. For instance, the Karuk treebank uses only a COMP relation for any sort of complement, whereas the UD system has a CCOMP relation for clausal complements and a XCOMP relation for other types of complements. The Karuk system, owing to its basis on a robust morphological annotation, can recover information about whether a complement is clausal or not without needing to add additional relations. The addition of such relations increases the complexity necessary to perform a search which finds all complements, including clausal and non-clausal ones.

Secondly, the Karuk treebank includes annotation of 'syntactic status,' which are syntactically relevant properties that do not involve binary relations, such as PRED or predicate status. Predicates are not given a consistent annotation in the UD system, with main clause predicates given the ROOT relation, and predicates of subordinate clauses or relative clauses given relations that simultaneously describe their clausal status and relation to their head.

Thirdly, the UD system lacks at least one relation that is necessary to adequately annotate the Karuk data, namely POSS or possession. The necessity of this syntactic relation will be discussed in the POSS section of the annotation guidelines below. It should be stated that the UD system can address this deficiency, however, with the use of language-specific subtypes of relations. The UD treebank for Mbya Guarani, for example, includes a language-specific subtype for question particles, which the general UD system does not have.

All in all, the differences between the UD relations and the Karuk Treebank relations are not very substantial; partially that is due to the influence of UD on the Karuk Treebank. Several relations inspired by those used in the UD system were added to the Karuk Treebank in its most recent revision, including the DET relation as a special relation for determiners and demonstratives (which used to be subsumed in the general nominal modification relation, ATR).

A more substantial difference between UD and the Karuk Treebank, and the main reason for the Karuk Treebank not adopting the UD system, is the morphological annotation. The Karuk corpus already has a robust format for annotating the morphology of the language, that includes splitting most multi-morphemic words into their component morphemes, and thus being able to search the corpus for specific morphemes. The UD morphological annotation is comparatively more shallow, involving only features assigned to lexical forms, without decomposing the lexemes into component morphemes. The universal set of features they propose, moreover, does not include enough detail to account for all aspects of Karuk morphology (for instance, the Karuk directional system is extraordinarily complex - see the COMP relation section in the annotation guidelines below for more detail - and there appear to be no UD morphological features related to verbal expressions of directionality). UD allows for the addition of language-specific features, but crafting such a system would be a substantial amount of duplicated effort, given the fact the information is already annotated in the Karuk corpus in a format that better represents the agglutinative tendency of the language. Interestingly, typologically similar languages with UD treebanks, such as Mbya Guarani, do not appear to add very many new morphological features, with the Guarani treebank seemingly only adding the major feature of clusivity, to represent differences between inclusive and exclusive agreement. (This information was gleaned from the project descrip-

tion at this link: `https://universaldependencies.org/gun/index.html`. The annotation guidelines for this treebank (Thomas 2019) do not mention morphological features.)

The differences are not insurmountable, however, and it is hoped that at least a subset of the Karuk treebank can be converted to the UD format (with some loss of information, most likely) to contribute to this important new initiative.

1.3 History of the Karuk Treebank

The Karuk treebank did not begin with this dissertation project. The Karuk treebank was originally conceived by UC Berkeley linguists Andrew Garrett and Line Mikkelsen and development began with funding through the National Science Foundation under Grant No. 1065620 "Karuk [kyh] and Yurok [yur] syntax and text documentation" with a start date of June 15th, 2011 and end date of November 20, 2014. As part of this project, Professors Garrett and Mikkelsen assembled a team of graduate and undergraduate students, to produce a set of annotation guidelines and utilize them to annotate a pilot corpus of around 500 sentences. I worked with them on this project until its end in 2014. The treebank project was dormant until 2016, when I started the first of two LRAP[13] projects to expand the pilot corpus and refine the annotation guidelines. This first LRAP project had two apprentices, Zeynep Özselçuk and Andrew Baker. In 2019, I began a second LRAP project with the goal of completing the annotation of Bright's texts, and this project involved the contributions of 8 undergraduate apprentices: Sammy Keyes-Levine, LaLa Speights-Barhatkov, Stuart Litjen, Jessica Wang, Cindy Yang, Ciara Agrella, Jessica Butcher, and Kevin Yu. Annotation of the texts was completed in the summer of 2020.

1.4 Structure of the dissertation

Following this introduction, Chapter 2: Annotation Guidelines presents the annotation guidelines which were both used as a manual to guide annotators in their annotations tasks and now serve as an explanation and description of the use of every element found in Karuk treebank annotations.

Chapter 3: Argument and Predicate order presents the first case studies utilizing the treebank, focusing on the word order of arguments and predicates. Karuk word order had often been claimed to be 'free' with every or most orders of subject, direct object, and verb attested, but their relative prevalence and the word orders of clauses with other types of argument (complements and indirect objects) or non-verbal predicates is eludicated for the first time in detail in this chapter. Methodologically, I argue in Chapter 3 for the utility of treebanks in allowing easier study of large-scale, quantitative properties of corpora than

[13]LRAP, standing for Linguistics Research Apprenticeship Practicum, is an initiative by the UC Berkeley linguistics department that pairs a graduate student with a research project with undergraduate students interested in being involved in research.

comparable, treebank-less methods. In the case studies meant to showcase this, I describe broad, word order patterns of the treebank corpus and elucidate three trends found in this data: that subordinate clauses tend to have less expressed arguments than main clauses; that subjects are less likely to be expressed in transitive clauses; and that the prevalence of pre-verbal S is driven partially by a likelier-than-expected tendency for the presence of both a subject and object to lead to pre-verbal subjects.

Chapter 4 presents the second set of case studies, focusing on two phenomena where observed agreement in the corpus does not match the agreement expected from Bright (1957)'s description of the agreement system: sentences with plural agreement where singular was expected, and vice versa. These two mismatches (and some inconsistency with one particular agreement prefix for 3pl subjects and 3pl objects) turn out to the only systematic cases that differ from Bright's description of the agreement, thus confirming his accuracy overall. Methodologically, in this chapter I argue for the utility of treebanks in locating rare phenomena in corpora that may escape notice by comparable treebank-less methods. In terms of the case studies, I elucidate in Chapter 4 a heretofore undescribed phenomenon whereby the use of plural subject agreement with a singular subject indicates subject demotion.

Chapter 5 concludes with some thoughts about the future of the treebank project. Appendix A and B contain the full datasets of sentences which informed the case studies in Chapter 4.

Chapter 2

Annotation Guidelines

This chapter contains the annotation guidelines for the Karuk treebank. These guidelines were used by annotators of the treebank to guide their decisions during the annotation process. As such, in many places this chapter will read like a manual, as it was, in essence, used as a manual by the annotators. This chapter will also serve to describe what the annotations are like, which is crucial for being able to use the treebank to investigate the corpus and also to understand the results of such investigations.

These annotation guidelines are structured as follows: Section 2.1 describes what sorts of information are encoded in the treebank and gives a general overview of the syntactic relations and statuses utilized in annotations. Section 2.2 presents some general principles for annotating sentences when multiple possible annotations may be possible. Section 2.3 describes how the dependency graphs which visually represent the annotations of the treebank are meant to be read. Section 2.4 details some definitions of common technical terms that will be used throughout these guidelines. 2.5 details glossing conventions for these guidelines and throughout the dissertation. Section 2.6 presents the syntactic statuses, going through each in order and explaining their use. Section 2.7 does the same for the syntactic relations. Section 2.8 describes the person/number annotation of the treebank. Section 2.9 features discussion of common constructions in the treebank and how they are annotated, and Section 2.10 concludes.

2.1 Types of information encoded in the treebank

- **Syntactic relation**: These are relations that hold between a head and a dependent. The name of the relation expresses the nature of the dependency. A prototypical example is SBJ - the subject relation, which holds between a predicate (the head) and its subject (the dependent). These are binary relations and always hold between two elements - no more, no less. Any given element can only be the dependent in one relation; in other words, each element can only have one head. Any syntactic element can be the head in multiple syntactic relations, however. Unlike in other dependency

grammars, whether something is a predicate is not encoded as a syntactic relation, but rather as a syntactic status.

- **Syntactic status**: These are used to encode syntactically relevant information that does not involve a dependency or relation between two elements. A prototypical example is PRED, the predicate status, which is given to any element in the sentence which is a predicate (not just the root or main predicate of the sentence.) Whether a particular word is a predicate depends not on a single relation it holds with any other particular element, but on a number of facts, including the types of dependencies the word is the head of. A syntactic element can have a syntactic status and a syntactic relation at the same time. A syntactic element can have only one syntactic status.

Syntactic relations	
SBJ	Subject. Holds between predicate and its subject.
OBJ	Object. Holds between predicate and its direct object.
IO	Indirect object. Holds between predicate and its indirect object.
DET	Determiner. Holds between a noun and a determiner.
QUANT	Quantifier. Holds between quantifiers (including numerals) and nominals.
POSS	Possession. Holds between a noun and its possessor.
ATR	Attribute. Holds between a nominal and a modifier.
TAM	Tense/aspect/mood/polarity. Holds between a predicate and a tense/aspect/mood/polarity particle.
NEG	Negation. Holds between a predicate and a negative particle.
COMP	Complement. Holds between a predicate or postposition and its complement.
QUOT	Quote. Holds between a quote-introducing predicate and its quote.
ADV	Adverbial. Holds between a predicate and a modifying adverbial.
SUB	Subordinator. Holds between a subordinate predicate and its subordinator *pa=*
APPOS	Apposition. Holds between two apposed elements.
COORD	Coordination. Holds between two coordinated elements.
COP	Copula. Holds between a non-verbal predicate and copular *uum*.

Syntactic statuses	
PRED	Predicate. Given to all (primary) predicates.
SPRED	Secondary predicate. Given to depictive and resultative secondary predicates.
FRAG	Fragment. Given to (the head of) a fragment.
VOC	Vocative. Given to (the head of) a vocative.
INT	Interjection. Given to interjections.

2.2 General principles for resolving parse multiplicity

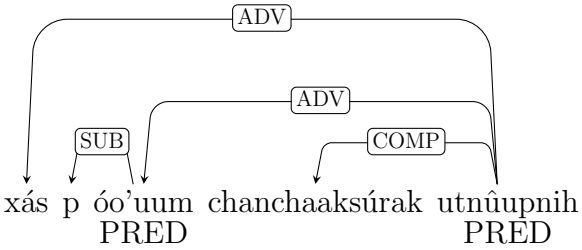
There are a variety of occasions in which a given sentence will have two or more possible parses which both satisfy the demands of the annotation guidelines. The following principles should be used to adjudicate between the options in these ambiguous cases.

2.2.1 Attach High

Often, it may be ambiguous whether a certain word α is a dependent of a particular word β or of another word γ that dominates β (through however many heads). If one is unable to decide between the two parses on linguistic evidence or on the translation, then α should be made a dependent of the highest head to which it can - in the abstract situation outlined above, α should be made a dependent of γ.

A common archetype of this principle at work can be seen in the following example, where a sentence-initial word, the coordinator *xás*, could in principle be an ADV dependent of the subordinate clause verb *óo'uum* or the main clause verb *utnûupnih*.

(1) xás p-óo-'uum, chanchaaksúr-ak u-t-nûupnih
and SUB-3SG-arrive roof.hatch-LOC 3SG-look-through
'And when he arrived, he looked in through the smokehole.' (Nettie Ruben, WB_KL-57:20)



We know that *xás* can be an ADV dependent of the main clause in that initial position, and we have no reason given the translation (where the same ambiguity holds in English of *and*) to choose one verb over the other as the head of *xás*. At this point, the principle ATTACH HIGH comes into play, and we choose the higher of the two options; in this case, the higher option for head is *utnûupnih*, as it dominates the other option, *óo'uum*.

It should be noted that the choice of attaching high, as opposed to, say, attaching to the leftmost option, is an arbitrary choice, but it reflects the generally accepted primacy of hierarchical rather than string-order relations in (generative) syntactic theory.

2.2.2 Follow the Translation

No one who has worked on the treebank is a native speaker of the Karuk language, and as such none of the annotators have had the option of referring to native speaker judgments

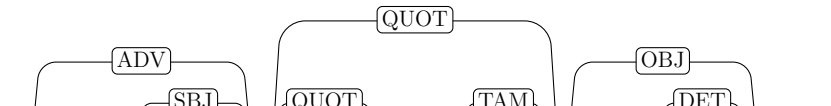
regarding any aspect of the language. This means that the main clue we have to the semantics of each Karuk sentence are the translations provided. These translations, presumably, were made with input from native speakers of the language (though we have little information about the creation of translations for most if not all work on the language prior to the contemporary period). Given the lack of other sources of semantic information, annotations must defer to the meanings indicated by the translation.

This principle becomes particularly important for understanding the person and number of a verb's arguments, as the agreement marking on the verb occasionally does not match the translation. We know that agreement systems cross-linguistically are not "perfect" in that they do not always reflect the actual semantic person or number of arguments, modulo processes which can impoverish agreement or voracious agreement. It is also unknown prior to the writing of this dissertation whether the agreement system described in Bright (1957) is an accurate reflection of agreement across the corpus or language. With an eye towards studying agreement mismatches, the translation must be the guide to determining the person/number of arguments for annotation, else inaccuracies or misunderstandings of the agreement system could become calcified in the very tool meant to study the system.

2.3 Dependency graphs

Throughout this text, dependency graphs or trees (example in (2)) will be used to represent the syntactic annotation of example sentences. Syntactic relations are represented as an arrow from the head to the dependent (a small arrowhead appears above the dependent), with the arrow bearing the label of the relation. Syntactic statuses are written below each word.[1]

(2) xás pihnêefich u-piip, "ishávaas, tá ni-tápkuup pa-mi-pákurih.
and Coyote 3SG-say nephew PER 1SG>3-like the-2SG.POSS-song
And Coyote said, "Nephew, I like your song!" (Nettie Ruben, WB_KL-07:28)



[1]These graphs are created using the LaTeX package tikz-dependency (`https://ctan.org/pkg/tikz-dependency?lang=en`). The LaTeX code is generated by a script I wrote which converts the morphological and syntactic tagging of the treebank into the appropriate code, modulo some manual editing to remove or change characters that could cause a LaTeX error.

2.4 Definitions

In this section I provide definitions of important terms that will recur throughout the text below. These terms may have different meanings than they have in other work, and readers are encouraged to familiarize themselves with the definitions below so as to reduce misunderstanding.

Sentence: A SENTENCE is a unit of Karuk text contained within a <s> element in the Karuk texts XML document. These units do not always conform to what we may call syntactic sentences, especially in the case of texts from Bright (1957), where they generally conform to a prosodic unit. In other words, these could be called database sentences, and if it is necessary to disambiguate they will be referred to as such.

Word: In this text, the term WORD will be used to refer to elements which can enter into dependency relations and bear syntactic statuses. Thus, WORD refers both to strings typically identified as words (i.e. a string between two spaces) and also to those clitics which enter into dependency relations with other words. Essentially, these are words from the syntactic point of view: syntactic words. To ease readability, these will generally be referred to only as words. If the need arises to disambiguate, these will be called syntactic words.

Parse: A parse is the syntactic annotation of a sentence, including all the syntactic relations and syntactic statuses assigned to particular words in that sentence. This will also be sometimes referred to as an 'annotation.'

Dominate: A word α dominates a word β if α is the head of β or is the head of another word γ that dominates β.

2.5 Glossing conventions

Agreement prefixes on verbs are glossed according to the following schema: SubjectPersonNumber > ObjectPersonNumber. For example, a gloss of 3PL>3 indicates agreement for a third plural subject and third person object. If only one set of features is provided (e.g 3PL alone), it means the agreement only tracks the subject (as for an intransitive verb). Possessive prefixes are glossed with the person and number plus .POSS, for example 3PL.POSS for a third plural possessor. Pronouns follow a similar glossing scheme: 3SG.PRO means a third person singular pronoun. Glossing abbreviations are given in the table below:

Gloss	Meaning
ANC	Ancient tense
ANT	Anterior tense
CAUS	Causative
COMPL	Completive
DIM	Diminutive
DUR	Durative aspect
IMP	Imperative
INT	Intensifier
IRR	Irrealis
ITER	Iterative
LOC	Locative
MOD	Modal
NEG	Negative
OBV	Obviative
PAST	Past tense
PER	Perfect aspect
PL	Plural
PL.ACT	Plural Action
PROSP	Prospective aspect
RESULT	Resultative
SUB	Subordinator

2.6 Syntactic statuses

2.6.1 VOC

The syntactic status VOC is given to words which are vocatives. Generally, vocatives are nominals, whether descriptive nouns or proper names, that refer to the addressee of an utterance. Crucially, vocatives have no syntactic relation that ties them to the predicate of the clause they appear to be in. In other words, vocatives are never arguments.

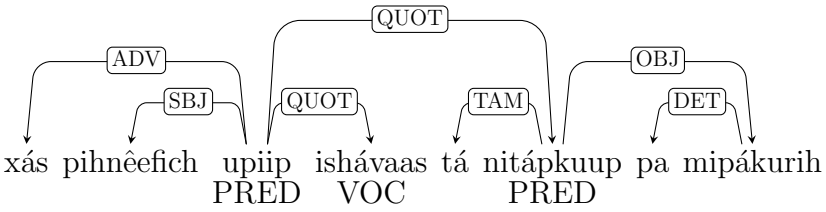
In (3), *êev*, a term of intimate address between women, is used to refer to the addressee of the utterance, just like the 'dear' in the English translation which renders *êev*. In this sentence, *êev* would be given the syntactic status VOC, and have no syntactic relation to any other word in the sentence.

(3) íim êev hûum kích i-xú-tih?
2SG.PRO dear Q only 2SG-think-DUR
'What do you think, dear?' (Nettie Ruben, WB_KL-11:43)

If a vocative is part of a quote, it can bear a syntactic relation. In (4) below, the noun *ishavaas* 'nephew' refers to the addressee of Coyote's utterance, but is not syntactically

related to the predicate of its apparent clause, *ni-tápkuup*.

(4) xás pihnêefich u-piip, "ishávaas, tá ni-tápkuup pa-mi-pákurih.
and Coyote 3SG-say nephew PER 1SG>3-like the-2SG.POSS-song
And Coyote said, "Nephew, I like your song!" (Nettie Ruben, WB_KL-07:28)



In this case, *ishávaas* is a dependent of the main predicate *upiip*, by virtue of the vocative being included within a quote introduced by *upiip*.[2] As in (3), this is not always the case, though vocatives are generally contained within quotes in the narratives which form the core of the Karuk corpus.

One may be moved to consider *ishávaas* in (4) as the possessor of *pamipákurih*, but this would be a mistake. Applying the principle FOLLOW THE TRANSLATION, we see that *ishávaas* is not treated as the possessor of 'your song', or the translation would have read "I like Nephew's song." Such a translation would indicate a non-vocative use of *ishávaas*, as being a vocative is incompatible with having syntactic relations to any other word in the vocative's immediate clause.

2.6.2 PRED

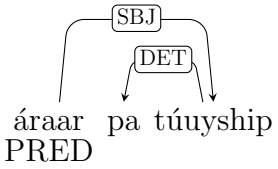
The syntactic status PRED is given to predicates. Generally, each sentence will have at least one word with the PRED status, though there are occasionally cases (such as with fragments, see the FRAG syntactic status below) where a sentence will not have any PRED words. There are two classes of PRED words: verbs, and non-verbal predicates (which can be nouns or adjectives.)

Words with the PRED status are the only words which can have SBJ or OBJ dependents.

Certain words will always have a PRED status. Verbs with person-number agreement will always have the PRED status, whether they are the verb of the main clause or of a subordinate clause. In other cases, a word should be marked as the predicate if it is identifiable as such in the translation. For example:

[2]The result of *ishávaas* being a QUOT dependent is that *upiip* has two QUOT dependents. Because a quote, unlike a complement clause, can be made up of many independent clauses or phrases, it would be an error to say, just in quotes, that the vocative in these cases has a syntactic relation to the predicate of its clause. Having two QUOT dependents accurately reflects the independence of the vocative from the other material in the quote.

(5) áraar pa-túuyship
person DET-mountain
'The mountain is a person.' (WB_-KL-83)



In the English translation, 'person' is the predicate of this sentence. It is the property which is being asserted of 'the mountain.' Based on that, *áraar* 'person' should be given the PRED status, and *patúuyship* 'the mountain' should be made a SBJ dependent of it.

2.6.3 FRAG

The syntactic status FRAG is given to the heads of fragments (in other words, the status is given to the word in a fragment that dominates all other words in the fragment.) This syntactic status is essentially a catch-all to be used when no word in a particular clause-level string can be identified as PRED, VOC or INT. This can include such elements as noun phrases which are not clearly predicates or arguments of predicates, or vocables without lexical meaning (such as used in songs).

In (6), both instances of *ruup* are given the syntactic status FRAG.

(6) "rúup, rúup"
rope, rope
'Rúup, rúup' (WB_KL-66:16)

In (7), each of the nonsense words *kitâana kitâana íiyaa*, representing a song, are given the FRAG syntactic status.

(7) xás u-'árihishrih-een pa-mu-pákurih "kitâana kitâana íiyaa"
and 3SG>3-sing-ANT the-3SG.POSS-song kitaana kitaana iiyaa
'So he sang his song, "kitâana kitâana íiyaa" (WB_KL-07:31)

2.6.4 SPRED

The syntactic status SPRED is given to secondary predicates, generally encountered in resultive or depictive secondary predication. Generally, secondary predicates are adjectives or uninflected verbs. For example, in (8) below, the adjective *aachíchhar* is acting as a depictive secondary predicate, describing the state of the runners (the subject of the verb *kunpihmarápiithva*).

(8) ... aachíchhar vúra kun-p-ihmará-piithva
happy INT 3PL-ITER-run-around.PL.ACT
'They ran around again happily.'

SPRED is also given to the complements of *ikyav* in the periphrastic causative construction, discussed in section 2.9.8.

2.6.5 INT

The syntactic status INT is given to interjections. Generally, interjections are exclamations, and are used as standalone syntactic units. They generally do not have dependents, and unless quoted, also are not generally dependents of any other word. There are often cases of multiple interjections in a single sentence. For example:

(9) chémi, chôora.
all.right, let's.go
'All right, let's go.' (WB_KL-33)

In (9), both words are interjections. They have no clear syntactic relationship to each other, and are not quote dependents of any other word in their sentence (though, in this case, they are inside a multi-sentence quote.)

Like vocatives described in section 2.6.1 above, if interjections are included in a quote, they are treated as being separate QUOT dependents of the quotative verb, which will generally cause that particular verb to have multiple QUOT dependents.

A non-exhaustive list of some typical interjections is given below:

- *hãã̃* 'yes'
- *pûuhara* 'no'
- *ayukîi* 'hello'
- *chémi* 'alright'
- *yôotva* 'thanks'
- *chôora* 'let's go'

2.7 Syntactic relations

2.7.1 SBJ

The SBJ relation is given to subjects, and is a relation between the subject and whichever predicate it is the subject of. In fact, all SBJ dependents must have a head with the PRED status.

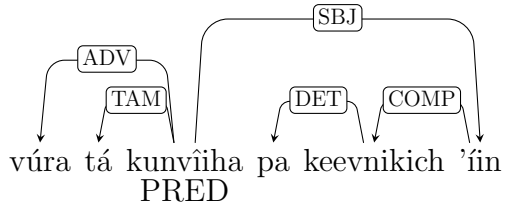
Subjects are generally the most agent-like argument of a predicate. In most cases, subjects in Karuk sentences can be accurately identified by comparing to the English translation, in that the subject of the English translation most often also is the subject of the Karuk sentence. In some cases, however, the English translation is less faithful to the observable syntax of the Karuk sentence, and merely following the translation will produce the wrong results. Thus, it is generally best practice to use a combination of the translation and, in the case of verbal predicates, the subject agreement on the verbal predicate in question to identify the subject.

There are some sub-cases where there is a consistent mismatch between the subject of the English translation and the subject of the Karuk sentence. These cases involve a class of 'subject postpositions' which affect the agreement patterns on verbs. These subject postpositions, a subset of the postpositions in the language, are:

- *xakaan* 'together (with one other person)'
- *koovan* 'together (with two or more)'
- *îin* OBVIATIVE[3].

These postpositions only ever take as complements the 'semantic' subjects of whichever predicate is heading the relevant clause - generally, the subject as indicated in the translation will be the complement of these postpositions. For reasons discussed below, these cases are analyzed as the subject postposition bearing the SBJ relation to the predicate, and the putative subject noun bears the COMP relation to the subject postposition. This can be seen in (10) below, where the final morpheme in the sentence is *iin*, one of the subject postpositions[4], the SBJ dependent of the predicate *kunvîiha* is *íin*, which has a dependent, the 'semantic' subject, *pakeevnikich*, bearing the COMP relation.

(10) vúra tá kun-vîiha pa-keevnikich-'íin
INTENS PER 3PL>3-hate the-old.woman-OBV
'The old woman disliked him.' (Nettie Ruben, WB_KL-57 'The Boy from Itúkuk')

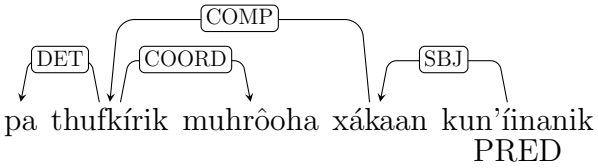


[3]For more information on *îin*, see Macaulay (2000)

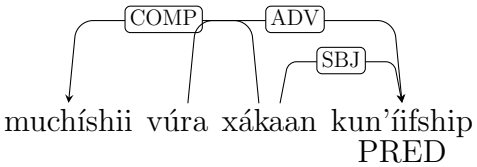
[4]In this particular example, *iin* is a clitic and forms a single phonological word with its complement. Other postpositions, such as *mûuk* 'with (instrumental)' display the same ability to form a single phonological word with their complement.

The postpositions *xakaan* and *koovan* appear with apparently coordinated noun complements, though overt coordinators are rare (if found at all) in these cases (11). In these cases, the nouns are treated as coordinated (see the section on the COORD relation below), and the leftmost of them (the head of the coordination) is the complement of the postposition. Even in cases where these postpositions occur with only one expressed complement, the translation will generally indicate that the expressed complement is to be understood as coordinated with a dropped pronoun (12).

(11) pa-thufkírik mu-hrôoha xákaan kun-'íin-anik
the-owl 3SG.POSS-wife together 3PL-live-ANC
'Owl and his wife lived together.' (Mamie Offield, WB_KL-24, 'The Greedy Father')

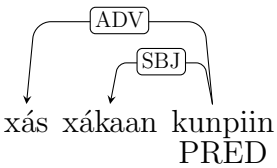


(12) mu-chíshii vúra xákaan kun-'íifship
3SG.POSS-dog INTENS together 3PL-grow.up
'He and his dog grew up together.'(Mamie Offield, WB_KL-54, 'Shinny Game Medicine')



In some cases, one of the subject postpositions is used without any complement. In these cases the postposition is still the bearer of the SBJ relation to the relevant predicate. Thus, in (13), despite having no complement (the putative subject pronoun being null), *xákaan* is the SBJ dependent of the predicate *kunpiin*.

(13) xás xákaan kun-p-iin
and together 3PL-ITER-live
'And they lived together again.' (Lottie Beck, WB_KL-61, 'The Kidnapped Child')



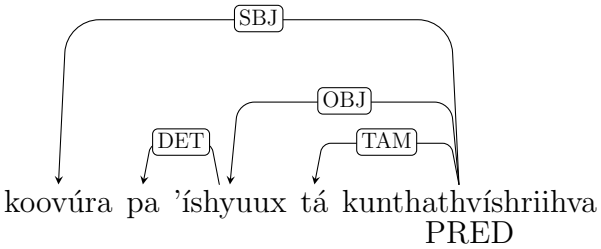
The examples above all showcase another feature of these subject postpositions: they generally cause the verbal agreement to show a plural subject, even in cases with *iin* where the complement of the postposition is clearly a single entity as in (10). The ability of the postpositions to trigger a change in agreement and the restriction to only taking putative subjects as complements are the reasons these are viewed as being the head of the subject, rather than being relegated to being the head of an adverbial PP or the like.

2.7.2 OBJ

The syntactic relation OBJ is given to direct objects of verbal predicates, with the relevant predicate as the head. Each predicate may have only one OBJ dependent.

Generally, the direct object is the more patient-like of the arguments, and is generally a nominal.[5] As for subjects, in general the direct object can be accurately identifying by matching the Karuk phrase with the direct object of the English translation, though annotators should always be sure such an analysis would not contradict any observable facts about the Karuk sentence in question. (14) is an example of the typical case, where the bearer of the OBJ relation is *íshyuux* 'elk' as the head of its noun phrase, which corresponds with 'the elk' being the direct object of the English translation.

(14) koovúra pa-'íshyuux tá kun-thathvishriih-va
all the-elk PER 3PL>3-carry.home-PL.ACT
'They all carried the elk home.' (Nettie Ruben, WB_KL-72:10)

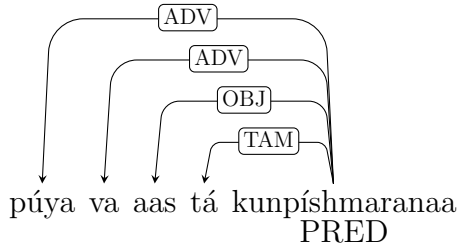


In (15), there is a case where the English translation does not match the observable Karuk syntax. (15) includes a Karuk idiom, *aas ... ísh*, literally rendered as 'drink water' but which is used to describe eating a meal. The suffix *-mara* on the verb indicates that the action is finished. The English translation chooses to render 'finish' as the verb, which does not represent the verb in Karuk, and uses the idiomatic reading of 'eat' rather than translating *aas ... ísh* literally. In this case, understanding the Karuk idiom to have *aas* as the direct object, we obtain the correct parse below.

(15) púya-va aas tá kun-p-ísh-mara-naa
so-thus water PER 3PL>3-ITER-drink-COMPL-PL

[5] See 2.9.3 for annotating objects without an expressed noun.

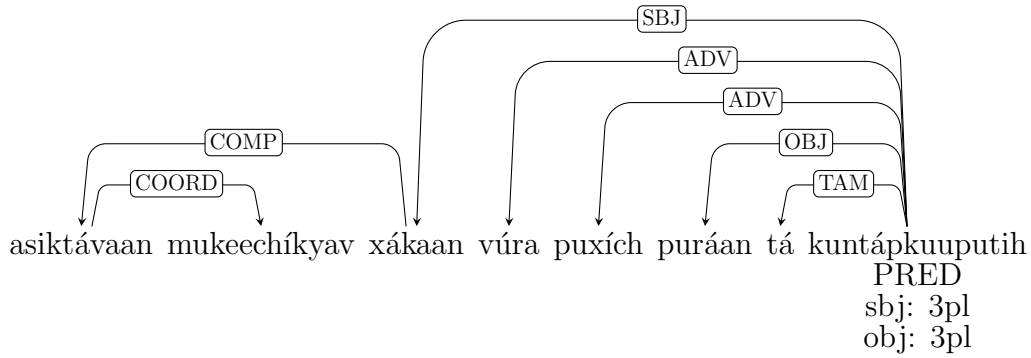
'Then they finished eating.' (Nettie Ruben, WB_KL-72:6)



In any case, all expressed noun phrases in the Karuk sentence must be included in the parse, and if there is a mismatch in the number of nominals between the Karuk and English sentences as in this case, that is a clear sign that the English translation cannot be a complete guide to the annotation.

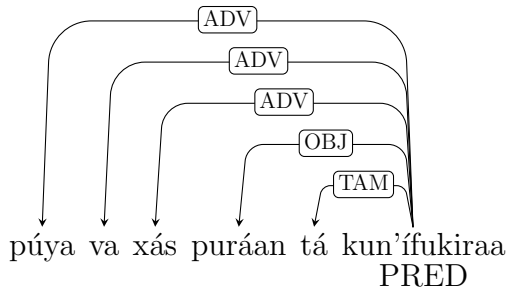
Like the subject postpositions discussed above in section 2.7.1, there is an object postposition, the reciprocal *puraan* 'each other.' This is called a postposition based on its morphological similarity with *xakaan*, though it has not been observed to actually take any complements. Regardless, when it appears, it is marked as the OBJ dependent of the relevant predicate. It must be an OBJ rather than SBJ because it can co-occur with *xakaan* as in (16). As described in the subject section above, *xakaan* must be the SBJ dependent, so the next argument slot available for *puraan* is OBJ. Also, the semantic contribution of *puraan*is clearly object-related: it says that the entities which make up the subject are acting on each other - it is telling us about how the patient argument or object is constituted.

(16) asiktávaan mu-keechíkyav xákaan vúra puxích puráan tá
woman 3SG.POSS-sweetheart together INT much each.other PER
kun-tápkuupu-tih
3SG.POSS-like-DUR
'A woman and her sweetheart loved each other very much.' (Mamie Offield, WB_KL-58:1)



Note that *puraan* does not only occur with *xakaan*, but it is given an OBJ dependent even when it occurs alone, as in (17).

(17) púya=va xás puráan tá kun-'ífukiraa ...
so=thus then each.other PER 3PL>3-grab ...
'Then they grabbed each other...' (Julia Starritt, WB_KL-78:14)

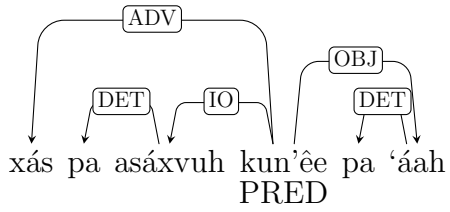


In some cases it may be difficult to tell if a given word is a OBJ dependent or a COMP dependent. Advice for distinguishing between the two is provided in the COMP section below.

2.7.3 IO

The relation IO is given to indirect objects of ditransitive verbal predicates, with the relevant predicate being the head of the relation. Most ditransitives are verbs which denote a transfer of possession, and the indirect object in these cases is generally the entity which gains possession of whatever is being transferred through the event (though, some verbs treat the one who gains possession as the subject, in which case the IO dependent is the one who had possession before the transfer). In (18), *asáxvuh* 'Turtle' gains possession of the direct object *aah*, 'fire', and is thus the bearer of the IO relation.

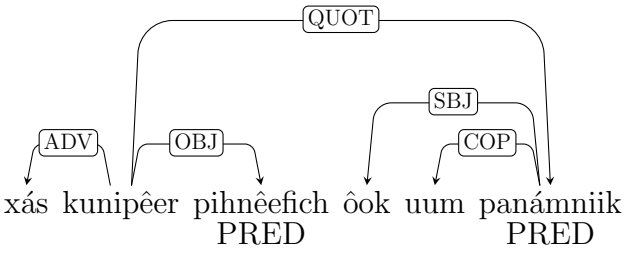
(18) xás pa-'asáxvuh kun-'êe pá-'aah
and the-turtle 3PL>3-give the-fire
'And they gave the fire to Turtle.' (Julia Starritt, WB_-KL-10:53)



The IO relation is only used with verbs which can potentially have three nominal arguments (even if these are dropped), like in the case of (18) above. This is in distinction to verbs which have two nominal arguments and a clausal argument, like *ipêer*, 'tell,' where

the non-agent nominal argument is given the OBJ relation. In (19), the verb *ipêer* has three arguments (two of them expressed): a dropped 3PL pronoun (indexed by the agreement prefix *kun-* on the verb) who are the speakers in the telling event described; the addressee of the telling event, which is *pihnêefich* 'Coyote' and is the direct object; and a quote told to Coyote.

(19) xás kun-ipêer pihnêefich "ôok uum panámniik"
and 3PL>3-tell Coyote here 3SG.PRO Orleans
'And they told Coyote, "Here's Orleans"' (Julia Starritt, WB_KL-04:175)



In essence, the IO relation is given only when the verb in question could also have had a separate OBJ dependent representing the entity which is transferred from the subject to the indirect object.[6] If instead the verb has a COMP dependent or QUOT dependent as above, the IO relation will not be used, to ensure that it is used only for clear cases of indirect objects.

A potential further aid to identifying the indirect object is the agreement on the verb. If the indirect object is a local person (i.e. either 1st or 2nd person), the agreement prefix will reference the indirect object rather than the direct object.[7] In (20) below, the agreement prefix on the verb, *nu-*, is used when the subject is 1st person, and the object (or indirect object in this case) is 2nd person. (Note, there is no expressed indirect object in this sentence, and thus there would be no IO dependent of the verb, but the agreement prefix still exhibits the relevant behavior.)

(20) miník koovúra nu-'ákih-eesh pa-nini-'ararí'uup
of.course all 1>2-give-PROSP the-1SG.POSS-treasure
'I'll give you all my Indian treasure.' (Mamie Offield, WB_KL-62:25)

[6]Note that this principle is not saying IO can only be used if there is an expressed OBJ dependent. Rather, it is whether the verb in question could possibly have an OBJ dependent, even if said object is dropped in the sentence being annotated.

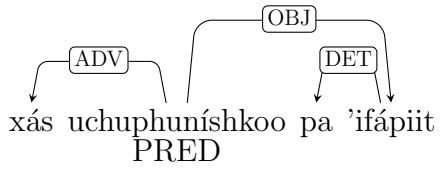
[7]No cases have been found where the direct object is a local person and the indirect object is 3rd person. Such a case would tell us whether the agreement is sensitive to local persons (if the prefix referenced the local person), or just indirect objects over direct objects (if it referenced the 3rd person indirect object). As it is, in cases of 3rd person direct and indirect objects, it is generally not possible to tell which of the objects the agreement is referencing.

2.7.4 DET

The DET relation, standing for 'determiner' is given to three words which act as determiners in noun phrases: the proximate demonstrative *pay*, the definite determiner clitic *pa=*, and rarely, the distal demonstrative *vaa*. These three are united in that they are all used to help specify the referent of a particular noun phrase, though, as will become clear in the following discussion, they are not entirely syntactically parallel (due in particular to the nature of *pa=* as a clitic). The head of the relation is the lexical noun (or whatever the head of the noun phrase is, if it is a headless noun phrase) that the determiner modifies.

The most typical case for the definite clitic *pa=* is that it is directly attached as a prefix to the lexical noun, as in (21), with *pa=* attached to the noun *ifápiit*. *pa-* bears the DET relation, the head of which is *ifápiit*.

(21) xás u-chuphuníshkoo pa-'ifápiit
and 3SG>3-talk.to the-girl
'Then he talked to the young woman.' (Nettie Ruben, WB_KL-57:9)

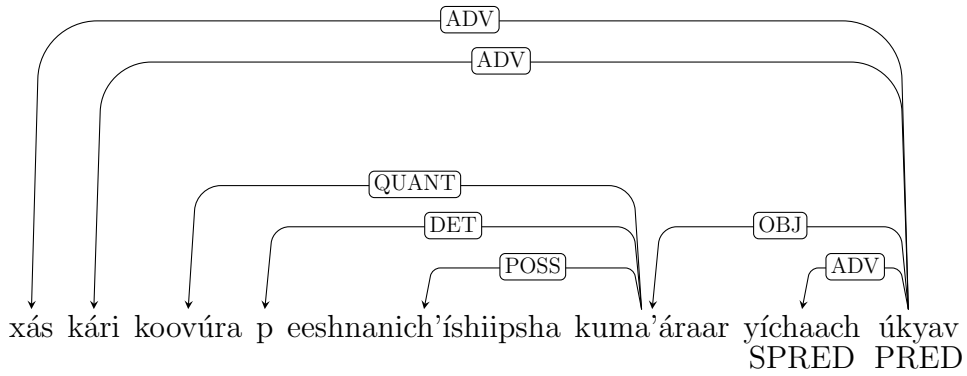


As a clitic, *pa=* can appear farther to the left of its head noun and attach to other words, though it must be attached to another word dominated by the head noun (in other words, it must attach to something else in the same noun phrase.) In (22), *pa=* attaches to *ishnanich'íshiipsha* 'swiftest-PL',[8] which is an adjective dependent of the noun[9]. *pa=*, too, is a modifier of the noun, not the adjective alone, and thus bears the DET relation, with the noun as the head, even though it is not directly attached to the noun.

(22) xás kári koovúra p-eeshnanich-'íshiip-sha kuma-'áraar yíchaach ú-kyav
and then all the-swift-most-PL kind.of-person together 3SG>3-make
'So he gathered together all the swiftest people.' (Julia Starritt, WB_KL-10:15)

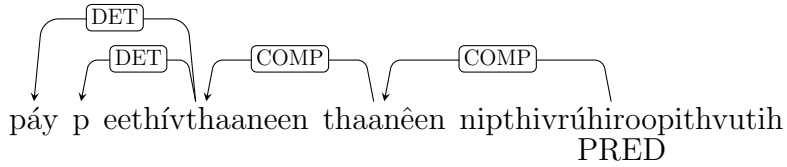
[8]The *ee* vowel at the beginning of *ishnanich'ishiipsha* is the result of vowel coalescence with the *a* in *pa=*.

[9]The particular relation the adjective has may be surprising. It is a POSS dependent, rather than an ATR dependent, due to the presence of the 'impersonal possessive' prefix *kuma-* on the noun *kuma'áraar*



The proximate demonstrative *pay* also generally appears to the left of the noun it is a dependent of, but as an independent word. In (23), it appears to the left of *eethívthaaneen* 'world' and is a DET dependent of it. Note that *eethívthaaneen* has two DET dependents, because *pa=* can coexist with the demonstrative.

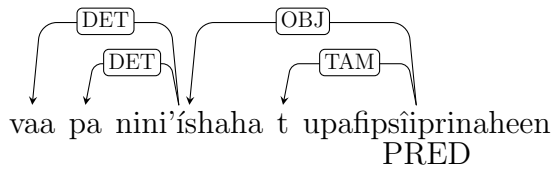
(23) páy p-eethívthaaneen thaanêen ni-p-thivrúh-iroopith-vu-tih
this the-world around 1SG-ITER-float-around-PL.ACT-DUR
'I float around and around this world.' (Chester Pepper, WB_KL-55:52)



páy has uses in a variety of idiomatic expressions or common collocations that do not involve being a DET dependent. It is generally safest to assume a DET interpretation for *páy* only if the translation uses 'this' as a determiner for one of the nouns.

The distal demonstrative *vaa* is rarely used as a determiner, but has a distribution like *páy* in these uses: occurring to the left of the noun as an independent word, as in (24).

(24) vaa pa-nini-'íshaha t-u-pafipsîiprin-aheen
that the-1SG.POSS-juice PER-3SG>3-finish.off-ANT
'He's drunk up that juice of mine.' (Chester Pepper, WB_KL-03:25)



As with *páy*, *vaa* is used in ways that do not involve being a DET dependent. *Vaa* is

marked as DET only if the translation uses 'that' as a determiner for a noun, as in (24). The most typical use of *vaa* is as an adverbial modifier generally meaning 'thus.'

2.7.5 QUANT

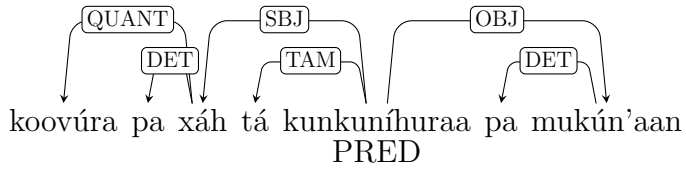
The QUANT relation is given to quantifiers in noun phrases, with the lexical noun which is quantified over being the head. The class of words which are generally given this relation includes:

- Numerals, such as *yítha* 'one', *áxak* 'two,' and so on.
- Derivatives of numerals, such as *itahanatapasich* 'a lot' (< *itráhyar* 'ten' + *-tapas* 'real' + *-ich* DIM)
- The universal quantifiers *koovúra* and (more rarely) *koo*
- The quantifier *táay*, 'many, a lot'
- The existential quantifier *kaakum* (only rarely used)

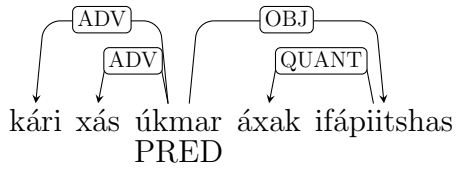
These are united in all expressing something about the quantity of the noun phrase they modify, though they may not necessarily have exactly parallel syntactic structure.

Typically, quantifiers precede the nouns they modify (as with *koovura* 'all' preceding *xáh* in (25) and *áxak* preceding *ifápiitshas* in (26)). (25) also showcases that a single noun can have both a QUANT dependent and a DET dependent.

(25) koovúra pa-xáh tá kun-kuníh-uraa pa-mukún-'aan
all the-spider PER 3PL>3-shoot-up the-3PL.POSS-string
'All the spiders shot up their string.' (Mamie Offield, WB_KL-33:58)

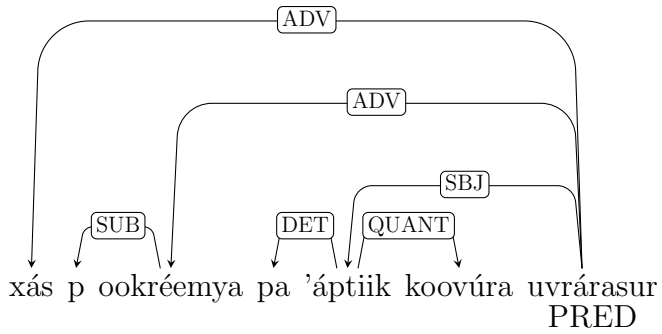


(26) kári xás ú-kmar áxak ifápiit-shas
and then 3SG>3-meet two young.woman-PL
'And he met two young women.' (Mamie Offield, WB_KL-09:2)

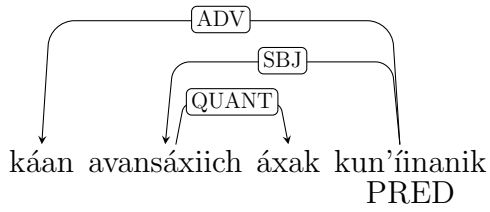


It is possible, however, for the order to be switched, and the noun to precede the quantifier, generally only if both of them are preverbal.

(27) xás p-oo-kréemya pa-'áptiik koovúra u-vrárasur
and SUB-3SG-blow the-branch all 3SG-fall.off
'And when it blew, the branches all fell off.' (Daisy Jones, WB_KL-20:32)

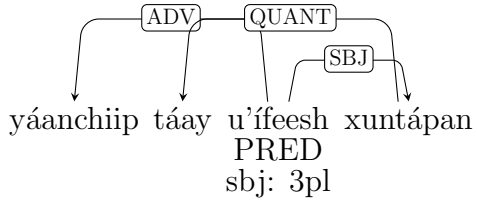


(28) káan avansáxiich áxak kun-'íin-anik
there boy two 3PL-live-ANC
'Two boys once lived there.' (Julia Starritt, WB_KL-08:1)



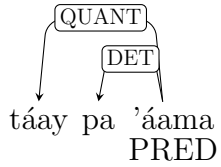
Additionally, quantifiers and their head nouns can be discontinuous: words that are not part of their phrase can intervene between the quantifier and noun. (29) exhibits this possibility, with the quantifier *táay* 'many' and the noun it modifies *xuntápan* 'acorn' not adjacent to each other, with the word *u'ífeesh*, which is not a part of the quantified noun phrase, intervening. The majority of cases involve the quantifier preceding the noun and the intervening word being the verb, but other patterns exist as well; see Maier (2016b) for more details.

(29) yáanchiip táay u-'íf-eesh xuntápan
next.year much 3SG-rise-PROSP acorn
'The next year many acorns will grow.', (Chester Pepper, WB_KL-47:18)



A common pattern that deserves mention is the use of quantifiers with predicational nouns in non-verbal predication sentences. These are analyzed just as quantified noun phrases which act as arguments like above, with the quantifier as a QUANT dependent of the noun. Thus, in (30), *táay* is a QUANT dependent of *áama* 'salmon.'

(30) táay pa-'áama
many the-salmon
'There were a lot of salmon.' (Daisy Jones, WB_KL-20:14)



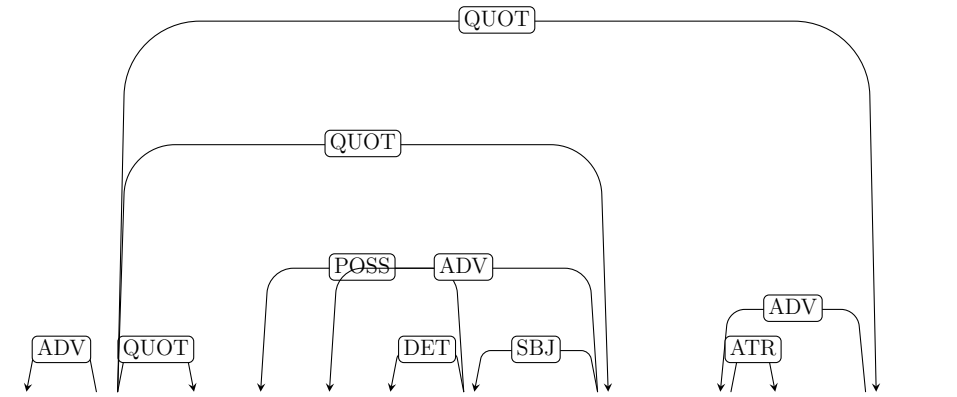
An alternate analysis not adopted here would be to treat the quantifier as the predicate. However, since we know that noun phrases can be predicates and that quantifiers can be dependents of nouns (and thus inside the noun phrase), it is reasonable to think the two patterns can combine and that a quantified noun phrase can be a predicate itself. Furthermore, treating this construction like other quantified noun phrases allows users of the treebank to find these cases in a general search for quantified noun phrases. Making the quantifier the predicate would leave it so there is no quantification relation between the quantifier and the noun (as the noun would be a SBJ dependent of the quantifier) and these cases would no longer be discoverable in a search for quantified noun phrases.

2.7.6 POSS

The syntactic relation POSS is given to possessors of noun phrases, with the possessum as the head of the relation. As Karuk is a generally head-marking language, possessors themselves are not marked by any particular morphology, but are generally free-standing noun phrases which are indicated in the translation as being a possessor. The possessum is these cases is

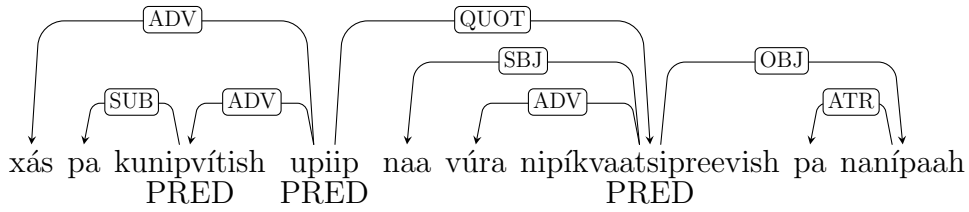
marked with a prefix indicating the person/number of the possessor. For example, in (31), *naa*, the 1st person singular pronoun, is the possessor of *nanipaa* 'my boat', where *nani-* is the 1st person singular possessive prefix. Thus, *naa* is a POSS dependent of *nanipaa*.

(31) xás u-piip "pûu, naa vúra pa-nani-paa nîinamich, hôoy íf
and 3SG-say no 1SG.PRO INT the-1SG.POSS-boat small where truly
nu-yâah-eesh
1PL-do.well-PROSP
'And he said, "No, my boat is little, we won't fit.' (Nettie Ruben, WB_KL-57:89)



Like with other noun phrases, Karuk can very freely drop possessors and only have the possessive prefix there to indicate possession. In these cases, there will be no POSS dependent. In other words, the POSS relation can only be used if there is an explicit possessor. Care must be taken to ensure that any putative possessor is actually the possessor, and not potentially filling in some other role in the clause. For example, in (32) there is a possessed noun phrase, *nanípaah* 'my boat', and like in (31), there is an expressed 1st person pronoun, *naa*. However, it would be a mistake to make *naa* a POSS dependent of the noun, as there is another potential parse: *naa* could be the subject (SBJ dependent) of the verb *nipíkvaatsipreevish*. Following the principle of ATTACH HIGH, *naa* should be analyzed as the subject, since that would make it a dependent of a higher head, the verb (which is higher, since it has the possessed noun as a dependent itself.)

(32) xás pa-kun-ip-vít-ish u-piip "naa vúra
and SUB-3PL>3-ITER-paddle-RESULT 3SG-say 1SG.PRO INT
ni-p-íkvaatsipree-vish, pa-naní-paah
1SG>3-ITER-put.on.shoulder-PROSP the-1SG.POSS-boat.
'And when he had beached his boat again, he said "I'll pick up my boat."' (Nettie Ruben, WB_KL-57:108)



Why POSS?

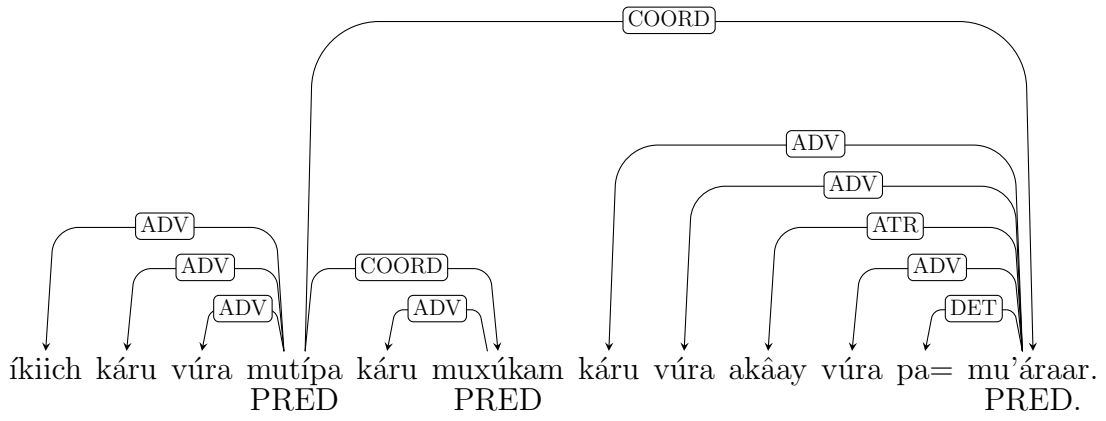
In the dependency grammar used in the Karuk Treebank, there is both a ATR syntactic relation, for nominal modifiers, and a POSS relation, which is used only for possessors. In a general sense, it is not unreasonable to think that possession is merely a type of nominal modification and the system could be simplified by collapsing POSS with ATR. Indeed, the system of Universal Dependency treats possession as merely another case of ATR (nmod in their labelling) and has no separate relation for possession. Why is this same scheme not adopted for Karuk Dependency Grammar? In short, collapsing ATR with POSS would cause an unacceptable loss of information, in that it would no longer be possible to isolate, using information encoded in the treebank, possessive relations from other types of modification.

Why is it important to distinguish possessive relations from other types of nominal modification? For one, there is the general principle guiding the treebank that all examples of potentially syntactically important constructions should be able to be isolated using search terms provided by the information encoded by the treebank, in conjunction with lexical and morphological annotation contained in the Karuk corpus. Insofar as there may be syntactic differences between possession and say, adjectival modification, the treebank should be equipped to separate out possessive structures from others. Indeed, research on Karuk (cf. Maier (2016b)on split NPs) has tentatively suggested there may be some differences between possessive structures and other types of nominal modification, in terms of their tendency to split.

The previous discussion only argues for possession being able to be isolated, not that this isolation must result from the introduction of a separate syntactic relation. Given that Karuk possession is head-marking, with a posessive prefix on the possessed noun, one might expect that possessive structures could be isolated by a search which combined looking for ATR relations between nouns, where one noun has a possessive prefix - thus utilizing the part-of-speech tagging and morphological annotation in addition to syntactic relations provided by the treebank to isolate this particular construction. However, there are cases that exactly match those search criteria, where the relation between the nominal ATR dependent of the possessed noun is nonetheless not the possessor of the head noun. Consider the following example:

(33) íkiich káru vúra mu-típa káru mu-xúkam káru vúra akâay
maybe and INTENS 3SG.POSS-brother and 3SG.POSS-uncle and INTENS anyone

vúra pa=mu-'áraar.
INTENS DET=3SG.POSS-relative
'Maybe it would be his brother or his uncle or any relative of his.' (Julia Starritt, WB_KL-0:13)

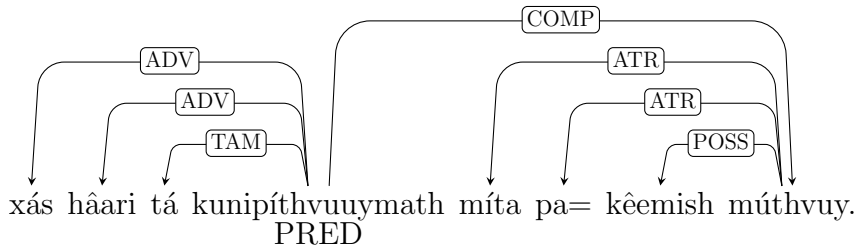


The crucial phrase in this sentence is *akâay ... pa=mu-'áraar* 'any relative of his.' Note that the head noun, *pa=mu-'áraar* contains *mu-*, the possessive prefix indicating a 3rd person possessor, and has a ATR dependent, *akâay* 'anyone'. However, based on the translation of this sentence, it is clear that *akâay* is not the possessor of *pa=mu-'áraar* - such a meaning would instead be rendered as "anyone's relative," rather than ' any relative of his' as it is rendered. It is clear from the context of this sentence that a reading of *akâay* as the possessor cannot be the intended meaning. The text from which this sentence originates, titled 'Swearing,' concerns the traditional Karuk taboo against saying the name of a deceased person. As described in the text, if someone commits this act, the family of the deceased person would ask for remuneration, and if this was rejected, would potentially kill the offender or one of the offender's relatives as retribution. It is inconceivable that they would kill some random person's relative, which a reading such as 'anybody's relative' would seem to suggest, given that the purpose is retaliation against a specific offending person. Based on this, we can be reasonably sure that *akâay* is not the possessor of *pa=mu-'áraar*, but instead is acting somehow as a modifier which, in this case, could be translated into English as *any*.

If there was no separate POSS syntactic relation to separate out clear examples of possession (as in (34)) from examples like (33) above, any search for possessive structures based on the combined part of speech, morphological, and syntactic information above would produce results which include a false positive - namely, (33) above.

(34) xás hâari tá kun-ip-íthvuuy-math míta pa=kêemish
and sometimes PERF 3PL>3-ITER-be.called-CAUS former DET-deceased.person
mú-thvuy.
3SG.POSS-name

'Sometimes they name someone again with the name of the deceased.' (Julia Starritt, WB_KL-0:4)



2.7.7 ATR

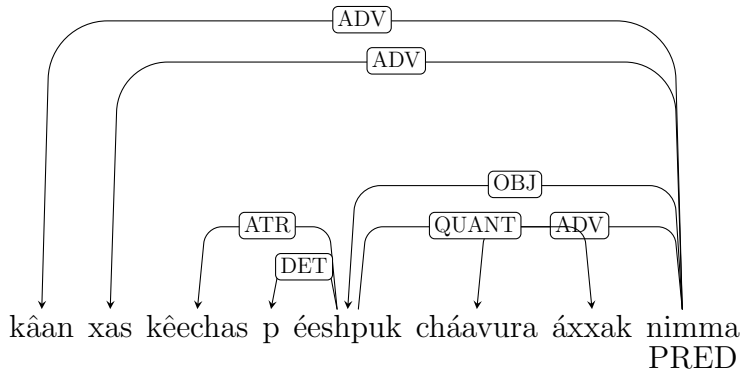
The ATR relation, standing for ATTRIBUTE or ATTRIBUTIVE, is a general category for types of nominal modification not covered by the DET, QUANT, or POSS relations discussed above. As such, the ATR relation is used for adjectival modification, nominals acting as modifiers for other nominals, conjunctions which coordinate noun phrases, and relative clauses, as well as any other case not mentioned elsewhere which involves modification of a noun by some other word. The function of the ATR relation, as a catch-all for modification in the nominal domain, is analogous to the ADV relation in the verbal/predicational domain.

Adjectival modification

Attributive adjectival modification is relatively rare in the Karuk corpus (as opposed to predicational uses of adjectives) but not unattested. (35) shows one example, where *kêechas* 'big-PL'[10] is modifying *éeshpuk* 'money' and is thus an ATR dependent of it.

(35) kâan xas kêech-as p-éeshpuk cháavura áxxak ni-mma
there and big-PL the-money finally two 1SG>3-find
'And right there I found two big gold nuggets' (Benonie Harrie, DAF_KT-05b:10)

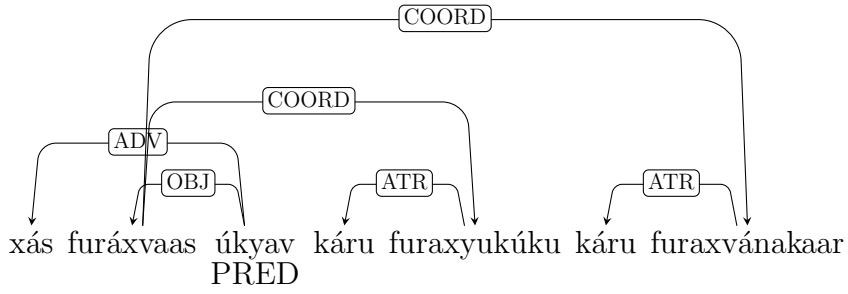
[10] Along with animate nouns, adjectives can be plural-marked in Karuk.



With *káru*

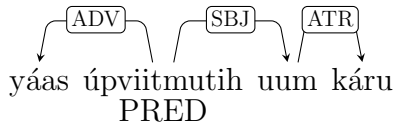
The ATR relation is also used for the dependency between the coordinator *káru* and a noun phrase in a coordinate structure. In these uses, *káru* precedes a noun phrase which is coordinated, as with *káru* preceding *furaxyukúku* 'woodpecker-scalp shoes', and is an ATR dependent of *furaxyukúku*. For more information about how coordination is annotated, see the COORD section 2.7.15 below.

(36) xás furáx-vaas ú-kyav, káru furax-yukúku káru
and woodpecker.scalp-blanket 3SG>3-make and woodpecker.scape-shoe and
furax-vánakaar
woodpecker.scalp-type.of.shirt
'So he made a woodpecker-head blanket, and woodpecker-head shoes and a woodpecker-head vánakaar (a shirtlike garment).' (Nettie Ruben, WB_KL-57:130)



káru can also be used in cases without coordination, meaning 'also.' In these uses, it follows the noun phrase it modifies, but retains the ATR relation with the noun as the head. Thus, in (37), *káru* is an ATR dependent of the pronoun *uum*.

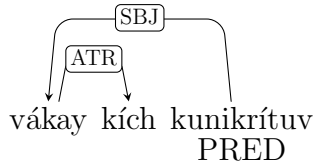
(37) yáas ú-p-viit-mu-tih, uum káru.
then 3SG-ITER-paddle-to-DUR 3SG.PRO also
'Then he paddled back there too.' (Nettie Ruben, WB_KL-57:107)



With kích

When modifying a noun, the focus particle *kich* 'only' generally closely follows the noun it modifies. In this usage, it is a ATR dependent of the noun. Thus, in (38), *kích* is an ATR dependent of *vákay* 'worm.'

(38) vákay kích kun-ikrítuv
worm only 3PL-be.in.pile
'Only worms lay there.' (Nettie Ruben, WB_KL-57:149)

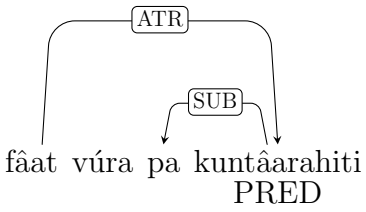


Relative clauses

Relative clauses are subordinate clauses which modify nouns. They exhibit a similar structure to subordinate clauses which act as adverbials or complements to verbs, in that the predicate of one generally has a *pa=* subordinator as a SUB dependent and, if a verbal predicate, is fully inflected. The head of the relative clause, which will generally be the predicate (and thus have a PRED status), will be an ATR dependent of whatever noun it is modifying. In (39), the noun to pay attention to is *fâat*, which is underlined, and the relative clause is *pakuntâarahiti*, bolded.

(39) xás âapun vúra uum p-oo-tâayhi-ti
and on.the.ground INT 3SG.PRO SUB-3SG-be.many-DUR
pa-mukun-'ásip káru vúra fâat vúra **pa-kun-tâarahi-ti**
the-3PL.POSS-cooking.basket and INT something INT SUB-3PL>3-have-DUR
pa-kun-imnísh-eesh
SUB-3PL-cook-PROSP
'And on the floor were their cooking baskets and whatever else they had when they were going to cook.' (Julia Starritt, WB_KL-77:15)

A subtree showing only the analysis of the relative clause is below:



As with other modifiers of nouns, relative clauses can be used on their own as arguments without an expressed lexical noun. See section 2.9.3 for how to annotate noun phrases without lexical nouns.

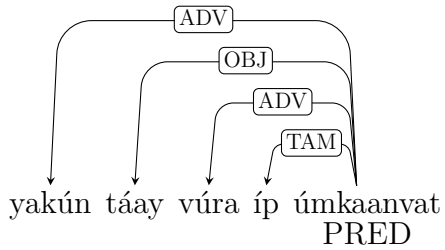
2.7.8 TAM

The TAM relation, standing for TENSE, ASPECT, MODALITY, is used for a variety of grammatical particles, including tense particles, aspect particles, modal particles, polarity particles, and some others. The head of TAM dependents is the predicate of the clause which the particle is in. In general, these particles are preverbal and often matched by morphology on the verb expressing similar meaning. The details of the syntax of each particle may differ, however. What follows is a discussion of each of the particles analyzed as TAM dependents.

Past tense particles *ipa* and *mit*

The particles *ípa* 'recent past' and *mit* 'past' are preverbal particles that are often used in conjunction with the past suffix *-at* on the verb (though they may also appear with non-past marked verbs). These particles are given the TAM relation and the predicate whose tense they express is treated as the head.

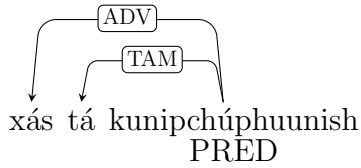
(40) yakún táay vúra íp ú-mkaanv-at
you.see a.lot INT REC.PAST 3SG>3-gather-PAST
'You see, she gathered a lot.' (Julia Starritt, WB_KL-32:19)



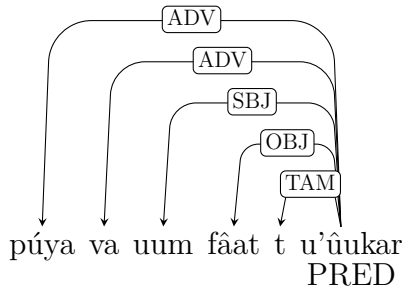
Perfect particle *tá/t-*

The perfect particle *tá* is an extremely common aspect particle, which appears preverbally as a separate word, *tá*, as in (41), but also in a prefixal *t-* allomorph which attaches to a verb directly, as in (42), or to a separate word consisting of *tá* plus certain agreement prefixes.

(41) xás tá kun-ip-chúphuunish
and PER 3PL>3-ITER-talk.to
'Then they spoke to him again.' (Julia Starritt, WB_KL-0:9)



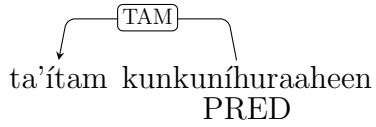
(42) púya-va uum fâat t-u-'ûukar
so-thus 3SG.PRO something PER-3SG>3-pay
'Then he paid something.' (Julia Starritt, WB_KL-0:8)



Anterior particle *ta'itam*

The particle *ta'ítam* is a TAM dependent of a predicate if the predicate is marked by the anterior tense suffix *-aheen*. Generally, the two, particle and suffix, co-occur, though it is not necessary. (43) is an example of this typical pattern of co-occurence, where *ta'ítam* is a TAM dependent of the predicate.

(43) ta'ítam kun-kuníh-uraa-heen
ANT 3PL>3-shoot-up-ANT
'So they shot it up.' (Mamie Offield, WB_KL-33:67)

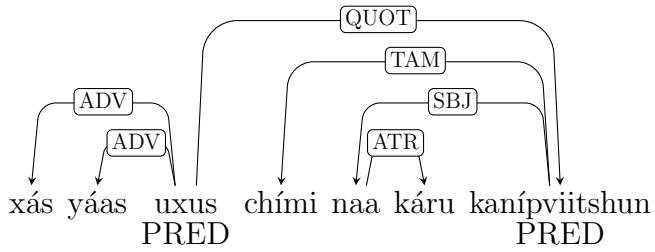


In uses where the predicate does not have the anterior suffix, *ta'ítam* is acting as a discourse particle and is made an ADV dependent of the verb.

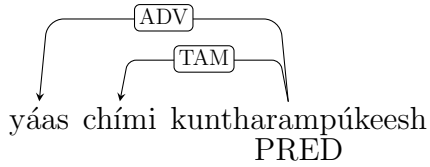
Imperative/prospective particle *chími*

The particle *chími*, glossed as 'soon,' is typically used (without obviously contributing the meaning 'soon') with imperative (44) and prospective-marked predicates (45). In these uses, *chími* is a TAM dependent of the predicate.

(44) xás yáas u-xus "chími naa káru kan-íp-viitshun
and then 3SG-think soon 1SG.PRO also 1SG.IMP-ITER-paddle.away.IMP
'So then he thought "Let me paddle away again too!"' (Nettie Ruben, WB_KL-57:83)



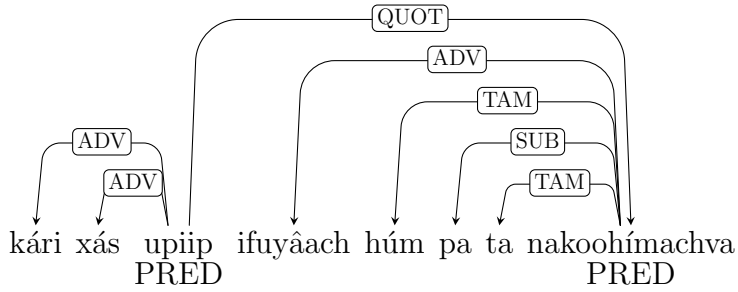
(45) yáas chími kun-tharampúk-eesh
then soon 3PL>3-cook.acorn.soup-PROSP
'Then they were about to make acorn soup.' (WB_KL-73:22)



Polar question particle *húm*

Polar questions are formed in Karuk by the addition of the particle *húm* to a clause. The particle is annotated as a TAM dependent of the predicate.

(46) kári xás u-piip "ifuyâach húm pa-ta na-koohímachva
and and 3SG-say true POL SUB-PER 3/2SG>1-grieve
'And he said, "Is it true that you grieve for me?' (Mamie Offield, WB_KL-58:10)

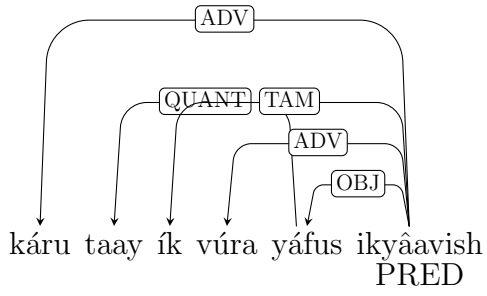


Occasionally, *húm* attaches as a clitic to the right edge of a preceding word, or occurs in the allomorph *úm*. These cases are analyzed the same.

Modal particle *ik*

The modal particle *ik* is likewise a TAM dependent of the predicate of its clause. Occasionally, this particle attaches as a clitic to the right edge of the preceding word, but in such cases it retains the TAM analysis.

(47) káru taay ík vúra yáfus i-kyâa-vish
and many MOD INT dress 2SG>3-make-PROSP
'And you must make many dresses.' (Mamie Offield, WB_KL-58:21)



The particle *ik* is often identifiable as a frozen clitic in several other particles. It is only annotated in the treebank when it is represented as a separate morpheme in the morphological annotation.

Others

A few other particles are routinely annotated as TAM dependents. They are listed here:

- *xâat* 'may'

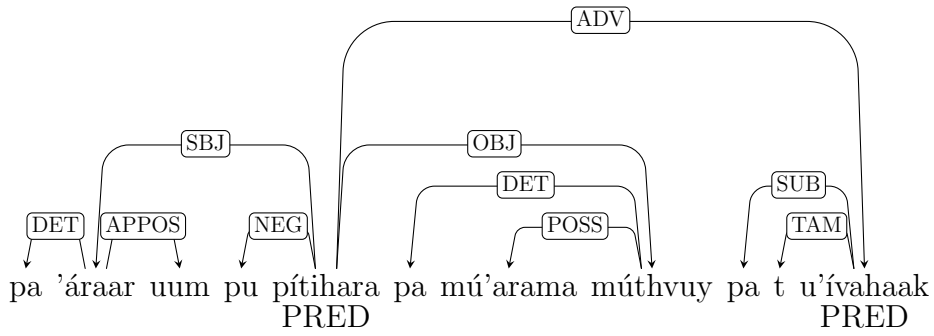
- *xâatik* 'it's better that...'
- *tîi* 'let' (expressing wish)
- *kíri* 'let' (expressing wish)
- *xayfaat* 'don't!' (expressing prohibition)

2.7.9 NEG

The NEG relation is given to the negative clitics *pu=* and *pura=*. The head of the relation is the predicate being negated. The negated predicates generally are marked also with the suffix *-ara*, meaning that identifying the head of the NEG relation is possible sans translation by finding the predicate marked with *-ara*. If the predicate is a verb, the agreement will also reflect a different paradigm used with negation. The negative suffix *-ara* does not occur in some morphological environments, however, and so it is not a guaranteed way to identify the negated predicate.

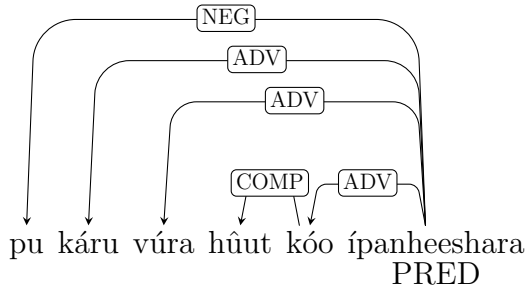
The typical pattern is for *pu-* to be attached to the left edge of the predicate it is negating, as seen in (48), where *pu-* sits on the left edge of *pitihara*, the negated predicate.

(48) pa-'áraar uum pu-pí-tih-hara pa-mú-'arama mú-thvuy
the-Indian.person 3SG.PRO NEG-say-DUR-NEG the-3SG.POSS-child 3SG.POSS-name
pa-t-u-'ív-ahaak
SUB-PER-3SG-die-IRR
'The Indian did not say his child's name when it died.' (Julia Starritt, WB_KL-0:1)



However, as a clitic, *pu-* can also attach to the left edge of any word preceding the negated predicate in that predicate's clause. In (49), *pu-* attaches to the left edge of the discourse particle *káru*, at the left edge of the entire clause. Note that, even with *pu-* being quite a distance from the negated predicate, the predicate is still marked with the negative suffix *-ara*: *ípanheesh**ara***. In these cases as in the simpler ones, *pu-* is a NEG dependent of the negated predicate, in this case *ípanheeshara*.

(49) pu-káru vúra hûut kóo ípanh-eesh-ara
NEG-also INT how as.much.as extend-PROSP-NEG
'It won't reach that far either' (Mamie Offield, WB_KL-33:62)



2.7.10 COMP

The COMP relation, standing for COMPLEMENT, is given to subcategorized dependents of verbs and postpositions that do not fall into any of the three main argument categories described above (SBJ, OBJ, or IO). The word which subcategorizes for the complement is the head of the relation. The COMP relation is specifically used for complement clauses of verbs, complements of verbs added by an applicative suffix, and noun phrase 'objects' of postpositions, as well as some more individual cases which are discussed in the subcategorization frames section of specific constructions, 2.9.11. Note that, while subcategorized for by verbs of speaking, direct quotation clauses are not marked as COMP dependents, but with the special syntactic relation QUOT described in the next section.

Complement clauses to verbs

Davis et al. (2020, p.845-849) describe Karuk complement clauses in depth, but some basic facts are presented here. Complement clauses in Karuk look essentially like main clauses, though generally somewhere in the clause to the left of the predicate will be the subordinator clitic *pa-*, though this is not necessary and there are some cases of complement clauses which lack *pa-* marking. Verbal predicates in complement clauses inflect exactly like main clauses, with the same agreement prefixes.

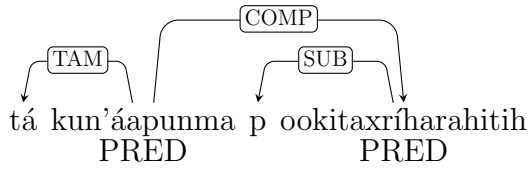
As with main clauses, the head of the complement clause will be its predicate. This predicate will bear the COMP relation to whatever verb subcategorized for it. As Davis et al. (2020) describe, complement clauses always follow the verb they are the complement of. A list of verbs which can take clauses as complements in this way is given below, reproduced from Davis et al. (2020, p. 846) :

- aachíchha 'to be glad'
- áapunma 'to know'
- ímus 'to look at'
- ikrûunti 'to wait for'

- ikyâavarihva ‘to try’
- ipêer ‘to tell’
- ipshinvárihva ‘to forget’
- ítap ‘to know’
- káriha ‘to be ready’
- kôoha ‘to stop’
- kúupha ‘to do’
- mah ‘to see, to find’
- pasúpiichva ‘to reveal’
- piip ‘to say’
- pikrôok ‘to remember’
- pikyaar ‘to finish’
- tápkuup ‘to like’
- thitiv ‘to hear’
- ûurih ‘to be unwilling’
- víiha ‘to dislike’

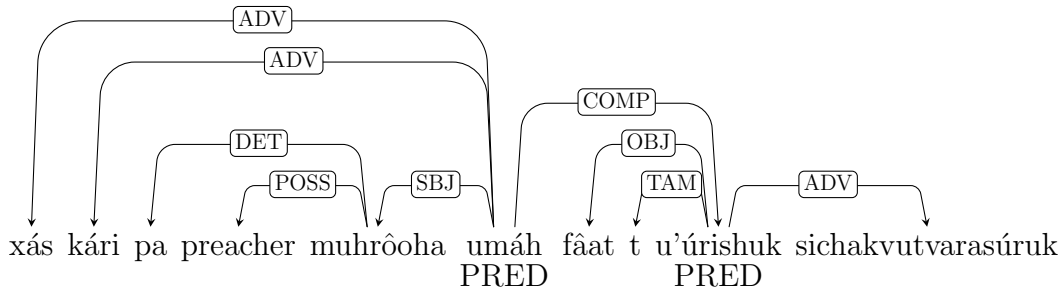
(50) is an example of the annotation of a complement clause, with the predicate of the complement clause, *ookitaxríharahitih* as a COMP dependent to the main clause predicate *kun’áapunma* which introduces it.

(50) tá kun-’áapunma p-oo-kitaxríharahi-tih
PER 3PL-know SUB-3SG-be.unfaithful-DUR
‘They found out that he was being unfaithful.’ (Nettie Ruben, WB_KL-39:24)



(51) shows an example of a complement clause which lacks the subordinator *pa=*. The complement clause, headed by *u’úrishuk*, is a dependent of the verb *mah* ‘to see,’ but there is no subordinator *pa=* in the sentence.

(51) xás kári pa=preacher mu-hrôooha u-máh “fâat t=u-’úrishuk
and then the=preacher 3SG.POSS-wife 3SG-see what PER-3SG>3-take.out
sichakvutvara-súruk”
belt-under
‘And then the preacher’s wife saw her take something out of the belt.’ (Julia Starritt, WB_KL-67:7)



Note that Bright enclosed the complement clause in quotation marks; this shows that Bright understood it was a complement, but interpreted the lack of *pa=* as meaning it was a quote. However, there is no direct evidence for it being a quote, as there is no shifting of person from third to first or second. It is also unexpected for *mah* to take a QUOT dependent, since QUOT dependents should be representations of actual speech, not of something seen.

It is possible that there are two independent clauses here, though one might expect in that case that the subject of *u'úrishuk* would be interpreted as the preacher's wife, rather than another character as here, since series of verbs with the same agreement and no introduction of new noun phrases as possible subjects typically share the same subject. There is likewise no direct evidence these are two separate clauses; such an interpretation would have been suggested by the presence of a comma separating the two clauses, but no comma is found.

Further, it is a more interesting result if the sentence above does in fact have a complement clause which lacks *pa=*. Such a sentence should be made to be found easily using the treebank, so that further research on this possibility can be expedited. Annotating the sentence with a complement clause achieves this, since it can be found with a search of complements which are verbs and lack a *pa=* dependent.

Applicative complements

Karuk has a large amount of directional applicative suffixes which add an extra, generally locational, argument to the subcategorization of a verb. When one of these suffixes occurs on a verb in a clause where the locational argument added by the applicative is also expressed, the locational argument will be a COMP dependent of the verb (as opposed to a ADV dependent, which it would be if not subcategorized for by the verb+applicative.) A list of these applicative suffixes, including some which are not directional and adapted from Macaulay (2004), is given below:[11] Note that directional adverbs, including *káruk* 'upriver,' *yúruk* 'downriver,' *sáruk* 'downhill,' *máruk* 'uphill,' *ithyáruk* 'across-river' and their derivatives often co-occur with directional affixes, but are themselves not treated as complements. They instead will be given the ADV relation.

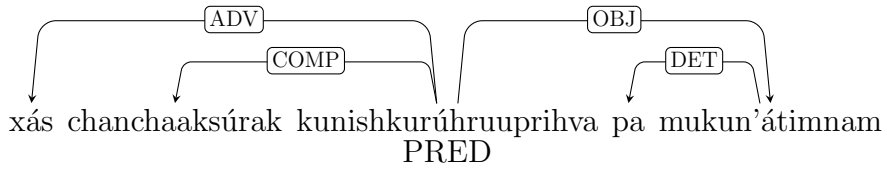
[11]The 0 in several of the suffixes in this list represents a copy vowel: it will have the quality of whatever vowel is in the preceding syllable.

- *-ara* 'with, INSTRUMENTAL'
- *-ihi* 'for, BENEFACTIVE'
- *-kiri* 'in, on, by means of'
- *-koo* 'to'
- *-path/-iroopith* 'around'
- *-rav/-ram* 'in, into'
- *-rip* 'off, out'
- *-ruprin* 'through'
- *-sar/-san* 'along with'
- *-suru* 'off'
- *-unish* 'to, at, about'
- *-0vraa* 'over'
- *-ku* 'on a vertical surface'
- *-ramnih* 'in a container'
- *-taku* 'on a horizontal surface'
- *-furuk* 'into an enclosed space'
- *-kurih* 'into cavity, into water'
- *-eep* 'away from a person'
- *-rishuk* 'out of a container'
- *-rupuk* 'out of an enclosed space'
- *-0thuna* 'here and there in an open area'
- *-varayva* 'here and there within an enclosed space'
- *-ruprih* 'in through a solid'
- *-vara* 'in through a tubular space'
- *-kiv* 'out through a tubular space'
- *-ruprav* ' out through a solid'
- *-0vruk* 'down over the edge of something'
- *-kara* 'toward center of a body of water, into one's mouth'
- *-kirih* 'into fire'
- *-tunva* 'towards each other'
- *-0vrath* 'into a sweathouse'
- *-ripaa* 'away from center of body of water'
- *-rupaa* 'out of one's mouth'
- *-0vrin* 'in opposite directions'
- *-mu* 'to'
- *-raa* 'to here'
- *-uk* 'to here'
- *-faku* 'to here from uphill'
- *-raa* 'to here from downhill or downriver'
- *-rina* 'to here from across a body of water'
- *-varak* 'to here from upriver'
- *-kath* 'from here across a body of water'
- *-rupu* 'from here downriverward'
- *-unih* 'from here downhillward'
- *-roovu* 'from here upriverward'
- *-uraa* 'from here uphillward'

(52) is an example of an applicative complement. The verb, *kunishkurúh**ruuprih**va* includes the suffix *-ruuprih*, 'through,' and thus introduces into the subcategorization of

the verb space for an argument (the thing which is moved through). What is being moved through in this case is *chanchaaksúrak*, the roof-hatch of a house, and this word, introduced by the applicative suffix on the verb, is thus the COMP dependent of the verb. Note that complements of applicatives, though sometimes marked with *-ak* LOCATIVE, as in (52), are not always marked as such.

(52) xás chanchaaksúr-ak kun-ishkurúh-ruuprih-va
and roof.hatch-LOC 3PL>3-carry.by.handle-through-PL.ACT
pa-mukun-'átimnam
the-3PL.POSS-burden.basket
'Then they pulled their burden baskets up through the smokehole.' (Nettie Ruben, WB_KL-57:23)



Occasionally, a postpositional phrase will be the complement to an applicative, in which case the postposition has its typical annotation (see example (53) in Complements to postpositions below), but the postposition itself, as head of the postpositional phrase, will be the COMP dependent of the verb.

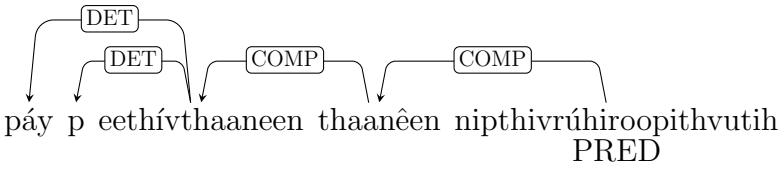
Complements to postpositions

Most postpositions in Karuk are, exceptionally, some of the only words whose subcategorized dependents must be expressed (though this does not hold for all of the postpositions). The 'objects' of these postpositions, which are generally nominals and closely precede the postposition, are COMP dependents of the postposition. COMP, rather than OBJ, is used to keep the OBJ syntactic relation dedicated only to verbal direct objects. A list of postpositions in the language follows:

- *îin* OBVIATIVE
- *koo* 'as much as'
- *xákaan* 'together with (one other person)'
- *kóovan* 'together with (2 or more people)'
- *kumá'ii* 'because of'
- *kuth* 'because of'
- *kuuk* 'to'
- *mûuk* 'with, by means of'
- *thaanêen* 'around'
- *váa* 'for'
- *véenik* 'ago'
- *vêetak* 'ago'

In (53), the postposition *thaanêen* 'around' has *eethívthaaneen* 'world' which immediately precedes it as its COMP dependent. The postposition is itself a COMP dependent of the verb because it is the argument of the applicative suffix *-iroopith* on the verb - see the preceding subsection for details.

(53) páy p-eethívthaaneen thaanêen ni-p-thivrúh-iroopith-vu-tih
this the-world around 1SG-ITER-float-around-PL.ACT-DUR
'I float around and around this world.' (Chester Pepper, WB_KL-55:52)



2.7.11 QUOT

The QUOT relation is used for direct quotes of speaking or thinking verbs, with the speaking or thinking verb as the head of the relation and the head(s) of the quote as the bearer of the relation. Direct quotes are characterized separately from other clausal complements like described in the COMP section by their lack of the subordinator *pa-*, and by the usage of local-person (1st and 2nd) pronouns or verbal agreement in them that reflect the perspective of individuals otherwise referred to with third-person reference in the narratives which form the corpus. Because not every direct quote necessarily involves local-person reference, the lack of *pa-* is a more reliable indicator. Also, in the texts in Bright (1957) which form the core of the corpus, direct quotes are generally indicated by "..." (double quotation marks) in the text (though Bright's indication may be incorrect, if a subordinator is present.) A further difference from non-direct quote complement clauses is that direct quotes do not necessarily have to follow the verbs they are complements of.

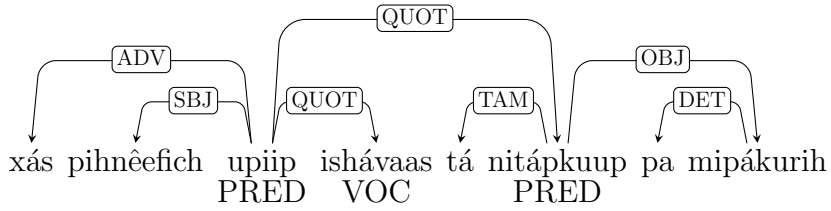
Some verbs of speaking and thinking which can take QUOT complements are listed below:

- piip 'to say'
- ipêer 'to tell'
- xus 'to think'
- xusêer 'to think X about Y'

At least two of these, *piip* and *ipeer*, can also take indirect quotes as COMP dependents, so care must be taken to correctly identify the type of complement these verbs have in each specific case.

In (54), the verb *upiip* takes the predicate of the quoted clause, *nitápkuup*, as a QUOT dependent.

(54) xás pihnêefich u-piip, "ishávaas, tá ni-tápkuup pa-mi-pákurih.
and Coyote 3SG-say nephew PER 1SG>3-like the-2SG.POSS-song
And Coyote said, "Nephew, I like your song!" (Nettie Ruben, WB_KL-07:28)



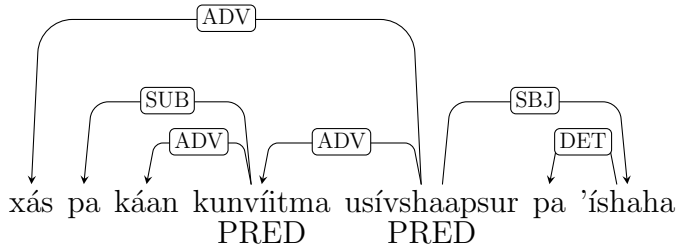
In (54), an exceptional property of the QUOT relation is also exemplified. The verb *upiip* has two QUOT dependents - the vocative *ishávaas*, and the quoted clause discussed above. Direct quotes do not necessarily include only one syntactic sentence - quotes may indeed include whole stretches of discourse made up of independent syntactic units. To preserve the independence of these quoted units, they are all made QUOT dependents of the relevant verb. This is analogous to the fact that these independent units, if not part of a direct quote, would have no syntactic relations connecting them. Note that there are many cases where the actual content of a quote extends over multiple database sentences in the Karuk corpus, and the only elements given the QUOT relation are those parts of quotes which occur in the same database sentence as the verb of speaking/thinking that introduces them.

2.7.12 ADV

The ADV relation, standing for ADVERBIAL, is given to any word which serves to modify a predicate and is not covered by any of the other syntactic relations. The predicate being modified is the head of the relation. This is analogous to the use of the ATR relation in the nominal domain. Common bearers of the ADV relation include discourse particles (like *kári* and *xás*, both meaning something like 'and, then'), adverbial clauses, and actual adverbs (in terms of part of speech), but ADV is essentially a catch-all category, and as such has very wide usage, with nearly every lexical part of speech category being able to be used adverbially. If a particular word in a sentence does not appear to conform to any of the patterns discussed in these guidelines, a safe annotation is to make the word an ADV dependent of the highest possible predicate (generally the main clause predicate, but can be a subordinate predicate if the word in question is, for example, between the subordinator *pa-* and a subordinate predicate).

(55) showcases several cases of ADV dependency. *xás*, a discourse particle, is an ADV dependent of the main clause predicate *usívshaapsur*, as is the adverbial clause *pákaan kunvíitma*, whose head *kunvíitma* is the bearer of the ADV relation. Within the adverbial clause, the adverb *káan* 'there' is an ADV dependent of the adverbial clause predicate, *kunvíitma*.

(55) xás pa-káan kun-víit-ma u-sívshaapsur pa'-íshaha
and SUB-there 3PL-paddle-to 3SG-become.unplugged the-water
'And when they paddled to there, the water opened ' (Nettie Ruben, WB_KL-57:102)



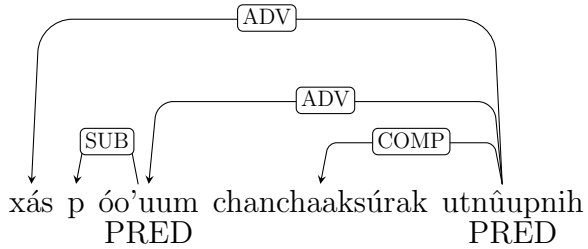
It should be noted that adverbial clauses exhibit essentially the same internal structure as complement clauses and relative clauses, in that they involve subordination using *pa-* and having full inflection marked on any verbal predicates heading them. Unlike complement clauses, adverbial clauses may occur before or after the predicate they modify (Davis et al. 2020).

2.7.13 SUB

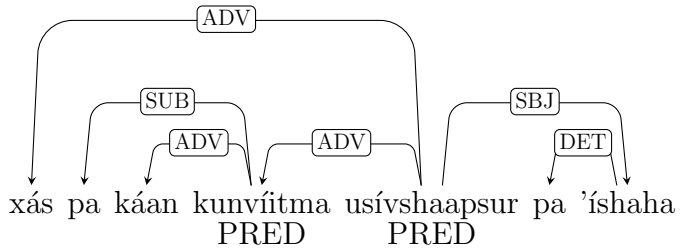
The SUB relation is given only to the subordinator clitic *pa-*, with the head of the relation being the predicate which *pa-* is subordinating. The nature of the subordination (i.e. whether the subordinate clause is a complement, acting adverbially, a relative clause, etc.) is expressed in the relation that the predicate of the subordinate clause has with its own head. There can only be one SUB relation, and only one subordinator *pa-*, per subordinated predicate.

As with the other clitics in Karuk, subordinator *pa-* can either attach to the left edge of the predicate it subordinates (56), or on the left edge of any word in the same clause which precedes the predicate (57). If multiple clitics appear on a word, subordinator *pa-* is always the outmost.

(56) xás p-óo-'uum chanchaaksúr-ak u-t-nûupnih
and SUB-3SG-arrive roof.hatch-LOC 3SG-look-through.
'And when he arrived, he looked in through the smokehole.' (Nettie Ruben, WB_KL-57:20)



(57) xás pa-káan kun-víit-ma u-sívshaapsur pa'-íshaha
and SUB-there 3PL-paddle-to 3SG-become.unplugged the-water
'And when they paddled to there, the water opened ' (Nettie Ruben, WB_KL-57:102)



Distinguishing the subordinator *pa-* from the determiner *pa-*

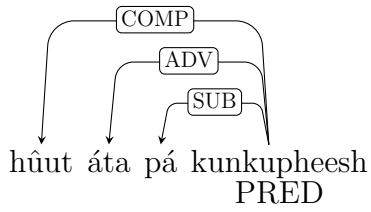
The subordinator and definite determiner have the same phonological form, *pa-*, which can present difficulties in identifying what role a particular *pa-* is playing in a sentence. Complicating matters further, there is never a case where the two *pa-*'s stack on a word, even if it would be expected. Bright (1957, p. 122) analyzes this as a case of haplology; for the purposes of annotation, this essentially means that, if one is unsure which *pa-* is being used in a sentence, there is also the possibility that both are present, with only one pronounced.

Following the principle ATTACH HIGH, if a given *pa-* could be analyzed as either the definite determiner or the subordinator, it should be analyzed as the subordinator, since the subordinator will necessarily be a dependent of the predicate, and thus head, of the clause, whereas the definite determiner will likely only be a dependent of one of the arguments of that predicate, and thus dominated by the predicate.

Main clause usage

Occasionally, predicates which must be the main clause of an sentence (most readily identified when they are the only predicate in a sentence) appear marked with the subordinator *pa-*. This is especially common in constituent questions. In these cases too, *pa-* is a SUB dependent of the predicate. The fact that the predicate is not, in fact, subordinated to anything will come from the fact that it has no head or syntactic relation. In (58), the only predicate is the verb *kunkupheesh*, but it is marked by the subordinator *pa-*, despite there being no head to which the predicate could be a dependent.

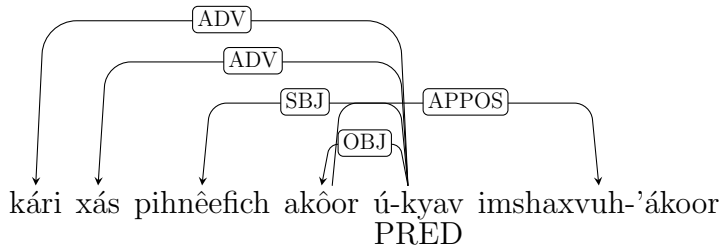
(58) hûut áta pá-kun-kuph-eesh
how maybe *sub*-3PL-do-PROSP
'I wonder what they're going to do?' (Nettie Ruben, WB_KL-57:31)



2.7.14 APPOS

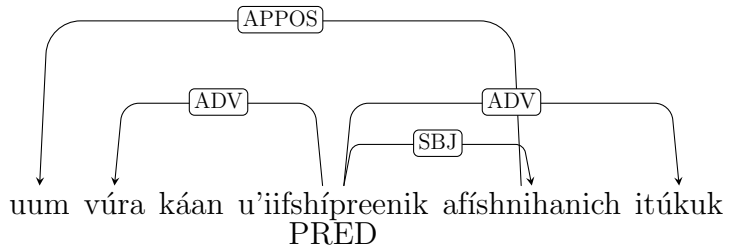
The APPOS relation, standing for APPOSITION, is used as a relation between two expressed noun phrases which refer to the same referent (or referents in a subset-superset relation). If the apposed noun phrases are both non-pronominal, the first in linear order is the head of the relation, and the bearer of the APPOS relation is the latter noun phrase.

(59) kári xás pihnêefich akôor ú-kyav, imshaxvuh-'ákoor
and and Coyote axe 3SG>3-make pine.gum-axe
'Then Coyote made an axe, a pine-gum axe.' (Mrs. Bennett, ALK_14-35:21)



If the apposed phrases are a pronoun and a non-pronominal noun phrase, the non-pronominal phrase is the head, and the pronoun the bearer of the APPOS relation. This "pronominal" apposition is quite common. See 2.9.5 for more information about how the third person *uum* is annotated specifically.

(60) uum vúra káan u-'iif-shípree-nik afíshnihan-ich, itúkuk.
3SG.PRO INT there 3SG-grow-up-ANC young.man-DIM Itúkuk
'A young man lived there at Itúkuk.' (Nettie Ruben, WB_KL-57:2)



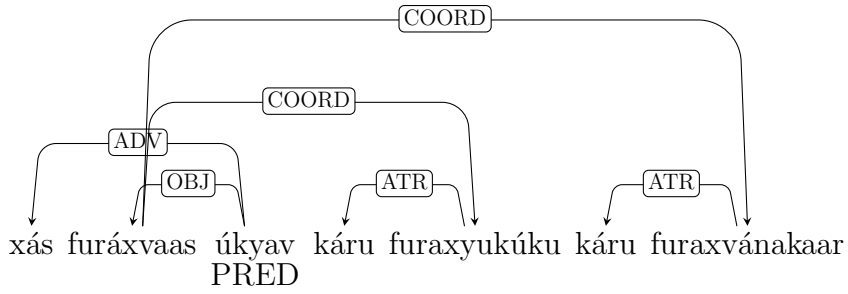
The pronoun *vaa* 'that way, thus' is often appositional to an entire clause. For details on that construction, please refer to section 2.9.6.

The APPOS relation is used in similar contexts as the COORD relation, so there may occasionally be difficulty in distinguishing the two. See section 2.7.15 below for guidelines on differentiating the two.

2.7.15 COORD

The COORD relation, standing for COORDINATION, holds between two phrases which are coordinated by the particle *káru*. The first of the phrases is the head of the relation, and succeeding phrases are COORD dependents of it. Following the principle ATTACH HIGH, if there are more than two coordinated phrases, the first in the linear order will be the head of all of the other coordinated phrases.

(61) xás furáx-vaas ú-kyav, káru furax-yukúku káru furax-vánakaar
and woodpecker.scalp-blanket 3SG>3-make and woodpecker.scape-shoe and woodpecker.scalp-type.of.shirt
'So he made a woodpecker-head blanket, and woodpecker-head shoes and a woodpecker-head vánakaar (a shirtlike garment).' (Nettie Ruben, WB_KL-57:130)



As seen in (61), the coordinator *káru* bears an ATR relation to the coordinated phrase to its right, and does not bear the COORD relation, which holds only between the actual phrases which are coordinated. If predicates or clauses are being coordinated, *káru* will bear the ADV relation to the latter of the coordinated predicates, rather than ATR.

Note that the COORD relation is only used for clausal coordination when there is an overt coordinator particle *káru* between the two coordinated elements. For cases where a sentence is translated as having coordination but there is no *káru*, refer to 2.9.2.

Differentiating COORD from APPOS

The COORD and APPOS relations apply to similar scenarios, and care must be taken not to confuse the two. The APPOS relation is used only for noun phrases which refer to the same referent or two referents that are in a subset-superset relationship, whereas coordination of noun phrases will involve nouns which have distinct referents.

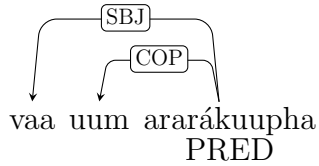
2.7.16 COP

The COP relation, standing for COPULA, is given to the third person pronoun *uum* in a particular configuration where it appears to act like a copula mediating between a subject and a non-verbal predicate. The non-verbal predicate is the head of the relation.

It is a common use of *uum* to occur between a non-verbal predicate and the subject of that non-verbal predicate, like as to where a copula might be found. Given that it is an attested grammaticalization pathway for third person pronouns to develop into copulas (Heine and Kuteva 2002, p. 235), it is possible Karuk is undergoing this grammaticalization. In any case, to study these particular uses of *uum*, they are given a unique relation.

(62) shows an example of how this annotation works. *uum* appears between *vaa*, 'that,' the subject, and *ararákuupha*, 'Indian law', the predicate, and *uum* is made a COP dependent of the predicate. Note that the predicate and subject can be in the opposite order; as long as *uum* is between them (not necessarily adjacent to both), it will be a COP dependent of the predicate.

(62) vaa uum arará-kuupha
that 3SG.PRO Indian.person-doing
'That was Indian law' (Julia Starritt, WB_KL-0:14)

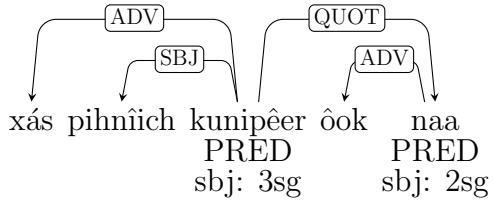


More information on the annotation of *uum* in different contexts is given in section 2.9.5.

2.8 Person/number annotation

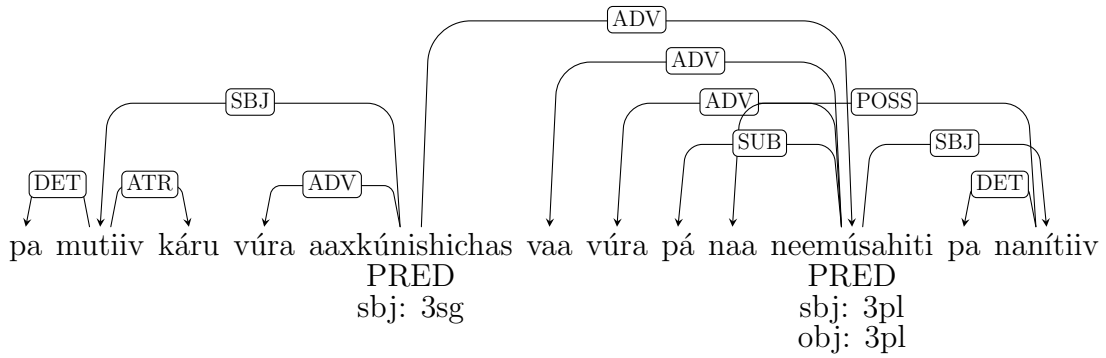
In addition to syntactic statuses and relations, predicates in the Karuk treebank are annotated for the *semantic* person/number of their arguments (subjects, direct objects, and indirect objects). This is different from the morphological expression of person/number in the agreement prefix system. The person/number represented by each of the agreement prefixes is already accounted for by the morphological annotation of the corpus, but occasionally, what we can surmise about the person/number of the arguments on the basis of the translation of a sentence or other facts about the morphology in the sentence or the particulars of the narrative does not match the person/number of the agreement prefixes. For example, in (63), the prefix on the verb *kunipêer* is *kun-*, indicating a 3pl subject, but the translation and the expressed noun phrase of the subject are both singular: 'old man.' We also know in this case from the story of the text this sentence is from that there is indeed only one old man. Thus, in this case, the agreement prefix, at least from our current understanding, appears to not match what the *semantic* number of the subject actually is.

(63) xás pihnîich kun-ipêer "ôok naa"
and old.man 3PL-tell here come
'And the old man said, "Come here!"' (Daisy Jones, WB_KL-20:42)



This is true also of person agreement, though to a lesser extent. For example, in (64), the second verb *neemúsahiti* has the prefix *nee-*, an allomorph of *na-* indicating, typically, a 1sg object and occasionally a 1sg subject of certain stative verbs. In any case, the important thing here is that it indicates 1sg. However, as the translation makes clear, the actual argument here is not 1sg, but 'my ears,' which, though possessed by 1sg, is 3pl. In such a case, the prefix does not strictly match the *semantic* person.

(64) pa-mu-tiiv káru vúra aaxkúnish-ich-as vaa vúra pá-naa
the-3SG.POSS-ear also ITNS red-DIM-PL that INTS SUB-1SG.PRO
nee-músahi-ti pa-naní-tiiv
1S-look-DUR the-1SG.POSS-ear
'His ears are reddish too, just like my ears look.', (WB_KL-16:15)



Because of these mismatches, the treebank provides annotations for each predicate, whether verbal or nonverbal, about what the *semantic* person/number of the arguments of that predicate are. This section will detail how this annotation was completed. However, it is important to make one caveat here at the beginning. Like with all aspects of the treebank, the goal is primarily to create a tool which allows for finding interesting sentences, not representing the actual syntactic structure completely faithfully. In terms of person/number annotation, this primarily means finding any sentences that potentially have a person/number mismatch between meaning and morphology, and annotating them as such even if it is also possible that there is no mismatch. As such, and because the person/number annotation is so dependent on the translation rather than more dependable semantic elicitation, person/number annotations should not be taken as necessarily reflecting what the Karuk semantics actually are, and more as a guide to finding sentences with interesting properties.

Before discussing the criteria by which this annotation was completed, I first describe the particular values of person/number that the annotation uses.

2.8.1 Person/number annotation tags

Table 2.1 presents the particular tags that can be assigned to a predicate for person/number annotation. The first two rows are the typical combinations of 1st, 2nd, and 3rd person combined with singular and plural. I will not discuss these further, but the tags 3MASS and EXPL require some more discussion.

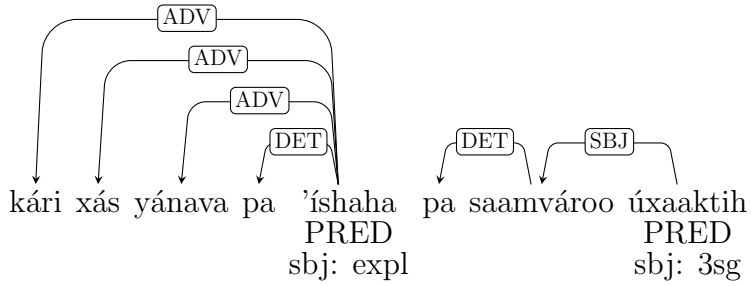
1sg	2sg	3sg	
1pl	2pl	3pl	
		3mass	
			expl

Table 2.1: Tags for person/number annotation

3MASS is used for annotating entities which could be said to be 'canonically' mass as opposed to count, including many entities that are clearly categorizable as substances: water, sand, acorn soup, among others. In some cases, a noun, like fire, may be treated as mass in some instances and as count in some other instances, dependent on the context. More detail about the annotation of 3MASS in specific sentences can be found in the Notes on Annotation section. It should be noted that there is not clear morphological evidence that Karuk has a mass/count distinction[12] Plural marking is not available for most nouns since it is restricted to animates, and so plural marking cannot be used as a diagnostic. Numerals are pretty rare in the corpus and not all nouns have been tested to see if they combine with numerals, so co-occurrence with numerals is also not a diagnostic which can be used at this time. However, to assume from this that there can then be no potential effect of countability on the agreement system would be folly. The treebank should be a tool which allows us to assess whether a potential mass/count distinction affects the agreement system, and so are treated in this annotation as 3MASS so they can be compared to the more typical 3SG and 3PL cases.

EXPL is used for expletive sentences and existentials, where there is no actual semantic argument but, if the predicate is verbal, there is still agreement.[13] In (65), we see a non-verbal existential predicate. The sentence is translated as 'He saw the water' but the 'saw' in the English translation does not actually correlate with a Karuk predicate; it instead is a translation of the evidential *yánava*, which is not a predicate. The predicate is actually *íshaha*, 'water,' and the sentence could be translated more literally as 'visibly there was water.' Since the predicate here is merely asserting the existence of water, it does not really have a semantic subject per se, and so it is given the EXPL tag for subject person/number agreement.

(65) kári xás yánava pa='íshaha, pa=saamvároo ú-xaak-tih
and then visible the=water the=creek 3SG-make.sound-DUR
'He saw the water, the creek was sounding.' (WB_KL-01:27, Nettie Ruben)



[12]Note however that not having a morphological count/mass distinction does not necessarily mean the language lacks a count/mass distinction. Deal (2017) argues from data in Nez Perce that all languages in fact have a semantic count/mass distinction, even in languages like Nez Perce where there is no major morphosyntactic evidence for mass/count.

[13]The EXPL value is still used for non-verbal existential predicates which don't have agreement.

2.8.2 Annotating person/number

Typically, the process for annotating the person/number of a predicate proceeded as follows. Generally, identifying the person information for an argument is simple and the main challenge is in identifying number information.

- Identify the predicate and its correlate in the translation
- Identify the valence of the predicate. If intransitive, only the subject person/number needs be annotated. If transitive, subject and direct object need to be, and if ditransitive, subject, direct object, and indirect object need to be. This annotation is done regardless of whether the argument is expressed or not.
- Identify the person/number of each argument and assign it that tag
  - To do so, first look for morphological cues in the Karuk sentence itself. Typically this means looking for plural morphology on any expressed arguments. If any of the arguments is pluralized, make sure it is pluralized in the tag.
  - Next, if any particular argument is an entity previously mentioned in the text, that may give a hint as to what the number annotation should be.
  - Finally, look at the English translation. If the person/number of the argument in the translation does not conflict with any cue from the Karuk morphology (other than the verbal agreement) or the understanding from the narrative, follow it and assign the arguments their person/number tags on the basis of the translation.

Person/number annotation in this way can become complicated, as the example in (65) illustrates. For such complicated examples, descriptions of the thought process behind the annotation can be found in the Notes on Annotation section.

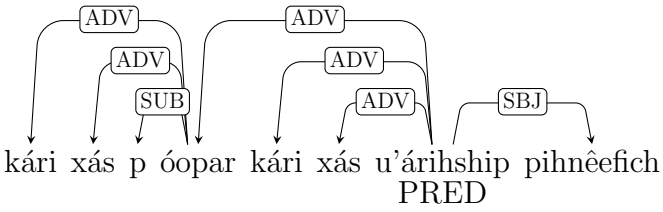
2.9 Annotation of specific constructions

2.9.1 Identifying whether discourse particles are dependents of subordinate or main clauses

Following the principle of Attach High (2.2.1), one should generally make sentence-initial adverbs or discourse particles dependents of the main predicate of the sentence, even when there is an intervening subordinate clause, if either predicate is in principle a possible head given the translation. However, there are systematic exceptions to this generalization, when a sentence-initial discourse particle should be made a dependent of an intervening subordinate clause instead of attaching higher. Namely, if the main or higher clause would have an identical discourse particle as a dependent, the sentence-initial discourse particle will be a dependent of the subordinate clause, to avoid a single predicate having two redundant dependents.

(66) exemplifies this case. The sentence begins with two discourse particles, *kári* and *xás*, and then has a subordinate adverbial clause *póopar*, followed by the main clause. Notice that, after the subordinate clause, there is another *kári* and *xás*. Strictly following ATTACH HIGH, we would make the main predicate of this sentence, *u'árihship* have two *kári* and two *xás* dependents. However, given that subordinate clauses can have dependents appearing to the left of the subordinator *pa=*, we could expect that the initial *kári* and *xás* are dependents of the subordinate predicate, *oopar*. To avoid making the main predicate have redundant identical dependents, we instead make the initial discourse particles dependents of the subordinate predicate, as shown below.

(66) kári xás p-óo-par, kári xás u-'árihship pihnêefich
and then NOMZ-3SG>3-bite and then 3SG-jump.up Coyote
'And when he bit it, then Coyote jumped up.' (Nettie Ruben, WB_KL-01:131)



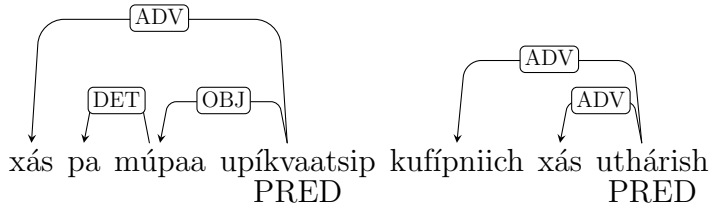
If a suitable other predicate is not present for duplicated discourse particles to be attached to, follow ATTACH HIGH.

2.9.2 Multiple clauses with no coordinator or a non-*káru* coordinator

In many cases, a single sentence in the corpus will contain multiple clauses with no coordinating particle or a discourse particle other than *káru* between the two clauses. These cases are not, however, analyzed as coordination, and do not involve a COORD relation between the predicates of the two clauses (unlike sentences with *káru* coordination which are analyzed as described in 2.7.15 above). Rather, in these cases, there will be no relation connecting the two clauses together, and each will be analyzed as if it was an independent sentence. Given that the sentence breaks in the sources of much of the corpus were determined on prosodic grounds and not on syntactic grounds, it is expected that the corpus sentences may include one or more independent clauses, and for this reason only in the presence of explicit *káru* coordination should the COORD relation be used. (67) exemplifies this sort of case.

(67) xás pa=mú-paa u-p-íkvaatsip kufípn-iich xás
then the-3SG.POSS-boat 3SG-ITER-pick.up.boat willow.grove-DIM then
u-thárish
3SG-put.down

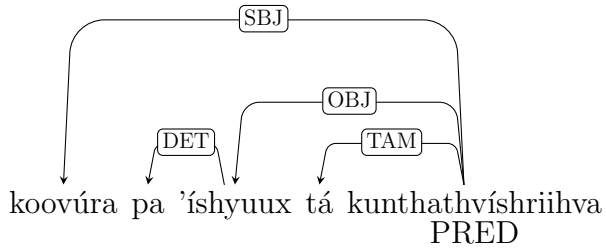
'And he picked up his boat, and put it down in a willow grove.' (Nettie Ruben, WB_KL-57:66)



2.9.3 Noun phrases without a lexical noun

Analogous to the fact that Karuk clauses can generally not express arguments, it is also possible even in expressed arguments for the lexical noun to be dropped or unexpressed, while leaving behind elements which would have been dependents of the noun if the noun were present. This includes words, at least, of the following types of nominal dependency relations - DET, QUANT, and ATR. In these cases, whatever is left behind in the noun phrase takes on the syntactic relation and head that the noun would have had, if it were present. Thus, in (68), the subject, as indicated by the English translation and the agreement on the verbal predicate, is third person plural, and is quantified over by 'all.' The 'all' is represented in the Karuk sentence, by *koovúra*, but there is no expressed pronoun for third person plural, leaving the *koovúra* as the sole syntactic word representing the subject noun phrase. *koovúra* would have been a QUANT dependent of the pronoun/lexical noun, if one were present, but since it is the only representative of the subject in the sentence, it is given the SBJ relation and made a dependent of the predicate, essentially taking on the relation that a lexical noun would have had if it were expressed as the subject.

(68) koovúra pa-'íshyuux tá kun-thathvíshriih-va
all the-elk PER 3PL>3-carry.home-PL.ACT
'They all carried the elk home.' (Nettie Ruben, WB_KL-62:11)



The most common cases are like (68)- with only one non-nominal word representing the relevant argument. It is conceivable that there could be cases there are two, however, (an adjective and determiner, for example) in which case only one of them should take on the

relation the lexical noun would have had, and the other should be a dependent of that one. In these cases, the more 'lexical' of the words should be treated as the head of the others, and take on the relation the lexical noun would have had.

2.9.4 Ellipsis and other missing verbs

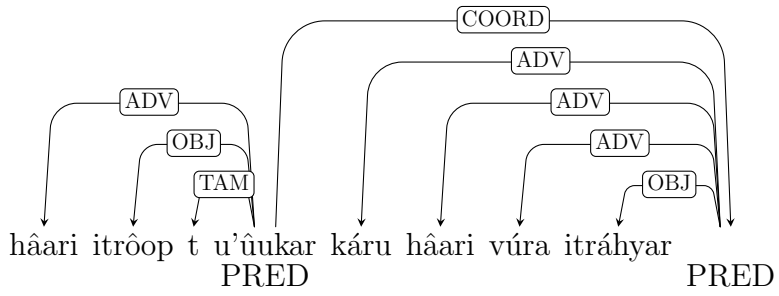
Occasionally, Karuk sentences will lack predicates where they are expected or necessary in order to complete the annotation. Predicates are in fact the most important word in any annotation following the system presented in these guidelines, since they usually form the root of any sentence and dominate all the other words. Predicates also are the only possible heads for various important relations, such as the argument relations SBJ, OBJ, and IO. Thus, in cases where a word, which would be analyzed as a dependent of a predicate, is in a sentence that lacks the appropriate predicate, the typical guidelines for annotation do not appear to easily apply.

In these situations, a 'null' verb is added into the sentence, which acts as the necessary predicate in all respects. This solution has the benefit of requiring no special dependency relations or statuses be added into the system to account for these cases of apparently 'missing' predicates. All the usual guidelines apply - the only difference is that the predicate simply is not expressed. There are no other 'null' elements in the system of syntactic annotation presented here, and the cases where null verbs can be invoked are particularly constrained. In fact, only in two situations is a null verb necessary: ellipsis and verb of speaking omission.

Ellipsis

In cases of ellipsis, a predicate which would be identical to another, earlier predicate can be left out, while leaving behind any of its dependents. This is the case in (69). It is clear that the intended meaning of the second clause of this sentence, which lacks a predicate, is "he paid ten dollars," with the meaning of the same predicate as in the first clause - the ellipsis in the English translation admits of no interpretation where the predicate is somehow a different meaning from the first. The second clause also is quite parallel to the first, including the adverb *hâari* and a numeral as the object. Adding a null verb, as shown in the tree below, allows this parallelism to be accurately captured, as the relevant aspects of structure are in fact the same - *hâari* is an ADV dependent of the predicate in both, and the numeral is an OBJ dependent of the predicate in both. All that is needed is the simple addition of the null verb in the annotation, represented as sentence-final in the tree below.

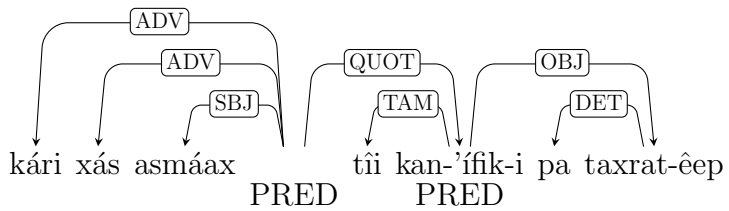
(69) hâari itrôop t-u-'ûukar, káru hâari vúra itráhyar
sometimes five PER-3SG>3-pay and sometimes INT ten
'Sometimes he paid five dollars, and sometimes ten.' (Julia Starritt, WB_KL-0:10)



Verb-of-speaking omission

Occasionally, an expected verb of speaking or thinking (in other words, a verb which can have a direct quote complement/QUOT dependent) will not be present, but the quote and other would-be dependents of the verb will be present. (70) is an example; there is no expressed verb-of-thinking (though the translation makes it clear that that is the interpretation), but there is a direct quote (which must be direct, given the use of first person agreement on the quote's predicate), a word which is identifiable as the speaker of the quote and which would have been the subject of a verb-of-thinking, *asmáax*, and two discourse particles which would clearly have been ADV dependents. The structure of this sentence is, in fact, remarkably similar to the structure of sentences with expressed verbs-of-speaking in what dependents appear. As such, to capture that parallelism, a null verb is added to represent the unexpressed or dropped verb-of-speaking/thinking, and the dependencies that would be expected of such a predicate are as they would be in a case where the predicate was expressed. Namely, *asmáax*, the thinker of the direct quote, is made the SBJ dependent of the null verb-of-thinking, and the predicate in the direct quote, *kan'ífiki*, is made the QUOT dependent of the null verb. Like in cases of ellipsis, the only difference in sentences with null verbs isn't syntactic, it is only in the fact that the verb is not pronounced.

(70) ... kári xás asmáax _ "tîi kan-'ífik-i
and then Towhee _ let 1SG.IMP-pick.up-IMP
pa-taxrat-êep
the-to.flake.arrowheads-refuse.from
'... then Towhee (thought) "Let me pick up the scraps."' (Nettie Ruben, WB_KL-36:4)



There is the potential for overzealous addition of null verbs, and as such only in cases like the above, where there is both a quote dependent and another dependent of the omitted verb-of-speaking, should null verbs be added. There are cases where there is a clear 'direct quote' with no verb-of-speaking and no other possible dependents of a null verb-of-speaking, but null verbs are not added in these cases. Narratives may shift from being from the narrator's perspective (using third-person agreement in reference to individuals in the narrative) to one of the character's perspectives (using first or second person reference) without syntactic signalling of this switch through means of subordination to a verb-of-speaking. The nature of direct quotes being, generally, full clauses on their own means that it is not unexpected that they do occasionally stand on their own. The sequence of two sentences in (71) shows an example of this, with (71-b) being a direct quote of Coyote (as evidenced by the first person agreement), who is the third person subject of the previous sentence (71-a), but there is no verb-of-speaking and no other potential dependents of one beyond the quote itself. In such a case, there would be no addition of a null verb to (71-b).

(71) a. xás íshaha t-óo xrah.
and water PER-3SG be.thirsty
'And he got thirsty'

b. "vúra pu-na-'ísh-eesh-ara"
INTENS NEG-1SG>3-drink-PROSP-NEG
' "I won't drink." ' (Nettie Ruben, WB_KL-01:15-16)

However, a non-quote dependent, like the subject *asmáax* in (71) above, not being a full clause, is not expected to generally occur on its own, without a head. Thus, only in those cases with both non-quote dependents and quote dependents, like (71), should a null verb be used; otherwise, the non-quote dependent would not have a suitable head.

2.9.5 Analyzing the third-person pronoun *uum*

The third person pronoun *uum* is a common word in sentences, and often the role it plays is obscure and not reflected in the translation. However, three contexts of use can be identified, and are each annotated differently.

Copular use

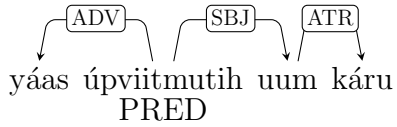
As described in the COP section above, *uum* occurs regularly in sentences with non-verbal predicates, where it is between the subject of the predicate and the predicate itself. In these situations, *uum* is annotated as a COP dependent of the predicate. See the COP section for examples.

Argumental use

If *uum* appears in a sentence with a verbal predicate with at least one of its arguments unexpressed, or a non-verbal predicate with no expressed subject, *uum* can be annotated as being the missing argument, if this is consistent with the interpretation of the sentence (i.e. as long as the missing argument is not 1st or 2nd person based on the translation). This, of course, is the expected case for nominals.

In (72), the verbal predicate of the sentence is intransitive and will have only a subject (though it can also take a locative complement, given the *-mu* directional applicative on it.) However, there are no expressed noun phrases other than *uum*. As such, *uum* is made the SBJ dependent of the predicate.

(72) yáas ú-p-viit-mu-tih, uum káru
then 3SG-ITER-paddle-to-DUR 3SG.PRO also
'Then he paddled back there too.' (Nettie Ruben, WB_KL-57:107)



In cases that involve transitive verbs, it may be impossible to tell which argument *uum* is representing, if multiple arguments are unexpressed. In that case, the following hierarchy should be followed. *uum* should be annotated as the leftmost type of dependent in the hierarchy that it can. (As such, if there is a ditransitive verb with the subject expressed, and one has a choice between making *uum* a OBJ or IO dependent, OBJ should be chosen, since it is to the left of IO in this hierarchy.)

(73) SBJ > OBJ > IO

Appositional use

If *uum* appears with a predicate that already has all of its arguments expressed by other phrases, *uum* is analyzed as appositional to one of them. The choice of which one should follow the hierarchy given in (73) above. See the APPOS section for examples.

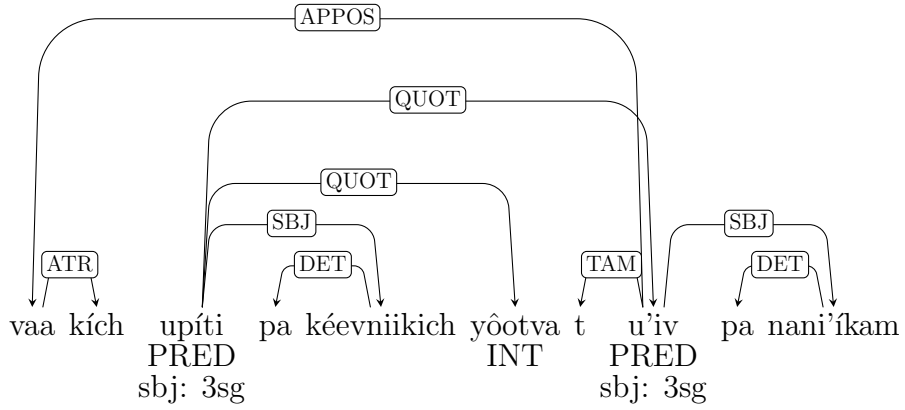
2.9.6 *vaa* used in apposition to a clause

The pronoun *vaa* 'that, thus' is often used in appositional constructions where it is in apposition to a subordinated clause appearing later in the sentence.[14] In many cases, the

[14]Davis et al. (2020) analyzes this type of appositional construction in depth, calling it backwards resumption.

annotation proceeds as normal following the guidelines: *vaa*, as a pronoun, is made an APPOS dependent of the head of the clause to which it is appositional, and that head itself takes whatever head and syntactic relation is appropriate, generally as a COMP or QUOT dependent of another verb. (74) exemplifies this scenario, with *vaa* appositional to the head of a quote, *u'iv*. Note also that *vaa* in (74) is modified by *kich*, and thus *kich* is an *atr* dependent of *vaa*. Since *kich* follows directly after *vaa* (in its usual position immediately after the word it modifies), it is treated as a dependent of *vaa* not of the clause *vaa* is appositional to.

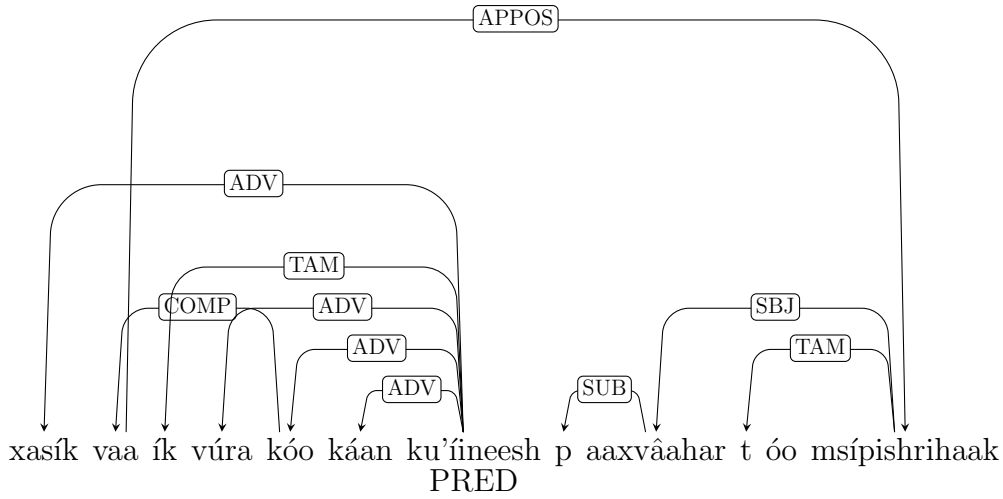
(74) vaa kích u-pí-ti pa=kéevniikich "yôotva t=u-'iv
thus only 3SG-say-DUR the-old.woman hurray PER=3SG-die
pa=nani-'íkam."
the=1SG.POSS-son.in.law
'The old woman was just saying, "Hurray, he's dead, my son-in-law." (Lottie Beck, WB_KL-18:18)



There are cases, however, where *vaa* is not made the APPOS dependent and the head of the clause is made the APPOS dependent, contra the typical pattern of having pronouns bear the *appos* relation. In these cases, *vaa* is clearly the complement of a postposition, and to preserve the annotation of postpositions as following their complements, *vaa* must be made the COMP dependent of the postposition, and the clause to which vaa is appositional has its head made the *appos* dependent to *vaa*. If the typical annotation scheme was followed, we would be forced to make the head of the clause the complement of the postposition, despite the head of the clause appearing much later in the clause. (75) exemplifies this, with *vaa* as the complement to *koo* and the head of the appositive clause, *msípishrihaak*, made an *appos* dependent of *vaa*.

(75) xasík vaa ík vúra kóo káan ku-'íin-eesh p=aaxvâahar t=óo
then thus MOD INT as.much.as there 2PL-do-PROSP SUB=pitch.wood PER-3SG

msípishri-haak ...
be.extinguished-IRR
'You must stay there until the pitch-wood is extinguished ... ' (Julia Starritt, WB_KL-32:39)



2.9.7 *xakaan*, *koovan*, *îin*, and *puraan*

For information on how to annotate the postpositions *xakaan*, *koovan*, and *îin*, please see section 2.7.1.

For information on how to annotate the postposition *puraan*, please see the section on objects on page 2.7.2.

2.9.8 Causatives

Causative suffix *-math/-vath*

Causatives in Karuk can be built with the addition of the suffix *-math* (or its post-vocalic allomorph *-vath*) to verbs, whether transitive or intransitive. The addition of the causative changes the annotation of the particular verb in systematic ways, as diagrammed in (76). Namely, what would've been the SBJ dependent of the verb without the causative is treated as the OBJ dependent of the verb with a causative. Prior OBJ dependents, if the verb was transitive, are treated as COMP dependents of the causative, and the causative can add a new SBJ dependent (the causer).

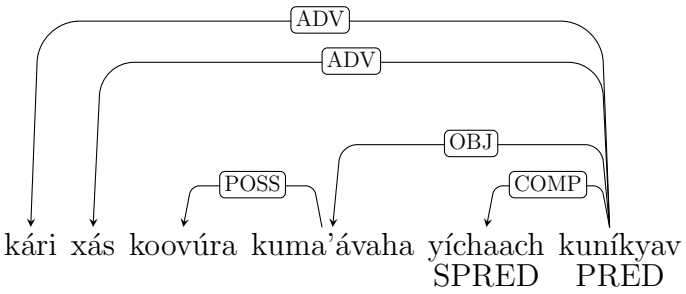
(76)

Before	→	After
SBJ	→	OBJ
OBJ	→	COMP
	→	SBJ

Periphrastic causatives

Periphrastic causatives are those built using a phrasal construction, rather than using the dedicated morphological causative suffix *-math/-vath* described above. In Karuk, periphrastic causatives are formed using the verb *ikyav* 'make' along with another word, typically an adjective or adverb, describing the state which is caused. For example in (77) the meaning 'gather' is expressed by a periphrastic causative glossed as 'make together,' with *ikyav* appearing with *yichaach* 'together.'

(77) kári xás koovúra kuma-'ávaha yíchaach kun-íkyav
and then all kinds.of-food together 3PL>3-make
'So they gathered all kinds of food.' (Mamie Offield, WB_KL-38:2)



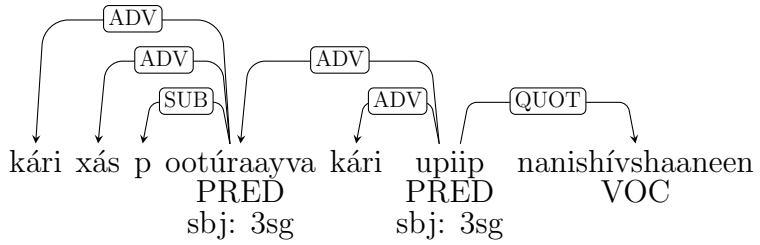
In these cases, the word denoting the caused state (*yichaach* in (77)) is made a COMP dependent of *ikyav*, and is also given the SPRED syntactic status.

2.9.9 Discourse particles separated from main clause by subordinate clause

Following the principle of Attach High, in most cases a discourse particle like *kari* or *xás* will be a dependent of the highest predicate possible, even if there is a subordinated predicate intervening between the discourse particle and the higher predicate. Indeed, this is what is shown in the example given in the section about Attach High, section 2.2.1.

There are occasionally exceptions to this general situation, however. The clearest case is when a discourse particle appears twice: once before a subordinate clause, and again before the main predicate. Note in (78) the position of the two *kári* particles - one preceding a subordinate verb *ootúraayva* and one preceding the main clause verb *upiip*. It would be redundant for a single predicate to have two identical dependents in different positions, and since in these situations there is another valid predicate for the first discourse particle to be a dependent of (namely, the subordinate predicate), we can avoid redundancy by making the first discourse particle a dependent of the subordinate verb. Note that the first *kari* in (78) is a dependent of *ootúraayva* and the second is a dependent of *upiip*.

(78) kári xás p=oo-túraayva kári u-piip “nani-shívshaaneen”
and then SUB-3SG-look.around and 3SG-say 1SG.POSS-land
‘And when he looked around, then he said, “My country!”’ (Mamie Offield, WB_KL-05:143)

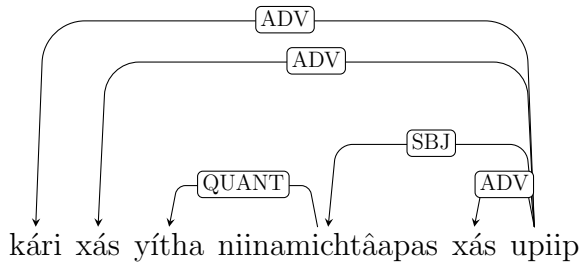


It is not just the first *kari* that is a dependent of the subordinate verb, however. The *xás* which immediately precedes *ootúraayva* is also a dependent of it rather than the main clause, despite there not being two instances of *xás* in the sentence. This is because, generally, discourse particles group together. *kari* and *xas* are frequently used in combination, and though I will not claim they form a constituent, it would be strange for one to be a dependent of the subordinate verb and one to be a dependent of the main clause despite their appearing in essentially the same position in the sentence. Thus, in cases where there is a split with some discourse particles being dependents of a subordinate clause and some being dependents of a main clause (generally due to doubling as described above), non-doubled discourse particles also preceding the subordinate clause will be made dependents of the subordinate clause, following the nearby particle which was forced to be a subordinate clause dependent by virtue of being doubled.

Note also that there are occasionally exceptions where discourse particles, though not doubled, are made dependents of a subordinate clause they precede as opposed to the main clause, because there are discourse particles after the subordinate clause but preceding the main clause which the first particle would have been expected to be grouped with. The reasoning behind such cases is often sentence-specific, and so will not be discussed further here.

A further addendum is necessary. There are cases where a discourse particle is doubled, but there are no possible other predicates which one of the particles can be a dependent of. In these cases, we are forced to treat both of the repeated particles as dependents of the same head despite the redundancy. (79) shows an example, where there are two *xás* particles, but only one predicate they could be dependents of.

(79) kári xás yítha niinamich-tâapas xás u-piip ...
and then one small-most then 3SG-say ...
‘Then the littlest one said ...’ (Chester Pepper, WB_KL-53:17)



However, the existence of such cases does not mean redundancy should not be avoided when it can. It is likely that even in cases where the treebank annotation is forced to be redundant that there is some different, finer-grained purpose the two identical particles are fulfilling which escapes the relatively coarse-grained annotation of the treebank.

2.9.10 Intensifier particle *vúra*

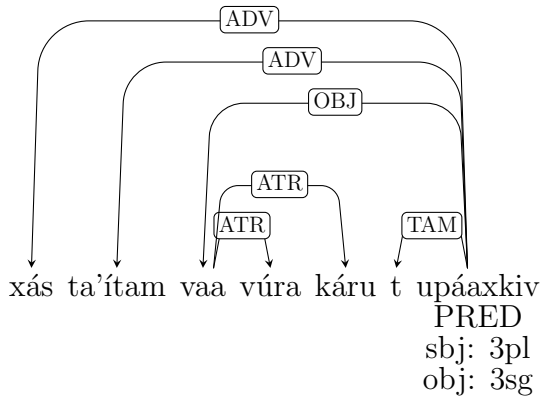
The word *vúra* is incredibly common in Karuk texts, but rarely appears to have any effect on the translation of any sentence it occurs in. It is glossed as an 'intensifier,' something like English 'very' or 'really,' but it rarely is rendered as such or as anything at all in the English translation. As such, it is difficult to determine what any given instance of *vúra* is modifying and thus what word it must be a dependent of. In most cases, this results in a default of *vúra* being annotated as an ADV dependent of whatever the highest available predicate is.

However, if *vúra* is immediately preceded and followed by words that share the same head or by a head and one of its dependents, *vúra* is made a dependent of the relevant head, with either an ADV syntactic relation (if the head is a predicate) or ATR syntactic relation (if the head is a nominal or acting as a nominal.) This is done based on the idea that *vúra* would be inside a constituent or phrase if in such positions, and thus likely to also be a member of that constituent.[15]

Example (80) below exemplifies the latter scenario of being surrounded by a head and one of its dependents: immediately preceding *vúra* is *vaa*, the OBJ dependent of the predicate *upáaxkiv*, and immediately following *vúra* is *káru*, an ATR dependent of the very same *vaa* that precedes *vúra*. As such, *vúra* looks to be inside the constituent headed by *vaa*, and is annotated as such by also being made a dependent of *vaa*. It is an ATR dependent in this case because *vaa* is acting as a nominal, as the OBJ dependent of the verb *upáaxkiv*.

(80) xás ta'ítam vaa vúra káru t=u-páaxkiv
and so that INT also PER=3SG>3-win
'And so (the upriver people) won that too.' (Julia Starritt, WB_KL-10:7)

[15]Split noun phrases are possible in Karuk, but generally such split NPs are split by either the predicate or a more contentful preverbal adverb (one that is not possible to interpret as being a modifier of the relevant noun, in any case, unlike *vúra*).



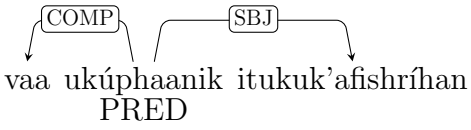
2.9.11 Subcategorization frames for particular verbs

This section contains information about the annotation of particular verbs, generally concerning how to identify their OBJ, COMP, and IO dependents. The verbs below typically exhibit argument structures that are not covered explicitly by the guidelines above and which thus need additional instructions for annotation.

kuupha

In addition to taking a complement clause as described in the COMP section above, the verb *kuupha*, 'do,' may also appear with non-clausal complements which are also annotated as COMP dependents of the verb. Most generally, these complements are the demonstratives *pay* and *vaa*, but occasionally other lexical complements are found. (81) is an example of *vaa* acting as the complement to *kuupha*. Sentences like (81) are common at the end of narratives in the corpus.

(81) vaa u-kúphaa-nik, itukuk-'afishríhan
that 3SG-do-ANC Itúkuk-young.man
'The boy from Itúkuk did that.' (Nettie Ruben, WB_KL-57:152)

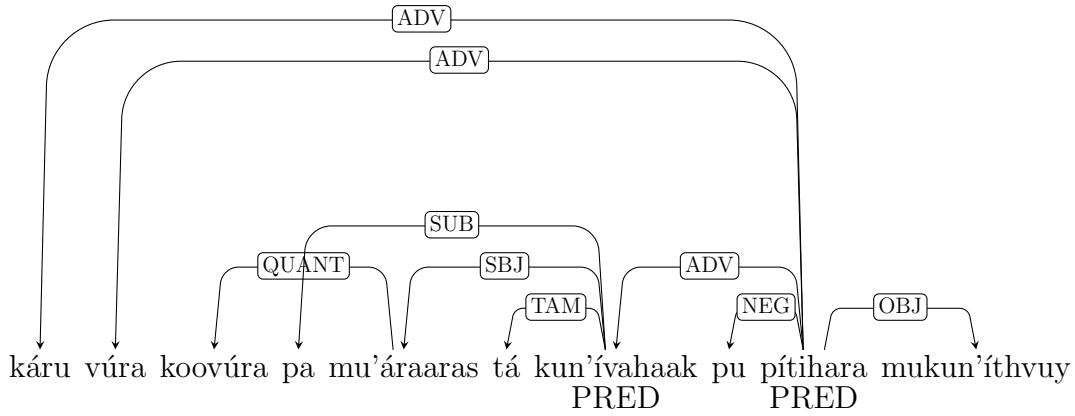


piip

The verb *piip* 'say' can take three types of objects:

- Direct quotes, treated as QUOT dependents. (see 2.7.11 for more information).
- Indirect quotes, treated as COMP dependents.
- Nominal descriptions of what was said, which are treated as OBJ dependents. (82) is an example of a nominal OBJ dependent of *piip*, with *mukun'íthvuy*, 'their names,' being a noun phrase describing, but not directly imitating, what is said.

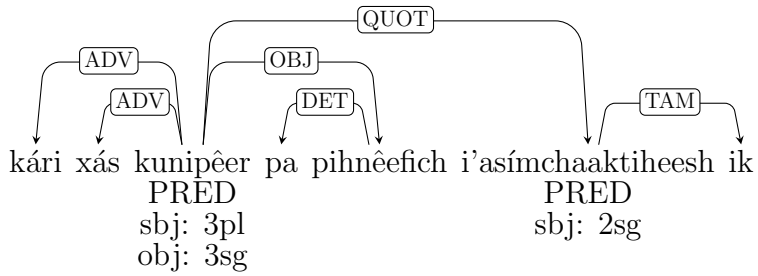
(82) káru vúra koovúra pa=mu-'áraar-as tá kun-'íva-haak pu=pítih-ara
and INT all SUB=3SG.POSS-person-PL PER 3PL>3-die-IRR NEG=say-NEG
mukun-'íthvuy
3PL.POSS-name
'And when any of his relatives died, he did not say their names." (Julia Starritt, WB_KL-0:2)



ipêer

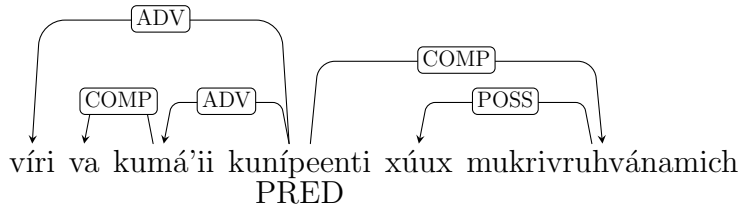
The verb *ipêer* 'tell' is much like *piip* from the section above in that it can take a direct quote as a QUOT complement. However, *ipêer* differs from piip in that it additionally takes an direct object representing the person being addressed by the act of speaking. This addressee is thus made an OBJ dependent of *ipêer*. (83) exemplifies this, with *pihnêefich* 'Coyote' being the one addressed in the speech act described by *ipêer* treated as an OBJ dependent, alongside the actual quoted speech itself (headed by *i'asímchaaktiheesh*) being treated as a QUOT dependent of *ipêer*.

(83) kári xás kun-ipêer pa=pihnêefich "i-'asímchaak-tih-eesh ik'.'
and then 3PL>3 the-coyote 2SG-close.eyes-DUR-PROSP MOD
'And they told Coyote, "You must keep your eyes closed."' (Nettie Ruben, WB_KL-02:39)



ipêer is also used at least once in a construction that looks similar to the usage of *ithvuuymath* 'to call, name' in which *ipêer* is used to describe how people call a particular place - in other words, to assign a name to a place, rather than address anyone. In such cases, *ipêer* is given an analysis like *ithvuuymath* (see 2.9.11 for details), with the entity or place being named treated as an OBJ dependent, and the name itself treated as a COMP dependent.

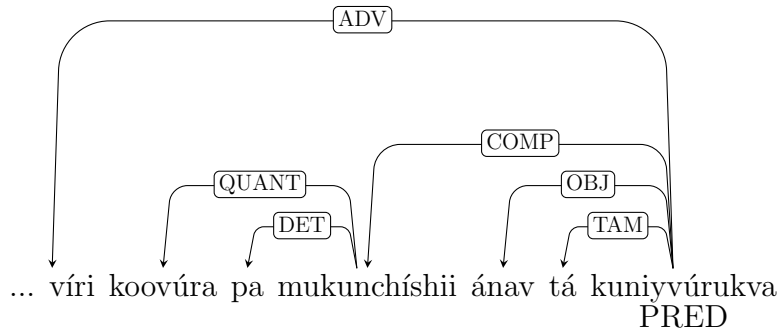
(84) víri=va kumá'ii kun-ípeen-ti xúux
so=thus because.of 3PL>3-tell-DUR legendary.animal
mu-krivruh-vá-nam-ich
3SG.POSS-roll-PL.ACT-place-DIM
'For that reason they call it 'xuux's little rolling place." (Chester Pepper, WB_KL-26:18)



iyvuruk

The verb *iyvuruk* 'to rub on' takes an OBJ dependent representing the substance which is rubbed on a surface, and a COMP dependent representing the surface or entity which the substance is rubbed onto. This is exemplified with (85) below, where *ánav* 'medicine' is the substance (and thus OBJ dependent) rubbed onto the surface of the dogs (the COMP dependent), *chíshii*.

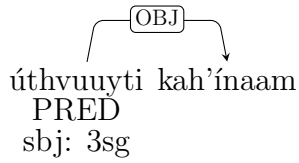
(85) ... víri koovúra pa=mukun-chíshii ánav tá kun-iyvúruk-va
... so all the-3PL.POSS-dog medicine PER 3PL>3-rub.on-PL.ACT
'... they (their dogs) would all be rubbed with medicine.' (Nettie Ruben, WB_KL-70:14)



ithvuy and *ithvuuymath*

The verb *ithvuy* 'to be called' takes an OBJ dependent representing the name which the subject is to be called. For example, (86) expresses how a certain place (the unexpressed subject) is called by the name *kah'ínaam*, the latter word treated as the OBJ dependent of the verb *úthvuuyti*.

(86) ú-thvuuy-ti kah'ínaam
3SG>3-be.named-DUR place.name
'(The place) is called kah'ínaam.' (Maggie Charley, WB_KL-84:2)

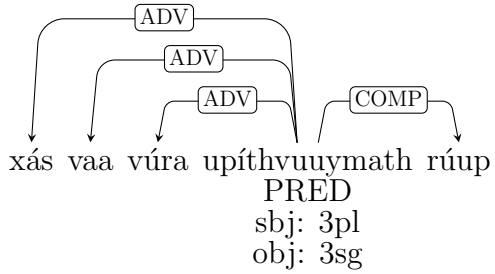


A relatively common form of *ithvuy* is its causative form, *ithvuuymath* 'to name something.' The analysis of the causative form is straightforwardly derived from the analysis of *ithvuy* given here plus the analysis of *-math* described in 2.9.8, but for concreteness this analysis will be explicated here.

The verb *ithvuuymath* 'to name something' can take an OBJ dependent representing the entity being named, and a COMP dependent representing the name being applied to the entity. Note that the entity being named would be a SBJ dependent for the non-causative form *ithvuy* and that the name being applied would be the OBJ dependent of the non-causative form - the causative changes these to OBJ and COMP dependents respectively. For example, in (87) below, the name being applied is RÚUP, which is treated as a COMP dependent of *upíthvuuymath*.

(87) xás vaa vúra u-p-íthvuuy-math rúup
and thus INT 3SG>3-ITER-be.named-CAUS Rube

'So they named him Rube.' (Julia Starritt, WB_KL-66:21)



2.10 Conclusion

This chapter presented the guidelines for the Karuk treebank, from its guiding principles to the minutiae of the analysis of specific verbs like *piip* 'say.' These guidelines were used by the annotators of the treebank while completing their annotations to faciiltate consistency across annotators. Now, they serve as a description of what the annotation consists of, which is crucial to being able to use the treebank to search through the Karuk corpus and understand the results of such searches. In the following chapters, I describe the results of treebank searches investigating two areas of inquiry: the word orders of arguments and predicates, and the Karuk agreement system.

Chapter 3

Argument and Predicate Order

This chapter investigates the relative order of arguments and predicates in Karuk. It seeks to answer the following questions:

1. What orders of argument and predicate are attested in the corpus, and how common is each? This includes not just the typical orders of subject, verb, and object in transitive sentences with two expressed arguments, but the orders in sentences with other types of arguments, namely indirect objects and complements, and sentences in which only a single argument is expressed.

2. What factors influence the distribution of the word orders? This includes examining interactions between basic properties like transitivity and subordination and the position or expression of certain arguments, as well as the effect that the expression of one argument has on the position of another.

The purpose of answering these questions, beyond the inherent value of establishing these facts, is to showcase the utility of the Karuk treebank by presenting some of the types of data that can be accessed using it. As such, this chapter seeks after a wide breadth of observations, rather than a deep exploration of any given one. Three primary generalizations will be formed from some of these observations, listed below.

(1) Subordinate Argument Suppression Tendency: Subordinate clauses have a tendency to have fewer expressed arguments than main clauses

(2) Transitive Subject Suppression Tendency: Subjects are less likely to be expressed in transitive clauses

(3) Object-Conditioned Pre-Verbal Subject Tendency: If a sentence has an expressed subject and object, the subject is more likely to be pre-verbal than if the object was not expressed.

To begin the discussion, I make a brief note about the general search methodology used to probe the treebank in this section. Following that, I summarize previous work on Karuk dealing with the question of argument-predicate order. Overall, the findings here largely align with the impressionistic descriptions of word order given in earlier sources, but put a

sharper point on the data and provide a surer empirical foundation for the observations.

Subsequently, I set the foundation for further discussion by presenting data on basic properties of clauses such as transitivity, main or subordinate status, and headedness by verb or non-verb. Afterwards, I present data on the word orders of clauses with two or more arguments and word orders of clauses with a single argument. Following that, I discuss several statistical interactions extant in the data, discussing each of the three generalizations above in more depth and ending with some additional interactions that don't fit into those generalizations.

3.1 General search methodology

Below I report various numerical counts of particular types of clauses in the corpus, such as the amount of transitive clauses. To derive this count in each case, I wrote a Python script which crucially utilizes the *ElementTree* package, a Python module for manipulating xml with Python, to iterate through the xml files containing the syntactic annotation of each text in the corpus. This script looks for sentences/words that have whatever attributes necessary for that search and returns a list and count of the matching sentences/words. For searches requiring some lexical information not specified in the text xml files, the script also accesses an xml file of the Karuk lexicon. For each search below I give a brief description of exactly what attributes were relevant. I present these descriptions in terms familiar from the annotation guidelines of the treebank, rather than in terms of what the Python script is actually doing in programming terms.

3.2 Previous work

Karuk is often described as a 'free' word order language, typically meant to indicate that most orders of subject, predicate, and direct object are possible. For instance, de Angulo and Freeland (1931) state that 'most of the business of the language goes on within the verb' and that 'the nouns... are interspersed between the verbs without syntactic cases or fixed order to show their relations.' (p. 194). Bright (1957), after discussing the composition of predicates and noun phrases, states that "the syntactic elements which have been described above" (meaning, predicates and noun phrases) "are combined into predications with nearly maximum freedom of word-order." (p.141). Neither DeAngulo and Freeland nor Bright, however, include any detail regarding exactly what word orders are found,[1] though Bright does mention one restriction on subject, object, and predicate order. Namely, Bright says that "when a transpersonal verb stem is present... the subject precedes the object." (p.134) 'Transpersonal stem' is the term Bright uses for those verbs which "occur with the full

[1]Harrington, in his 3 publications dealing with Karuk, makes no mention of word order or syntax at all; his only grammatical notes deal with phonetics and phonology (Harrington 1930; Harrington 1932b; Harrington 1932a).

total of sixteen personal morphemes [=agreement morphemes]" (p. 59) rather than just the agreement morphemes that indicate 3rd person objects - transpersonal stems can have local person objects, in other words. Bright gives a semantic description of these stems as well: transpersonal stems are verbs which "designate actions which may be performed by animate beings with other animate beings as objects." (p. 59), Interpreting this we can surmise that, with Bright's transpersonal verbs, only SOV, SVO, and VSO orders should be available, but Bright gives no further information about other orders available for other types of verbs or the relative prevalence of any orders for any verb type.

More recently, Garrett, Gehr, et al. (in press) describe the word order possibilities impressionistically:[2]

> As for word order, while clauses with multiple overt verbal arguments are not the norm, most permutations of Subject, Object, and Verb are attested on the surface. Only the SOV and SVO patterns... seem common; the absence of VSO examples may be an accident. (p.12)

These recent sources have confirmed the occurrence of most orders of subject, verb, and object, with the interesting exception of VSO, and in the case of Garrett, Gehr, et al. (in press) have provided impressions of how common certain patterns are. Further, Yu (2021) has delved in deeper to a subset of the Karuk corpus to explore what factors govern the appearance of different orders, finding that, for subjects, greater difference between mentions correlates to a higher rate of pre-verbal appearance, while for objects, animacy correlates to a higher rate of post-verbal appearance. However, to date there has been no systematic quantitative survey of the word orders attested in the Karuk corpus and what exactly their relative prevalence is. This chapter fills in that gap.

Further, no previous study has systematically discussed the occurrence or prevalence of sentences which do not include all three of subject, predicate, and object, or which include indirect objects or other kinds of complements. Macaulay (2000), in a discussion of 147 transitive sentences including the obviative marker *îin*, does report on the order of subject and verb and the expression of objects, including a number of transitive sentences that only have a subject and verb expressed and no object, but does not discuss these properties outside of that small subset of sentences. Previous work has indicated that sentences with two or more overt arguments are rare (as per the Garrett et al quote above), but has not been able to show how rare such sentences are, or what the composition of sentences with one or less overt arguments is. In addition, interactions between word order and other syntactic factors like transitivity or subordination have not been discussed for Karuk previously. This section brings such interactions to light for the first time.

More generally, this study is, to my knowledge, the first quantitative study of word order in an indigenous American language which utilizes a treebank. Reliance on the treebank

[2]Though I am a co-author on Garrett, Gehr, et al. (in press), all work on that chapter was completed before annotation for the treebank was completed, and as such does not reflect findings of the treebank, which are solely reported here.

offers significant benefits in the scale of accessible data, the breadth of syntactic properties covered, and the speed at which data can be collected and characterized for any given research question. This later benefit is a major one: most studies are laser-focused on only one specific question, and as such do not annotate any data unrelated to that question. However, this means that the annotations completed are thus not of much use for any further study beyond the original one. A treebank, on the other hand, is not designed for a particular question, but more for a particular field of possible questions and as such can prove useful for a large number of future studies, including ones the treebank creator did not foresee.

To make these benefits of treebanks clearer, let us consider two other quantitative studies of American languages: Meyer (1992) on Klamath and Tonhauser and Colijn (2010) on Paraguayan Guaraní. Both of these languages, like Karuk, exhibit free order of argument and verb, and these studies generally attempt to understand the order of S, V, and O in these languages. Like the Karuk treebank, both of these studies utilize manual annotation of texts according to some criteria, though these studies, as far as I am aware, do not utilize any computational tools in assessing or measuring the annotated corpus, unlike the present study with the treebank and search scripts.

As for scale, Meyer (1992) reports on a corpus of 7 texts (five traditional myths and two ethnographic/procedural texts, a similar sort of corpus to the Karuk treebank) consisting of 761 clauses. Tonhauser and Colijn (2010) worked with a corpus of 8 texts consisting of around 2,800 words; they do not report a clause number. These studies represent much smaller corpora than the Karuk corpus utilized in the treebank, which, as described later, numbers around 5300 clauses in 92 texts. Of course, treebanks are not inherently larger than any other corpora; one can of course have a treebank of comparable size to the corpora in Meyer (1992) and Tonhauser and Colijn (2010) or even smaller, though the utility of a treebank, with its generalized nature, is arguably more dependent on the size of the corpus and as such, treebanks are going to involve larger corpora just to ensure the utility is worth the time investment of creating the treebank.

As for breadth, let us consider what is annotated in Meyer (1992) and Tonhauser and Colijn (2010). Meyer (1992) follows the framework for assessing NP topicality laid out in Givón (1983), which crucially involves calculating the amount of clauses between each mention of a referent. This involves annotating the appearance of NPs and their referents, as well as identifying clauses in the corpus. This latter point is of interest; because Givón's system of tracking referents requires counting clauses where a referent is not mentioned, certain phenomena we would classify as clauses on grammatical grounds, namely relative clauses and (some) complement clauses, are not included in the count of clauses, else the count of clauses be inflated in a way not suitable for Givón's system. This is all well and good for the study itself, but it means that the data reported by the study are less useful for someone coming to the data later with a different research question in mind that is not committed to that particular system. Likewise, Tonhauser and Colijn (2010) exempt all subordinate clauses from annotation, because 'subordinate clauses have a tendency' cross-linguistically 'to exhibit a more rigid word order' and their study is only focused on understanding cases of word order flexibility. Of course, for a single study it is important to restrict one's attention

to those areas that are actually relevant, but it does mean the resulting annotations are not as useful for subsequent research. Tonhauser and Colijn (2010) are also only interested in the order of the arguments and the verb, and do not annotate any of the other elements in a clause, including nominal adjuncts. The Karuk treebank, and any treebank for that matter, takes a different approach: every element in every sentence is annotated for the desired syntactic features. Thus, although in this chapter I will not discuss the position of adverbs or nominal adjuncts, the data is nevertheless annotated in the treebank and thus easily accessible for further research without any need for extra time-consuming annotation. Of course, Meyer (1992) and Tonhauser and Colijn (2010) do annotate some features that treebanks typically do not and that the Karuk treebank does not, such as the measure of topic continuity in Meyer (1992) and the discourse status of argument NPs in Tonhauser and Colijn (2010). Overall, however, the breadth of a treebank, designed to be a general tool, is greater.

To discuss the prevalence of various orders of arguments and predicate, it is useful to first know the distribution of some basic clause types in the treebank. Namely, such a discussion will need to reference transitivity, main versus subordinate clause status, and verbal versus non-verbal clause status. As such, I will first discuss the prevalence of each of these categories in the treebank in the section below before turning to word orders.

3.3 Basic properties of clauses

In this section I present data on the prevalence of some basic properties of clauses in the treebank, namely the distribution of intransitive versus transitive predicates, main clauses versus subordinate clauses, and verbal versus non-verbal predicates, ending with a discussion of detectable interactions between these categories. The main point here is to provide some foundational information about the types of clauses present in the treebank, so that these facts are accessible for the later discussions of word orders that are affected by these properties.

The focus of this section is on presenting the data, rather than providing explanations of the reported distributions in each case. Very few other treebanks or corpora have readily accessible reports of this sort of basic information. Because of this, it will not be possible to robustly compare the Karuk results below to other languages, which would allow for more informed interpretations of the data. I report the variety of data I do here, even if the data leads to no particular conclusions, because the topic is so unexplored and, as such, expectations for what sorts of data matter are not yet set. Reporting these particular properties also should make clear part of the utility of the treebank in assessing corpus-wide properties that are otherwise too time-consuming to assess.

I start by examining transitivity, moving then to subordination, then to verbal versus non-verbal status, and ending with a discussion of interactions among these.

3.3.1 Transitivity

Table 3.1 below presents the breakdown of the corpus by transitivity and gives the total number of clauses in the treebank. Figure 3.1 displays the information in a bar graph.

Category	# of Occurrences	% of total clauses
Intransitive	3660	68.39%
Monotransitive	1645	30.74%
Ditransitive	47	0.88%
Transitive total	1692	31.61%
Total clauses	5352	100.00%

Table 3.1: Number of clauses by transitivity

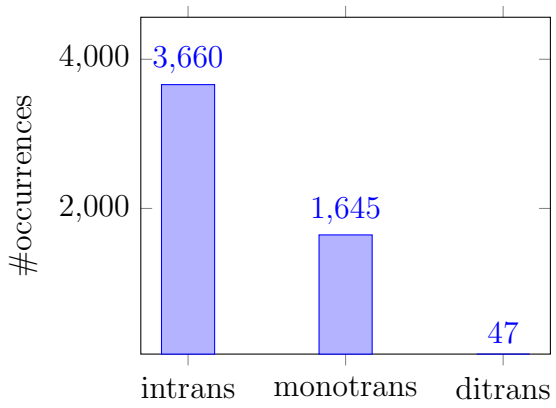


Figure 3.1: Number of clauses by transitivity

Discussion

What is immediately striking about the data above is the large bias towards intransitives, with intransitives making up over two-thirds of the corpus. Information on the distribution of transitivity in other languages is mostly lacking, though Roland, Dick, and Elman (2007) present data for English that, based on the Brown corpus, suggest about 57% of sentences fall into an intransitive category, while about 42% fall into a transitive category.[3] Thus,

[3]Roland, Dick, and Elman (2007) also report percentages for four other corpora, but I chose the Brown corpus for comparison. Roland, Dick, and Elman (2007) does not report these statistics directly; they are derived from more fine-grained statistics in their Table 15 (p.61), which report on the relative frequency of a large number of specific subcategorization types attested in sentences in the corpus. I categorized

in the English represented in the Brown corpus intransitives seem slightly favored, though not nearly as much as in the Karuk treebank. Whether this discrepancy is connected to linguistic differences between the two languages or merely reflects differences in the nature of the two corpora (the Karuk treebank being comprised mainly of narratives and entirely of texts which were originally spoken) is a question that would be best answered with more data on more languages.

Another striking fact about the data above is the stark infrequency of ditransitives in the treebank. Ditransitives (meaning verbs that take both a direct and indirect object) make up less than 1 percent of all clauses. This infrequency is comparable to the English data from Roland, Dick, and Elman (2007), which report for the Brown corpus that only about 1% of sentences have a ditransitive verb. This is despite the fact that ditransitives in the Karuk treebank are strictly defined as involving verbs that denote literal transfer of possession, whereas Roland, Dick, and Elman (2007) appear to include metaphorical transfers of possession (or base their coding on grammatical factors observable in English) as evidence by their use of the following sentence as an example of a ditransitive: 'The mayor of the town taught them English and French.'

Search methodology

The number of overall clauses was calculated by a Python script which counted all examples of the PRED synstat tag in the corpus. This includes both verbal and non-verbal predicates, and main and subordinate clauses. The number of PRED tags is a suitable measure for number of clauses (as well as number of predicates) because the PRED tag is given to only the heads of clauses. Note that there are sentences in the database which do not have a predicate (they may have a FRAG tag as the head, for example), but such examples are not clausal and have no bearing on the data reported above. To categorize each particular predicate in terms of transitivity, the script looked for the existence of indirect and direct object agreement tags on each predicate. These agreement tags are based on the meaning of a sentence, as described in the section on agreement tags in the annotation guidelines, section 2.8, and not on the presence of an object or indirect object. As such, even sentences with no expressed objects or indirect objects can be correctly identified as transitive or ditransitive if those arguments were dropped. If a predicate was found to have an indirect object (IO)

these types as being intransitive or transitive and added the percentages together to get the percentages reported above. The names of their categories which I assigned to intransitive are: simple intransitive, prepositional phrase, *to* infinitive verb phrase, prepositional phrase + *to* infinitive verb phrase, *wh* clause, sentential complement with complementizer, sentential complement (no complementizer), gerundive verb phrase, perception complement, and passive. The names of their categories I assigned to transitive are: simple transitive, ditransitive, transitive + prepositional phrase, transitive + *to* infinitive verb phrase, transitive + *wh* clause, transitive + sentential complement with complementizer, transitive + sentential complement (no complementizer). Note that verbs which take various complements in English that are not direct objects (e.g. sentential complements, prepositional phrases) are treated as intransitive like the comparable sentences in the Karuk treebank are. As such, the difference in prevalence of transitivity between Karuk and English cannot merely be due to differences in how clauses were coded as transitive or not.

tag, it was marked as a ditransitive. If no IO tag was found, the script looked to see if there was a direct object (OBJ) tag and if there was, marked the predicate as transitive. If no IO or OBJ tag was found, the script classified the predicate as intransitive. Note that classifying transitive by the presence of a direct object means that verbs which take what is coded in the treebank as a complement but not a direct object are treated as intransitive; two common verbal types like this are verbs with applicative suffixes introducing complements and most verbs of speaking/thinking.

Note that there is an important difference between categorizing transitivity by the agreement tags as I have done and by categorizing based on whether a verb root is thought to be transitive or not. The latter may seem to be the more obvious route, but would rely on that sort of categorization being available. In the Karuk treebank, there is no direct encoding of any particular verb's transitivity, chiefly because the previously-created lexical database which the treebank is built on top of does not include this information. However, this does not mean that relying on the agreement tags is a worse solution, meant only to find a backdoor way to get at transitivity in the absence of direct encoding. Rather, because the agreement tags are entirely dependent on the context of each individual clause, they are much more likely to capture the actual transitivity of each clause, whereas relying on a general encoding of each verb as transitive or intransitive will likely result in errors. Of course, labile verbs (that in some uses are transitive and in others are intransitive) are well-known and common in languages like English (For example, the verb 'lock' used transitively in 'I locked the door' and intransitively in "That door locks.'). If relying on lexicon-level categorization, a verb marked as transitive in the lexicon could be used intransitively but still marked incorrectly as transitive. Of course, the accuracy of the agreement tags is key to a correct categorization of transitivity. For more information, please see section 2.8.

3.3.2 Subordination

Table 3.2 below presents data on the distribution of main versus subordinate clauses in the corpus. Figure 3.2 summarizes the results in a bar graph.

Category	# of Occurrences	% of total clauses
Main clause	4946	92.41%
Subordinate clause	406	7.59%
Total	5352	100.00%

Table 3.2: Number of main vs subordinate clauses

Discussion

The results here are stark, with over 90% of predicates being the heads of main clauses, and only a relatively small number of true subordinate clauses. There is a dearth of studies

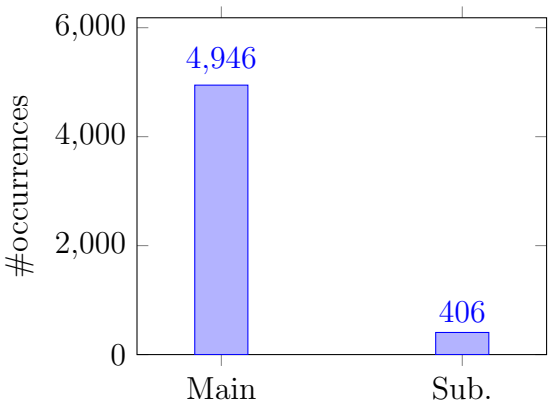


Figure 3.2: Number of main vs. subordinate clauses

specifically on the prevalence of subordinate clauses versus main clauses in an entire corpus for other languages, but Kempen and Harbusch (2019) do report on some frequencies of these types for German, Dutch, and English in the process of discussing a statistical relationship between clause type and high-frequency verbs. Table 3.3 summarizes data from Kempen and Harbusch (2019)'s Table 4 (p. 1145).[4]

Language	Main %	Sub %	Non-finite %
German	71.0%	8.6%	20.5%
Dutch	59.2%	16.3%	24.5%
English	46.5%	22.5%	31%

Table 3.3: Main and subordinate clause frequences in German, Dutch, and English (Kempen and Harbusch 2019)

It should be noted that Kempen and Harbusch (2019) separate out non-finite clauses from other subordinate clauses. This distinction is not relevant for Karuk, as Karuk lacks non-finite clauses and all Karuk subordinate clauses are finite. Kempen and Harbusch (2019) note that their subordinate clause category contains 'complement, adverbial, and relative clauses' (p. 1144), which is exactly the types comprising the subordinate clause category used for Karuk above. Whether the appropriate category for comparison to Karuk is only the subordinate category, or the sum of the subordinate and non-finite category (all of which are subordinate clauses), is unclear.

[4]Raw numbers are not presented in the table here for ease of comparison to Karuk, but the total number of clauses for each language is as such: German: 49,879; Dutch: 159,299; English, 165,027. These are much larger than the Karuk treebank's 5352 clauses, though note there is significant differences in size even among the three Germanic treebanks Kempen and Harbusch use.

Overall, there appear to be significant differences between the three Germanic languages Kempen and Harbusch studied in the prevalence of subordinate clauses, but all have more common subordinate clauses than Karuk, which has only 7.6% of clauses as subordinate. This remains true whether comparing the Karuk data only to the subordinate clause column or to the sum of the subordinate and non-finite columns. Karuk is closest to German in subordinate clause prevalence (7.6% vs. 8.6%) if one leaves out the non-finite category, but far below the German subordinate clause prevalence if comparing to the sum of non-finite and subordinate (7.6% vs. 29.1%).

Given this, all that can be said is that Karuk is potentially an outlier in its low frequency of subordinate clauses, but it is difficult to make a certain claim when comparing Karuk only to three languages from a single language family. It is also possible that genre differences in the corpora used in the Karuk treebank versus Kempen and Harbusch (2019)'s study make a difference; Kempen and Harbusch (2019) collected sentences from large corpora of the three languages, focused on data they describe as consisting of "sentences extemporaneously produced in varied dialogue situations (face-to-face or telephone conversations)" (p. 1143). In contrast, the corpus of the Karuk treebank is made up of stories told in monologue and with a bent toward traditional mythological narratives.

Search methodology

To obtain this data, the Python script looked first for whether the predicate had a head. Predicates were categorized as main clauses if they had no head, since those predicates would then be the highest word in their clause. If a predicate had a head, there were several options. First, if the predicate was a COORD dependent of another predicate that itself did not have a head, the predicate was coded as a main clause, on the reasoning that a predicate coordinated as a main clause is not really subordinate but only forced to be a dependent because of the nature of the dependency grammar used in the treebank. Second, if the predicate was a QUOT dependent, it was coded as a main clause, since quotes exhibit the syntax of main clauses and not of subordinate clauses in lacking the subordinator *pa=*. If neither of those conditions were met, the predicate was coded as a subordinate clause. Based on this classification, the category of subordinate clause includes adverbial clauses, complement clauses, and relative clauses.

3.3.3 Verbal and non-verbal predicates

This section presents data on the distribution of verbal vs non-verbal predicates in the corpus. Table 3.4 presents the data, and Figure 3.3 summarizes with a bar graph.

Discussion

Clauses with a verbal predicate far outnumber clauses with a non-verbal predicate, with over 90% of clauses headed by a verbal predicate. It is not known whether this is a surprising result

Category	# of Occurrences	% of total clauses
Verbal	4853	90.67%
Nonverbal	499	9.32%
Total	5352	100.00%

Table 3.4: Number of verbal vs. nonverbal clauses

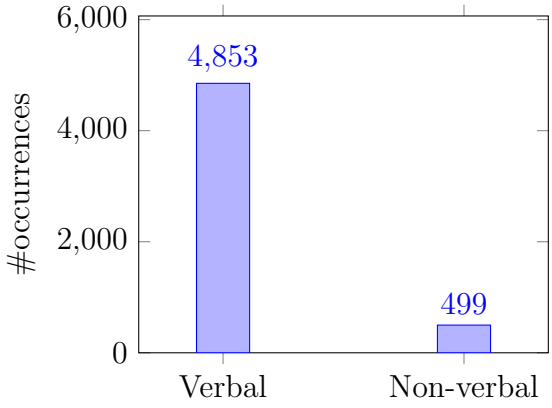


Figure 3.3: Number of verbal vs. nonverbal clauses

or an expected one. As with the other properties examined in this section, there is little accessible information from corpora of other languages about the prevalence of sentences equivalent to Karuk sentences with non-verbal predicates, whether copular sentences or sentences with direct non-verbal predicates like Karuk. Presumably, such information is possible to derive from treebank searches for other languages, though doing such searches is beyond the scope of this work. The information on Karuk is presented here as a first step towards exploring this topic in further cross-linguistic depth in the future.

Search methodology

To obtain this data, a Python script categorized predicates into verbal and non-verbal predicates on the basis of whether said predicates included any morphemes with a 'V' part of speech tag in the lexicon xml file. Because Karuk verbal predicates include a large amount of morphemes, each with their own part of speech tag in the lexicon, verbal predicates are not directly marked in any text as being verbal. This necessitated checking out the parts of speech of each morpheme. If a morpheme with a 'V' part of speech tag was found, the predicate in question was categorized as verbal. If no such morpheme was found, the predicate was categorized as non-verbal.

3.3.4 Interactions

Having established the distribution of basic clause types in the corpus, it is worth noting that, of course, these properties all overlap; transitive predicates can also be subordinate predicates, main clauses can be non-verbal, and so on. Because of this overlap, there is the potential that there are interactions between these properties. Since this is a relatively unexplored topic, we cannot know in advance what properties may influence others, and as such it is worth exploring these potential interactions systematically.

Two interactions are discussed here: the potential interaction of main versus subordinate clause status and transitivity, and the potential interaction of main versus subordinate clause status and whether the predicate is verbal or non-verbal. As all non-verbal predicates are intransitive, the interaction of verbal versus non-verbal predicate status and transitivity is not further discussed; we can firmly say in that case there is an interaction, because the nature of non-verbal predicates ensures they lack direct objects.

Interactions between transitivity and main/subordinate clause status

To assess the interactions between transitivity and main versus subordinate clause status, I present in Table 3.5 below the percentages of each transitivity category within each of the main and subordinate clause categories. If the null hypothesis, that these two properties are not significantly related, is true, we should see that the percentage of each transitivity category in the main and subordinate clause categories is similar to the overall percentage of that transitivity category in the corpus. Figure 3.4 presents the information in a bar graph.

		Intransitive	Transitive	Ditransitive
Main		3369	1533	44
	% of total Main (n=4946)	68.12%	30.99%	0.89%
Sub		291	112	3
	% of total Sub (n=406)	71.67%	27.59%	0.74%
All		3660	1645	47
	% of total clauses (n=5352)	68.39%	30.74%	0.88%

Table 3.5: Main/subordinate x Transitivity

Impressionistically, the null hypothesis of no interaction appears to be supported. For example, intransitives make up roughly 68% of main clauses and all total clauses, while making up roughly 72% of subordinate clauses. Transitives have a similar amount of variance, while ditransitives have very little variance across the main versus subordinate categories. A Chi Square test[5] on this data confirms the impression that these differences are not significant;

[5]Chi square test was performed with the *chi2_contingency* function of the *scipy.stats* module of the SciPy python package.

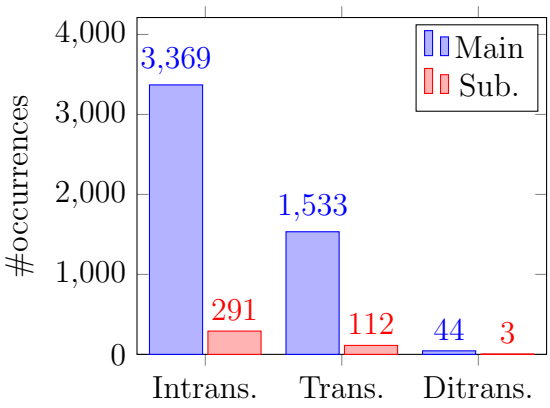


Figure 3.4: Clauses by transitivity and subordination

$X^2(2,N{=}5352) = 1.38$, $p = .50$. As such, this data is consistent with the null hypothesis that no interaction holds between whether a clause is a main or subordinate clause and the transitivity of that clause.

Interactions between main/subordinate clause status and verbal/non-verbal predicate status

To assess the interactions between main versus subordinate clause status and verbal versus non-verbal predicate status, I present in Table 3.6 below the percentages of verbal and non-verbal predicates in each of the main and subordinate predicate categories, with the bottom rows presenting the total number of verbal and non-verbal predicates for comparison. If the null hypothesis, that these two properties are not significantly related, is true, we should see that the percentage of verbal or non-verbal predicates in the main and subordinate clause categories is similar to the overall percentages in the corpus. Figure 3.5 presents the information in a bar graph.

		Verbal	Non-verbal
Main		4467	479
	% of total Main (n=4946)	90.32%	9.68%
Sub		386	20
	% of total Sub (n=406)	95.07%	4.93%
All		4853	499
	% of total clauses (n=5352)	90.68%	9.32%

Table 3.6: Main/subordinate x Verbal/Non-verbal

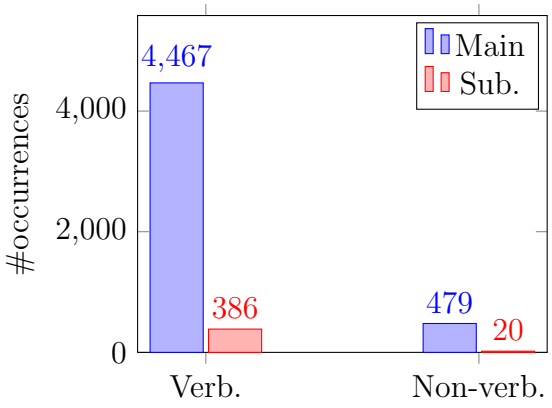


Figure 3.5: Clauses by transitivity and verbal versus non-verbal status

The percentage of subordinate clauses which have verbal predicates is roughly 4% greater than the percentage of total clauses with a verbal predicate. Impressionistically, this seems to be a small difference; however, the results of a Chi Square test on this data indicate that this difference is significant. $X^2(1,N{=}5352) = 9.49$, $p = .002$. This data suggests it is more likely for a subordinate clause to have a verbal predicate instead of non-verbal than is expected merely based on the overall bias towards verbal predicates. In other words, this amounts to a bias against non-verbal predicates in subordinate clauses. The small difference shows this is not anywhere near a categorical effect, though; it is not ungrammatical to have a subordinate non-verbal predicate, merely slightly less likely than expected. Regardless, some factor intervenes to make subordinate clauses a slightly less preferred environment for non-verbal predicates. I will not speculate on what this factor may be, but only note that the causes and robustness of this novel interaction should be addressed by future research in this heretofore unexplored topic of study.

3.4 Clauses with two or more expressed arguments

In this section, I move on to addressing the first question mentioned in the introduction. What orders of argument and predicate are attested in the corpus, and how common is each? To begin I present results on the prevalence of the various possible orders of clauses with two or more expressed arguments. The arguments under consideration are subjects (2.7.1), direct objects (2.7.2), indirect objects (2.7.3), and complements (2.7.10). Readers are encouraged to refer to the aforementioned sections in the annotation guidelines on each of these argument types to see which sorts of phenomena are counted under each.

3.4.1 Clauses with subject and direct object

The table below summarizes the results of a survey of the treebank regarding the orders of arguments and predicates in clauses with at least an expressed subject and direct object. The table gives the percentage that each order takes up in clauses with subject and direct object expressed, transitive clauses, and the total number of clauses. Note that the table below uses the letters S to represent subject, O to represent direct object, and V to represent verb, written in the order in which they occur in the sentences described. Both main and subordinate predicates are also counted in this survey, so that a subordinate clause with subject, object, and verb in that order is counted as an occurence of SOV order, exactly as a main clause would be. Because clauses with objects are necessarily verbal (non-verbal predicates do not take objects), the orders here are represented with V for verb.

Clause order	# of occurrences	% of S,O,V clauses (n=135)	% of trans. clauses (n=1645)	% of total clauses (n=5352)
S V O	61	45.19%	3.71%	1.14%
S O V	52	38.52%	3.16%	0.97%
O V S	9	6.67%	0.55%	0.17%
O S V	8	5.93%	0.49%	0.15%
V O S	4	2.96%	0.24%	0.07%
V S O	1	0.74%	0.06%	0.02%
S first	113	83.70%	6.87%	2.11%
Non-S first	22	16.29%	1.34%	0.41%
Pre-verbal O	69	51.11%	4.19%	1.29%
Post-verbal O	66	48.89%	4.01%	1.23%
All	135	100%	8.21%	2.52%

Table 3.7: Orders of clauses with S and O expressed

A few facts should be highlighted. Most strikingly, the amount of clauses with both a subject and object expressed is exceedingly small relative to the amount of clauses found in the corpus, with only 2.52% of clauses overall having both a subject and object expressed, and only 8.21% of transitive clauses. This matches with what Garrett, Gehr, et al. (in press) stated as quoted above, that such sentences were rare, but the extent of this rarity cannot be overemphasized.

We can also see that SVO and SOV, the two subject-first orders, are far more common than the other orders, with the two subject-first orders accounting for 83.7% of all clauses with subjects and objects both expressed and other orders accounting for only 16.29%. Garrett et al.' s claim that SOV and SVO sentences are most common among these is thus also well-supported here, though again the quantitative data here put a sharper point on it than was possible with impressions alone.

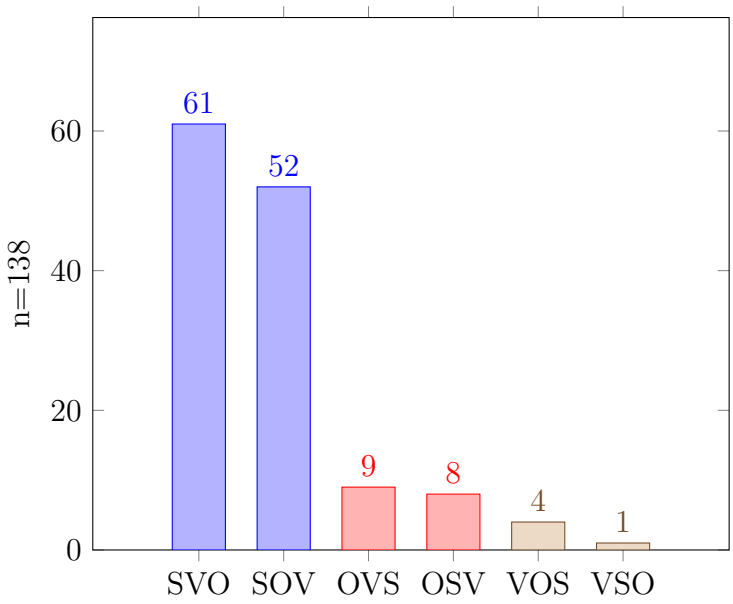


Figure 3.6: Clauses with S and O

A potentially interesting comparison to this data is the data on the cross-linguistic prevalence of basic word orders from the World Atlas of Linguistic Structures (WALS), specifically Feature 81A: Order of Subject, Object, and Verb (Dryer 2013). WALS categorizes a sample of 1,187 languages on the basis of their dominant word order of subject, direct object, and verb. If a language had no dominant order, it is coded as such. Karuk is included in their sample, categorized correctly as having no dominant word order. It would be striking indeed if languages with word order freedom like Karuk had a similar distribution of word order frequencies as languages do overall. Table 3.8 summarizes the WALS data. The third column, '% of languages with dom. order,' is the percentage out of the amount of languages in their sample that have an identified dominant order. In other words, that column excludes the 'No dominant order' category. This exclusion is included for the benefit of comparing to the clausal word order prevalences in Karuk, since, at the clause level, there can be no equivalent to 'no dominant order.'

Comparing the cross-linguistic data to the Karuk clausal data, we find that, in Karuk clauses, from most frequent to least, the orders are SVO > SOV > OVS > OSV > VOS > VSO, whereas in WALS the orders from most to least frequent are SOV > SVO > VSO > VOS > OVS > OSV. As such, the orders are not exactly the same, though there are some similarities. SOV and SVO are the most common orders in each, and in fact these categories do account for a similar percentage of the data both cross-linguistically and in Karuk clauses, with S-first orders accounting for 88.63% of all languages and 84.06% of all Karuk clauses. All the other orders are infrequent both in Karuk and in WALS, though the frequencies are

Order	N. of occurrences	% of dom. order lgs (n=1187)	% of all lgs (n=1376)
SOV	564	47.51%	40.99%
SVO	488	41.11%	35.47%
VSO	95	8.00%	6.90%
VOS	25	2.11%	1.82%
OVS	11	0.93%	0.80%
OSV	4	0.34%	0.29%
No dom. order	189	–	13.74%
All	1376	–	100%
S first	1052	88.63%	76.45%
Non-S first	135	11.37%	9.81%

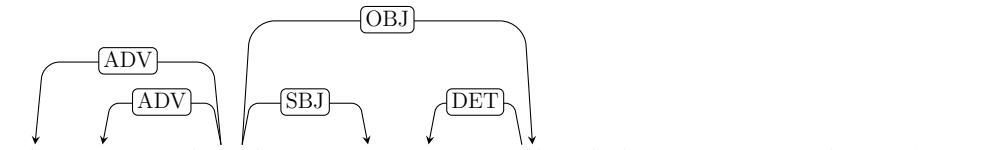
Table 3.8: WALS Feature 81A: Order of Subject, Object, and Verb

different, especially with regards to VSO, which is the least frequent order in Karuk and the third-most frequent in WALS.

If the trends had been more similar overall, we could have surmised that the same pressures driving word order frequency cross-linguistically may be the same pressures driving word order frequency across clauses in a single, no dominant order language. This idea can not be supported overall by this comparison between Karuk and the cross-linguistic facts, though the clear tendency for subject to be first in both datasets is striking. Further in-depth quantitative studies of languages like Karuk in terms of word order freedom would help illuminate this question.

In contrast to earlier work which reported the absence of VSO cases, 1 VSO sentence was discovered by the search. However, examination of the sentence reveals that this discrepancy is due to the way that the often ambiguous 3rd singular pronoun *uum* is annotated. The treebank prefers annotations that treat *uum* as an argument if possible, but a researcher looking for unambiguous examples of VSO order would likely not consider such an example due to the difficulty of understanding the function of *uum* in any given sentence. The ostensible VSO sentence is given below:

(4) vúra vaa u-pakurîihvu-ti uum pa-mu-pákurih haninuvêe naa
INTENS so 3SG>3-sing.songs-DUR 3SG.PRO the-3SG.POSS-song *haninuvêe naa*
hanuvêe naa
haninuvêe naa
'He was singing his song that way, "haninuvêe naa hanuvêe naa."', (WB_KL-07:18)



Note that the predicate in (4) is *upakurîihvuti*, the subject as annotated is *uum*, and the object is *mupákurih*, and they occur in that order. The key consideration here is the interpretation of *uum* as the subject. As noted in the *uum* section of the annotation guidelines (2.9.5), the annotation of *uum* is often difficult, since it is not always clear that any English translations with a pronoun are actually reflecting a given use of *uum*, given the fact that Karuk has frequent pro-drop that would be translated by an English pronoun anyway. To resolve this difficulty, annotating *uum* follows a simple algorithm which leads, in cases where *uum* is the only possible word which can be the subject, to *uum* being annotated as the subject. Note in the case of (4) that this is a perfectly interpretation of the sentence; there is no other noun phrase expressing the subject. An alternative possibility is that *uum* is actually the possessor of *mupákurih* in (4). The annotation guidelines forbid this annotation, since it would violate the principle of ATTACH HIGH (2.2.1) which says, if you have two possibilities, choose the one that attaches highest, and in this case attaching *uum* as subject to the main predicate is higher than attaching it as aPOSS dependent of that predicate's object. However, *Attach High* is merely a strategy for ensuring that annotation is consistent, not necessarily accurate in terms of the real syntactic structure. As such, this one example of VSO order is not unambiguous, and it is possible that it is not in fact an example of such. As such, the results of the treebank do not necessarily contradict the earlier findings that there were no VSO clauses.

3.4.2 Clauses with subjects and complements

Having discussed clauses with a subject and direct object, I move now to the second most common category of clauses with 2 arguments expressed: clauses with a subject and a complement. It should be emphasized that this information has not been commented on at all in any previous work on Karuk, as opposed to S, V, and O orders which have been discussed in previous work as described above.

Table 3.9 summarizes the data below. C is used to represent complement here. It is important to note that complements here comprise a heterogenous class, composed of a variety of different grammatical phenomena such as complement clauses, locative complements to directional suffixes, and others. For more information on what constitutes a complement, please see the section on complements in the annotation guidelines (2.7.10). Unlike in the S,V,O cases discussed above, S,V,C sentences need not be transitive predicates as defined in this treebank, since transitivity is specifically about whether a direct object can be a

dependent of the verb, and both transitive and intransitive verbs can take complements that are not direct objects.

Clause order	# of occurrences	% of clauses with S and C (n=147)	% of total clauses (n=5352)
S C V	77	52.38%	1.44%
S V C	31	21.09%	0.58%
C V S	27	18.37%	0.50%
C S V	10	6.80%	0.18%
V S C	2	1.36%	0.04%
V C S	0	0.0%	0.0%
S first	108	73.47%	2.02%
Non-S first	39	26.53%	0.73%
Pre-verbal C	114	77.55%	2.13%
Post-verbal C	33	22.45%	0.62%
All	147	100%	2.75%

Table 3.9: Orders of clauses with S and C expressed

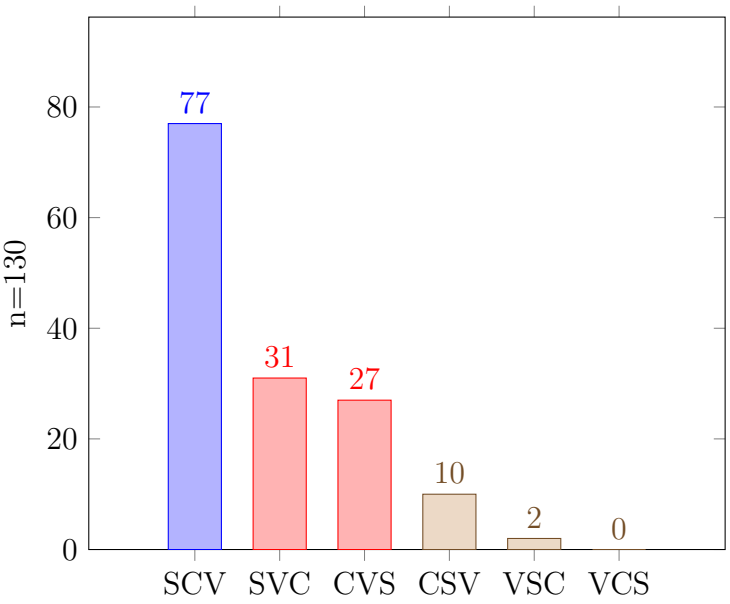


Figure 3.7: Clauses with S and C

Like in the clauses with S and O expressed discussed above, the subject first orders are most common here. However, while SVO was most common there at 45.65% of clauses with

S and O expressed and SOV the second most common at 38.41%, with complements the order SCV is most common at 52.38%, with SVC second most common at 21.09%. So, with objects, the verb-medial category is most common, whereas with complements the verb-final category is most common. Also, with complements, the difference between the two most common categories is more pronounced.

The third-most common here is CVS, the counterpart of OVS, which likewise is the third-most common of clauses with subjects and objects. However, CVS clauses make up a more significant percentage (18.37%) of clauses with complements and subjects than the OVS clauses did of their respective clauses (6.67%).

The first significant break in frequencies with the subject and complement clause orders is not between subject first and non-subject first orders, as it was with the subject and object clauses, but between SCV and everything else. Another break separates SVC and CVS, the verb-medial orders, from the remaining three.

3.4.3 Clauses with direct objects and complements

Table 3.10 presents the data regarding frequency of orders of direct object (O), complement (C), and verb. Figure 3.9 summarizes the data in a bar graph. Clauses with O and C expressed are not particularly common, with only 48 total examples. Unlike the previous two cases, where there were clear observable subgroupings of orders in terms of frequency, here there is one more common order, CVO, and then a slow trailing off of the other orders, as is clearer visually in Figure 3.9. However, though CVO being most common may make initially make one wonder if a post-verbal object is thus most common, it turns out the three options with pre-verbal objects together make up a slightly higher percentage of clauses than those orders with post-verbal objects. This is similar to the proportion of pre- versus post-verbal objects in the S,V,O category above. Like in every case of two arguments expressed we have looked at thus far, the verb-initial categories are the least frequent.

3.4.4 Clauses with indirect objects and one other argument

In this section I examine those orders which include an indirect object (I) and one other argument, whether object or subject. Both objects and subjects are included in a single table here because there is in fact only one extant example of a clause with an expressed subject and indirect object, with SIV order. There is of course also the logical possibility of a clause with an expressed indirect object and complement, but no such clauses were found. Unattested orders with subject or complement are thus left out of the table, for interests of presentation. Table 3.11 reports the data. Such sentences are quite uncommon, with only 6 examples out of the 5352 clauses in the corpus; of course, ditransitives at all are uncommon, and so is expressing arguments at all. However, note that of only 47 ditransitives in the corpus, 6 of them, nearly 13%, have an expressed indirect object in addition to one other expressed argument. This is a higher percentage than, for example, what clauses with subjects and objects expressed make of transitive clauses overall, 8.2%.Because of the low

Clause order	# of occurrences (n=54)	% of clauses with O and C (n=5352)	% of total clauses
C V O	23	42.59%	0.43%
O C V	12	22.22%	0.22%
C O V	9	16.67%	0.17%
O V C	8	14.81%	0.15%
V O C	2	3.70%	0.04%
V C O	0	0.0%	0.0%
Pre-verbal C	44	81.48%	0.82%
Post-verbal C	10	18.52%	0.19%
Pre-verbal O	29	53.70%	0.54%
Post-verbal O	25	46.30%	0.47%
All	54	100%	1.01%

Table 3.10: O, C, and V orders for clauses with 2+ arguments

numbers, it is hard to know whether the patterns here are robust, but, like in the other cases, the verb-initial orders are among the most uncommon; here, they are completely unattested. It is also interesting to note that more of the cases here have an indirect object preceding a direct object, though with the small number of examples it is not really possible to claim this is a robust tendency.

Clause order	# of occurrences	% of ditransitive clauses (n=47)	% of total clauses (n=5352)
I V O	3	6.38%	0.06%
I O V	1	2.13%	0.02%
O V I	1	2.13%	0.02%
S I V	1	2.13%	0.02%
O I V	0	0.0%	0.0%
V I O	0	0.0%	0.0%
V O I	0	0.0%	0.0%
I before O	4	8.51%	0.07%
O before I	1	2.13%	0.02%
All	6	12.77%	0.11%

Table 3.11: O, I, and V orders for clauses with 2+ arguments

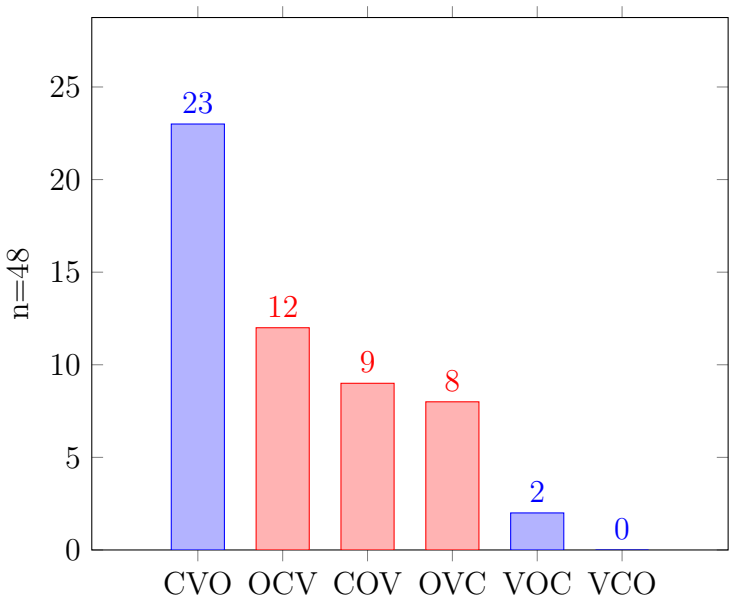


Figure 3.8: Clauses with O and C

3.4.5 Three arguments present

In addition to the examples just discussed of sentences with two expressed arguments, there are seven sentences in the corpus that include three arguments. Five of these sentences have subject, object, and complement expressed, and two of them have subject, object, and indirect object expressed. No other combinations of three and no sentences with all four types of arguments were found. Table 3.12 lists the seven sentences and gives their order and identification number. Because there are so few of these, the typical statistical presentation given for the two argument orders above is eschewed. Rather, in this section I discuss some more particular properties of these sentences. Until the present work, there was no examination of sentences like these, perhaps due to their extreme rarity. Thus, as the treebank allows for one to get a handle on large-scale statistical facts about the corpus, it also allows one to find needle-in-a-haystick examples like the following, which might otherwise have escaped notice.

As stated above, there are no sentences which include all of a subject, direct object, indirect object, and complement, which likely results from there being no verbs which have a subcategorization frame including all of these elements. There are also no sentences in which both a complement and indirect object are expressed along with either a subject or object and in fact there are no sentences with C and I as the only expressed arguments at all. Presumably this absence is also due to the lack of verbs which subcategorize for both an indirect object and a complement at once.

It is striking, though unsurprising, that there are so few examples. Despite there being

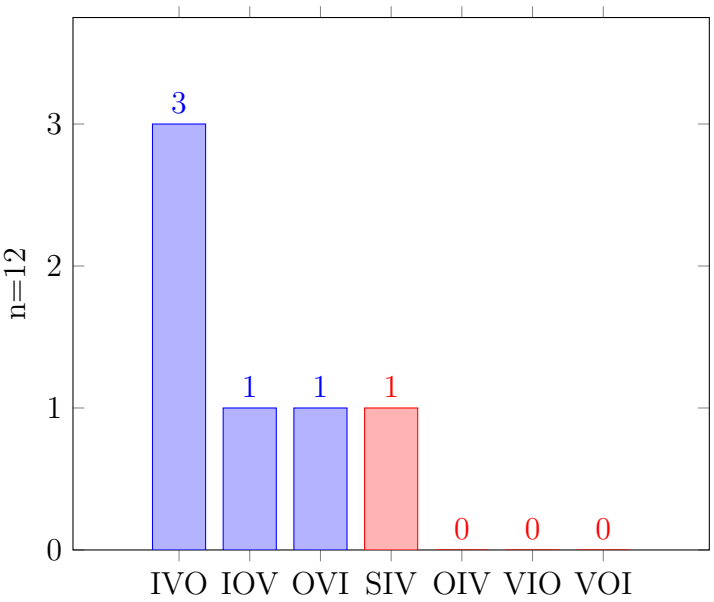


Figure 3.9: Clauses with S or O and I

Clause order	Id number of sentence
S V O C	WB_KL-21:32
S O V C	WB_KL-21:35
C S O V	WB_KL-81:11
C O V S	WB_KL-68:14
O S C V	WB_KL-73:10
I V O S	WB_KL-67:2
O V I S	WB_KL-37:2

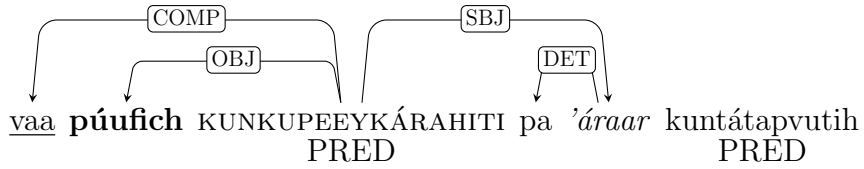
Table 3.12: Sentences with three expressed arguments

24 possible orders of C, S, O, and V or I, S, O, and V, only 5 and 2 of these possible orders respectively are attested. Further, none of the orders are doubled, and each sentence with three arguments has a different order from the others. Due to the rarity of clauses with 2 arguments overall, it is not surprising that clauses with 3 expressed arguments are this rare, and the non-existence of various orders may be merely an effect of this rarity combined with the relatively small size of the treebank.

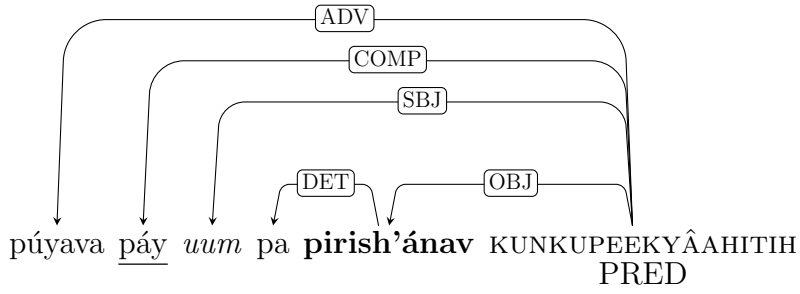
There are some patterns in this data worth examining, despite the rarity of examples. The two examples where the complement is first in the sentence involve cases where the complement in question is a manner complement either of the verb *kuupha* 'to do' (5) or of the manner applicative circumfix *kupa-...-ahi* (6). The complement is in underlined, the

subject is in *italics*, the object is in **boldface**, and the verb is in SMALL CAPS. A further similarity between these two is that the complements are demonstratives, *pay* 'this' and *vaa* 'that,' which are often used as manner complements.

(5) vaa **púufich** KUN-KUPEE-YKÁR-AHI-TI pa-*'áraar* kun-tátapvu-tih
so deer 3PL>3-MODAL-beat-MODAL-DUR the-human 3PL>3-trap-DUR
'The Indians killed deer that way, they trapped them.', (COVS order, WB_KL-68:14)

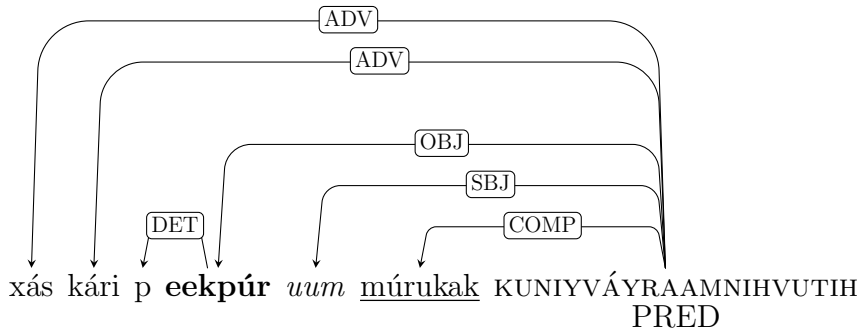


(6) púyava páy *uum* pa-**pirish-'ánav** KUN-KUPEE-KYÂA-HI-TIH
you.see this 3SG.PRO the-plant-medicine 3PL>3-MODAL-make-MODAL-DUR
'That's how they made plant medicine.', (CSOV order, WB_KL-81:11)



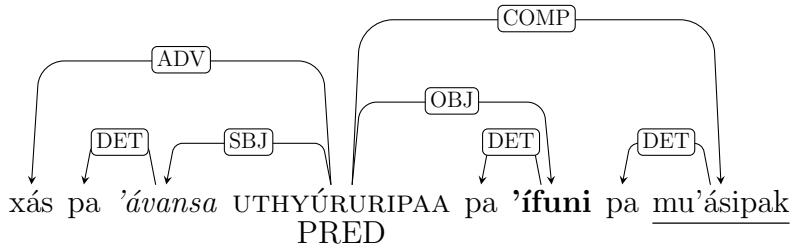
The other three examples all involve complements which are introduced by directional suffixes attached to the verb. Though two of them have post-verbal complements and one has a pre-verbal complement, all of them have the complement as the nearest argument to the verb. In (7), the complement is *múrukak* 'tray basket' and is a complement by virtue of the directional applicative suffix *-raamnih* 'in' on the verb. In this case, the complement is immediately prior to the verb.

(7) xás kári p-**eekpúr** *uum* múruk-ak KUN-IYVÁY-RAAMNIH-VU-TIH
then then the-acorn.flour 3.SG.PRO tray.basket-LOC 3PL>3-pour-in-PL.ACT-DUR
'Then they poured the flour into a tray-basket.', (WB_KL-73:10)

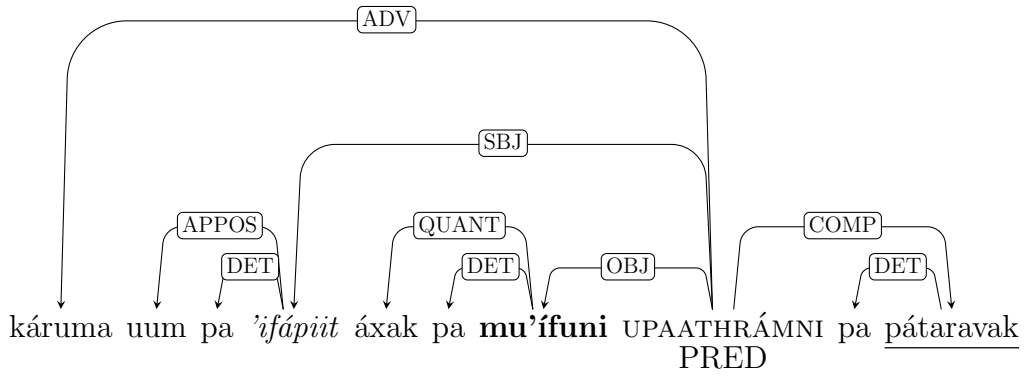


In the two sentences below, the complements *mu'ásipak* 'in his bowl' (8) and *pátaravak* 'in the soup basket' (9) immediately follow the verb, and they are introduced by the applicative suffixes *-ripaa* 'out' and *-ramnih* 'in' respectively.

(8) xás pa-*'ávansa* U-THYÚRU-RIPAA pa-**'ífuni** pa-mu-'ásip-ak
then the-man 3SG>3-pull-out the-hair the-3SG.POSS-bowl-LOC
'And the man pulled the hair from his basket.', (WB_KL-21:32)



(9) káruma uum pa-*'ifápiit* áxak pa-**mu-'ífuni** U-PAATH-RÁMNI
in.fact 3SG.PRO the-young.woman two the-3SG.POSS-hair 3sg(>3)-throw-into
pa-pátarav-ak
the-soup.basket-LOC
'The fact was, the young women had thrown two of their hairs into the soup-baskets.', (WB_KL-21:35)



There is no guarantee that these commonalities reflect the only possibilities for the position of these types of complements, but it is striking that, even with such few examples, these patterns emerge.

3.5 Clauses with a single expressed argument

In this section, I discuss the results of a treebank search targeting clauses with only a single expressed argument. This includes both transitive and intransitive clauses, since a transitive clause can have only one of its arguments expressed. The table below summarizes the results. Note that P for Predicate is used, rather than V for Verb as before, because these sentences include those with non-verbal predicates.

Clause order	# of occurrences	% of 1 arg. clauses (n=2308)	% of total clauses (n=5352)
S P	915	39.64%	17.10%
P S	291	12.61%	5.44%
O P	407	17.63%	7.60%
P O	267	11.57%	4.99%
C P	316	13.69%	5.90%
P C	108	4.68%	2.02%
I P	1	0.04%	0.02%
P I	3	0.13%	0.06%
P or V first	669	28.99%	12.50%
P or V second	1639	71.01%	30.62%
All	2308	100%	43.12%

Table 3.13: Relative prevalence of clauses with 1 argument

What jumps out from this data is that the number of clauses with a single argument is far greater than those with two expressed arguments: 43.12% of clauses in the corpus have

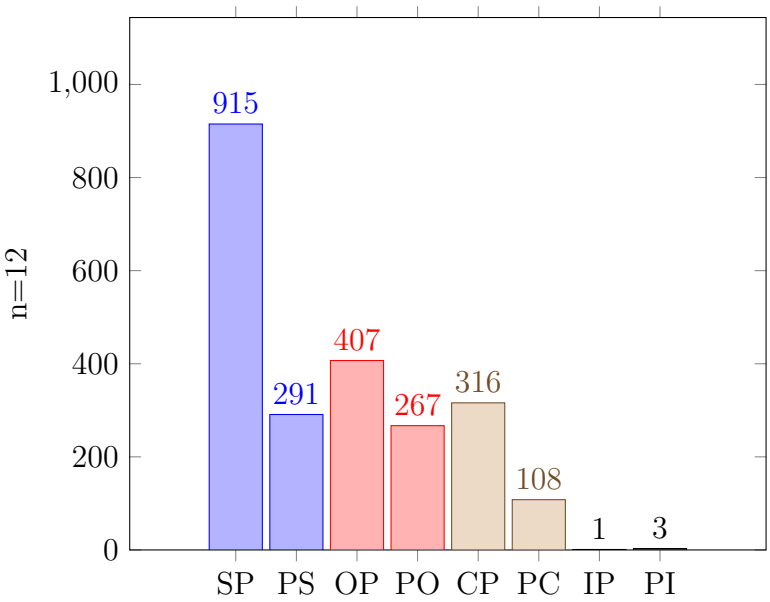


Figure 3.10: Clauses with a single expressed argument

a single argument expressed, versus only 6.52% of clauses in the corpus which have two or more arguments expressed. Of course, the pool of sentences that could potentially have 2 arguments is lower than the pool which could have one, since a large portion of sentences with two arguments (those with direct or indirect objects) are necessarily transitive. However, given that sentences with transitive verbs make up around 31% of the corpus, that difference alone cannot account for the discrepancy in argument expression. Strikingly, the majority of clauses in the corpus lack any argument. Table 3.14 compares the number of clauses with 2 or more, 1, and no arguments. Clauses with no expressed arguments make up 50% of clauses in the corpus.

Category	# of occurrences	% of total (n=5352)
2+ arg.	349	6.52%
1 arg.	2308	43.12%
No arg.	2695	50.36%

Table 3.14: Prevalences of clauses by number of expressed arguments

Of further interest is that of the clauses with a single argument, 71% have the verb or predicate appearing after the argument, driven mainly by the subject and object orders, which both have a majority of their examples with the argument first and predicate second. That distribution matches well with Davis et al. (2020)'s argument that Karuk is a verb-final

language. The only counterpoint to this is the indirect object orders, but the small number of indirect objects overall means we cannot be sure the pattern here would be maintained if there were a sufficient amount of examples.

3.6 Interactions

I turn now to addressing the second question mentioned in the introduction. What factors influence the distributions described above? This section first addresses the three generalizations from the introduction, going through the Subordinate Argument Suppression Tendency, the Transitive Subject Suppression Tendency, and the Object-Conditioned Pre-Verbal Subject Tendency in order, then moving on to an additional statistical interaction outside of those tendencies.

3.6.1 Subordinate Argument Suppression Tendency

The questions motivating this section are as follows: in terms of word orders, how are subordinate clauses different from main clauses? Do subordinate clauses have the same total range and prevalence of word orders that main clauses do? As previewed in the introduction, we find that, in fact, subordinate clauses differ significantly from main clauses in being less likely to have expressed arguments. I term this the Subordinate Argument Suppression Tendency, provided again below:

(10) Subordinate Argument Suppression Tendency: Subordinate clauses have a tendency to have fewer expressed arguments than main clauses.

Table 3.15 gives the data concerning the interaction of subordination with argument expression generally. In this table, I compare sentences with any argument expressed, whether a single argument or multiple, to those without any argument, separated on the basis of their main or subordinate clause status. This tabulation allows us to see if subordinate and main clauses are roughly similar in how often they have expressed arguments. If subordinate clauses were not significantly different from main clauses in their tendency to have arguments expressed, we should see a similar percentage of subordinate clauses with arguments as we see main clauses with arguments. This is the null hypothesis in this case. However, subordinate and main clauses do differ: only 37.19% of subordinate clauses have arguments compared to 50.67% of main clauses, and a Chi Square test of independence performed on this data confirms that this difference is indeed statistically significant. $X^2(1,N{=}5352) = 26.72$, $p = 0.0000002$. If the null hypothesis were true, we would have expected there to be something nearer to 201 subordinate clauses with arguments and 204 subordinate clauses without arguments, as opposed to the observed 151 and 255.

From this result, we know that subordinate clauses thus are less likely in general to have arguments than main clauses.[6] However, there are a multitude of ways in which this could

[6]Line Mikkelsen (p.c.) brought to my attention that this is potentially a result of the principle of Attach

		Sentence with argument(s)	Sentence without argument
Main		2506	2440
	% of total Main (n=4946)	50.67%	49.33%
Sub		151	255
	% of total Sub (n=406)	37.19%	62.81%
All		2657	2695
	% of total clauses (n=5352)	49.64%	50.36%

Table 3.15: Interaction between argument expression and subordination

arise. Is it that all arguments are equally unlikely with subordinate clauses, or just that, for example, subjects are especially unlikely and, due to there being a lot of subjects, it skews the data for the whole dataset? To investigate this, I look at the cases of sentences with a single expressed argument, comparing each set of sentences with a particular argument to the rest, to see if any particular argument is more significantly skewed against being subordinate than any of the others. Before doing this, however, we should be sure that clauses with a single argument exhibit the Subordinate Argument Suppression Tendency, to ensure that the effect is not due to the tendency merely being strong in other sentence types. Table 3.16 presents the relevant data. Clauses with a single argument are compared to all other sentences in the corpus, including sentences with more than one argument and no argument. The null hypothesis here is that there is no interaction between a single argument being expressed and subordination, which would mean that the distribution of subordinate and main clauses among the single argument clauses should not be significantly different from the distribution of subordinate and main clauses in the corpus overall. However, this turns out not to be the case. A chi-square test of independence indicates a significant relationship between subordination and whether a single argument is expressed. $X^2(1,N=5352) = 17.02$, $p = .00004$. There are less subordinate clauses with a single expressed argument than expected if the null hypothesis was true; the expected number was roughly 175, compared to the

High that guided the treebank annotation. Attach High states that, if a word can in principle be a dependent of a lower or higher head, we should choose the higher. This does mean that, if a noun phrase was ambigious between being the subject of a subordinate verb and the main verb that the annotators would make it a dependent of the main verb, potentially depriving subordinate clauses of arguments that might in truth belong with the subordinate verb. This is a potential confound, but one that I think would be too rare to have a significant effect, as the position of the noun phrase in question would have to be such that it was equally likely as a main clause and subordinate clause argument. However, any future work attempting to explain the Subordinate Argument Suppression Tendency should endeavour to investigate the extent to which Attach High is responsible for the tendency. As I am not attempting to explain the tendency for reasons of space here, I leave this avenue for future research.

observed 135.

		Sentences with 1 arg.	Other sentences
Main		2173	2773
	% of total main (n=4946)	43.93%	56.07%
Sub		135	271
	% of total sub. (n=406)	33.25%	66.75%
All		2308	3044
	% of total clauses (n=5352)	43.12%	56.88%

Table 3.16: Interaction between presence of a single argument and subordination

This result tracks with the result of looking at sentences with any number of argument, and shows that the Subordinate Argument Suppression Tendency holds for the subset of sentences with a single argument that I now turn to investigate in more depth.

The following tables look at each argument in turn, to see if any particular argument is less likely with subordinate clauses than the others. This includes subjects, direct objects, and complements; indirect objects are excluded due to their low number of occurrences. I find that no one argument is more unlikely than the others in subordinate clauses. Table 3.17 presents the data relating to the subject, comparing sentences with the subject as the sole argument to other clauses with a sole argument that isn't the subject. A chi-square test of independence on this data does not indicate a significant relationship. $X^2(1,N=2308) = 0.27$, $p = .60$.

		Sentences with S	Sentences with other single arg.
Main		1132	1041
	% of total main (n=2173)	52.09%	47.91%
Sub		74	61
	% of total sub. (n=135)	54.81%	45.19%
All		1206	1102
	% of total clauses (n=2308)	52.25%	47.75%

Table 3.17: Interaction between presence of subject as sole argument and subordination

Table 3.18 below presents the data relating to the object, comparing sentences with the direct object as a sole argument to all other clauses with a single argument. A chi-square test of independence on this data likewise does not indicate a significant relationship. $X^2(1,N=2308) = .92$, $p = .34$.

Table 3.19 presents the data relating to complements, comparing sentences with a complement as the sole argument to all other clauses with a single argument. A chi-square test of in-

		Sentences with O	Sentences with other single arg.
Main		640	1533
	% of total main (n=2173)	29.45%	70.55%
Sub		34	101
	% of total sub. (n=135)	25.19%	74.81%
All		674	1634
	% of total clauses (n=2308)	29.20%	70.80%

Table 3.18: Interaction between presence of object as sole argument and subordination

dependence on this data likewise does not indicate a significant relationship. $X^2(1,N=2308) = 0.15$, $p = .70$.

		Sentences with C	Sentences with other single arg.
Main		397	1776
	% of total main (n=2173)	18.27%	81.73%
Sub		27	108
	% of total sub. (n=135)	20.00%	80.00%
All		424	1884
	% of total clauses (n=2308)	18.37%	81.63%

Table 3.19: Interaction between presence of complement as sole argument and subordination

As none of these investigations have revealed a significant difference between any one argument and the others in terms of how often they are subordinate, we can infer that the Subordinate Argument Suppression Tendency is a general tendency affecting all arguments, and not an artifact of one particular common argument, like the subject, being uniquely unlikely in subordinate clauses.

This picture is complicated slightly by turning our investigations to sentences with two arguments expressed. We may expect, given the prior discussion, that every combination of two arguments is roughly as unlikely in subordinate clauses as any other combination. However, this turns out not to be the case, with only sentences containing the S+O combination to show the Subordinate Argument Suppression Tendency. Combinations of complements with either subject or direct object do not exhibit the tendency.

Table 3.20 below gives the raw numbers of how many sentences that include S and O are main or subordinate versus those sentences which do not include S and O (but may include other combinations of arguments.) The question here is whether there is a significant interaction between the presence of subject and object in a clause and whether the clause is a main or subordinate clause. In other words, is it more likely than expected for a main clause to have subject and object expressed than a subordinate clause or vice versa? Indeed,

a chi-square test of independence indicates there is a significant interaction. $X^2(1,N=5352) = 4.93$, $p = .026$. Specifically, the chi-square test indicates that the expected amount of subordinate S and O clauses is roughly 10, as opposed to the observed 3.

		Sentence with S and O	Sentence without S and O
Main		132	4814
	% of total Main (n=4946)	2.67%	97.33%
Sub		3	403
	% of total Sub (n=406)	0.74%	99.26%
All		135	5217
	% of total clauses (n=5352)	2.52%	97.48%

Table 3.20: Interaction between presence of S+O and subordination

Table 3.21 breaks the data down further by clause order, showing which orders of S, V, and O had which numbers of main versus subordinate clauses. Unsurprisingly, only the most common orders, SVO and SOV, have subordinate clauses. Subordinate clauses being rare overall, it is more likely for them to occur in a dataset with more sentences.

Clause order	# of occurrences	# main	# sub.
S V O	61	60	1
S O V	52	50	2
O V S	9	9	0
O S V	8	8	0
V O S	4	4	0
V S O	1	1	0
All	135	132	3

Table 3.21: Subordination and clause order of subject, verb, object

Table 3.22 presents the data related to combinations of subject and complement and their potential interaction with subordination. A Chi Square test of independence on this data indicated no detected significance. $X^2(1,N=5352) = 0.04$, $p = .84$. As such, even though subordinate clauses with S and C expressed are very rare, with only 10 occurrences over all, that rarity is not more than expected given the overall rarity of S and C clauses and subordinate clauses respectively. This lack of an interaction is notable concerning there was a significant interaction between S and O expression and subordination, as described above. For whatever reason, the Subordinate Argument Suppression Tendency does not seem to apply in this case.

Table 3.23 breaks the data down into clause order categories. As expected, the most common category, SCV, likewise has the most subordinate examples.

		Sentences with S and C	Sentence without S and C
Main		137	4809
	% of total main (n=4946)	2.77%	97.23%
Sub		10	396
	% of total sub. (n=406)	2.46%	97.54%
All		147	5205
	% of total clauses (n=5352)	2.75%	97.25%

Table 3.22: Interaction between presence of S+C and subordination

Clause order	# of occurrences	# main	# sub.
S C V	77	71	6
S V C	31	29	2
C V S	27	27	0
C S V	10	9	1
V S C	2	1	1
V C S	0	0	0
All	147	137	10

Table 3.23: Subordination and clause order of subject, verb, complement

Table 3.24 presents the data relevant to the combination of direct object and complement and their potential interaction with subordination. As with S+C combinations, there is no apparent significant interaction between the two variables per a chi-square test of independence based on the data in Table 3.24. $X^2(1,N=5305) = 0.09$, $p = .76$. As with subjects and complements, it appears that the Subordinate Argument Suppression Tendency does not hold of clauses that specifically have the combination of objects and complements.

		Sentences with O and C	Sentence without O and C
Main		51	4895
	% of total main (n=4946)	1.03%	98.97%
Sub		3	403
	% of total Sub (n=406)	0.74%	99.26%
All		54	5298
	% of total clauses (n=5352)	1.01%	98.99%

Table 3.24: Interaction between presence of S+C and subordination

Table 3.25 gives the breakdown of this data by specific order. Unsurprisingly, the rare

subordinate clause cases are found in two of the top three most common orders.

Clause order	# of occurrences	# main	# sub.
C V O	23	21	2
O C V	12	12	0
C O V	9	78	1
O V C	8	8	0
V O C	2	2	0
V C O	0	0	0
All	48	45	3

Table 3.25: Subordination and clause order of subject, verb, complement

As for clauses with indirect objects, it turns out that all of the clauses with an indirect object and one other expressed argument are main clauses. Because of the small number of these cases and the rarity of subordinate clauses overall, one cannot make a strong claim that this is due to an actual restriction, and we must await further data to come to any more sure-footed conclusions.

To briefly summarize, we have found that, across the whole corpus and with clauses with a single argument, there is a clear tendency for subordinate clauses to be less likely than main clauses to have any expressed arguments. This tendency also holds for the most common combination of two arguments, subject and direct object, but does not seem to hold, surprisingly, for sentences with a combination of a complement and subject or direct object.

3.6.2 Transitive Subject Suppression Tendency

The question motivating this section is as follows: do transitive clauses exhibit similar rates of argument expression as intransitive clauses? We know, of course, that only transitives are able to include direct and indirect objects (if ditransitive), but what about for word orders involving subjects and complements? I find that subjects are less likely than expected in transitive clauses, given their distribution in intransitive clauses. Complements, however, exhibit a bit more complicated pattern.

Table 3.26 summarizes the data relating to the potential interaction between transitivity and whether S is expressed or not. In this case, I compare the set of all sentences with an expressed S to the set of sentences that have no expressed arguments, with each set subgrouped on the basis of transitivity. I compare to the set of sentences without arguments as opposed to, say, all sentences without an expressed S, as I expect that the sentences with no arguments are likeliest to have the 'basic' distribution of transitives and intransitives, unaffected by any influence other arguments may exert.[7]

[7]A chi-square test was performed also on data that included all sentences, splitting them by S being

		S expressed	No argument expressed
Intransitive		1224	2110
	% of total intrans. (n=3334)	36.71%	63.29%
Transitive (monotrans and ditrans)		272	585
	% of total trans. (n=857)	31.74%	68.26%
All		1496	2695
	% of total clauses (n=4191)	35.70%	64.30%

Table 3.26: Interaction of transitivity and whether S is expressed

The null hypothesis of no interaction here would mean that sentences with S expressed showcase a statistically similar distribution of transitives and intransitives to the sentences with no argument. However, a chi square test of independence performed on the data is consistent with a rejection of the null hypothesis. $X^2(1,N=4191) = 7.13$, $p = .007$. Specifically, there are less transitive clauses with expressed S than expected given the distribution of transitivity across this data. Roughly 306 transitive clauses with S were expected, as opposed to the observed 272. This shows the tendency for transitives to have fewer expressed subjects, which I term the Transitive Subject Suppression Tendency.

(11) Transitive Subject Suppression Tendency: Subjects are less likely to be expressed in transitive clauses than in intransitive clauses.

Table 3.27 presents a similar summary related to complement expression and transitivity. As with the subject expression data, I compare sentences with an expressed complement, including sentences with other expressed arguments, to sentences with no expressed arguments.

A Chi Square test of independence performed on this data returns a significant result, but a surprising one. In fact, there are more transitives with expressed complements than expected, rather than less, as with subjects. $X^2(1,N=3325) = 15.73$, $p = .00007$. If there was no such skew, we would have expected roughly around 484 intransitive clauses with complements and around 145 transitive clauses with complements, as opposed to the observed 446 and 184 respectively.

I turn now to investigating whether these patterns hold for smaller subtypes of sentences with subjects and complements, first looking at clauses with both a subject and complement

expressed or not, but it gave a dramatic significant result: $X^2(1,N=5352) = 172.43$, p ¡ .000001. Given that I expect the presence of other arguments is likely to influence this, it seemed the data here might be unduly skewed and thus lead to producing an effect to this extent. As such, I think the comparison to the clauses with no arguments to be a stronger one.

		C expressed	No argument expressed
Intransitive		446	2110
	% of total intrans. (n=2556)	17.45%	82.55%
Transitive (monotrans and ditrans)		184	585
	% of total trans. (n=769)	23.93%	76.07%
All		630	2695
	% of total clauses (n=3325)	18.95%	81.05%

Table 3.27: Interaction of transitivity and whether C is expressed

expressed. Drilling down to these subcases is important to discover any possible skewing, just in case it is only a particularly common subcase which exhibits the tendency and not all of them. Table 3.28 summarizes data comparing transitivity in sentences with S and C expressed versus all sentences without S and C expressed together (but which can contain other combinations of arguments.) For this table, the category of transitive includes only monotransitives, since no ditransitives with S and C expressed were found. A chi square test on this data indicates a significant relation between transitivity and S and C expression. $X^2(1,N=5352) = 15.62$, $p = 0.00008$. The expected number of transitive S and C sentences, if there had not been an interaction, was roughly 46, as opposed to the observed 24.

So, transitive sentences with both a sentence and complement expressed are far rarer than expected just from the independent rarity of S and C sentences and transitive sentences respectively. This matches with the prior observation that subjects are rarer with transitives, but not with the observation that complements are actually more likely.

		Sentences with S and C	Sentence without S and C
Intrans.		123	3537
	% of total intrans. (n=3660)	3.36%	96.64%
Trans.		24	1668
	% of total trans. (n=1692)	1.42%	98.58%
All		147	5205
	% of total clauses (n=5352)	2.75%	97.25%

Table 3.28: Interaction between presence of S+C and transitivity

Table 3.29 breaks the data down by specific clause order. Unsurprisingly, the most common order has most of the transitive cases, and the distribution of transitives is roughly similar to the distribution of clause orders.

Clause order	# of occurrences	# intrans.	# trans.
S C V	77	64	13
S V C	31	28	3
C V S	27	22	5
C S V	10	8	2
V S C	2	1	1
V C S	0	0	0
All	147	123	24

Table 3.29: Transitivity and clause order of subject, verb, complement

I now move on to looking just at sentences with only a single argument, either S or C, expressed. This differs from the earlier data relating to S- and C-expression in that those datasets included all sentences with S or C respectively, not just those where S or C was the sole argument Table 3.30 presents the relevant data for sentences with the subject as the sole argument, and Table 3.31 does so for sentences with a complement as the sole argument. Sentence types that must be transitive, meaning those with O or I as the only expressed argument, are not considered here as above, since we can be sure already that there is a relationship between transitivity and direct/indirect object expression. As above, these cases are compared to sentences with no expressed argument, on the assumption that such sentences are neutral in terms of their split between intransitives and transitives. Because of the small amount of ditransitives, they are not separated from the transitive subset but included in it.

		Sentences with S as sole arg.	No argument
Intrans.		1098	2110
	% of total intrans. (n=3208)	34.23%	65.77%
Trans.		108	585
	% of total trans. (n=693)	15.58%	84.42%
All		1206	2695
	% of total clauses (n=3901)	30.92%	69.08%

Table 3.30: Interaction between presence of a single argument and transitivity

For subjects, a Chi Square test of independence indicates that there is a significant relationship, with there being far less subjects of transitives than expected given the typical

distribution of transitives and intransitives. $X^2(1,N=3901) = 91.86$, $p < 0.00001$. If there was a similar distribution of transitives in sentences with a sole S as argument as there are in sentences with no argument, we would have expected there to be 214 transitive sentences with sole S as opposed to the observed 108. This again exhibits the Transitive Subject Suppression Tendency.

		Sentences with C as sole arg.	No argument
Intrans.		323	2110
	% of total intrans. (n=2433)	13.28%	86.72%
Trans.		100	585
	% of total trans. (n=685)	14.60%	85.40%
All		423	2695
	% of total clauses (n=3118)	13.57%	86.43%

Table 3.31: Interaction between presence of C as sole argument and transitivity

For complements, on the other hand, there does not seem to be any relationship with transitivity in clauses with only a single argument. A Chi Square test of independence does not find a significant interaction between expression of C and transitivity. $X^2(1,N=3118) = 0.69$, $p = 0.41$. This is unlike the previous finding, where sentences with complements in general were slightly more likely to be transitive than sentences without any argument were. In this case, the distribution of transitive and intransitive in sentences with a complement as the only argument was similar to the distribution of transitivity in sentences with no argument. As such, it does not appear that there is really a tendency for clauses with complements to be transitive, and we can also be sure that the Transitive Subject Suppression Tendency is indeed only about the (transitive) subject.

The Transitive Subject Suppression Tendency bears resemblance to the central phenomenon discussed in Du Bois (1987), where, in Sacapultec Maya, subjects of transitives are rarer than expected. Sacapultec, like Karuk, is a heavily pro-drop language, and it is rare that any transitive clause will have both subject and object expressed: only 2.8% of transitive clauses in Du Bois' sample had both arguments expressed. Karuk, in comparison, has both arguments expressed in 13.36% of transitive clauses.[8] In addition, Du Bois found that only 2.9% of expressed nominals were subjects of transitives, with subjects of intransitives and direct objects accounting for 32.8% and 21.1% of expressed nominals respectively. Du Bois' account for this utilizes a combination of various grammatical and pragmatic constraints. I simplify the discussion here, but essentially he argues that, in general, it is preferable that clauses have only one expressed argument, that only one argument is new information, and that subjects of transitives are preferably given information, rather than new information; he calls this later point the Given A Constraint, A being the name of the subject of transitive position. Additionally, there is a tendency for new information to

[8]There are 226 transitive sentences with two arguments expressed out of 1692 total transitives.

be expressed as a full NP. Taken all together, if a clause preferably has only one argument, and that argument is preferably new information, transitive subjects are less likely to be expressed than other arguments since they are typically not new but given information, and the only 'slot' for an argument to be expressed is more likely to be filled by new information. The Karuk treebank does not include annotation of new and given status for NPs, and as such data exactly comparable to Du Bois' can not currently be obtained from the treebank. Nevertheless, the similarity is tempting.

There is potentially one wrinkle for the Du Bois view in Karuk. As described above, the most common word orders of transitives with two expressed arguments are those in which the subject is pre-verbal and occurs before other arguments. This pre-verbal position is generally where new information is located in Karuk, with given information being unexpressed or post-verbal (Mikkelsen 2014). Yu (2021) has investigated correlations between pre-verbal and post-verbal order of noun phrases and their referential distance (i.e. the distance in clauses between their last mention and the current mention) and found that, for subjects especially, higher referential distance correlated with preverbal mention, and lower referential distance with postverbal mention. Higher referential distance, meaning more clauses between mentions, suggests that the noun in question is not as activated or given as a noun with lower referential distance, so Yu's findings could be seen as supporting the idea of newer information being preferentially pre-verbal. If the Given A Constraint held in Karuk, why would the subject be able to hold the initial position in greater numbers than the object? Perhaps the solution to this wrinkle lies in the rarity of sentences with two expressed arguments, in that those sentences are maybe used when whatever pragmatic factors have colluded to require a new mention in subject position, and are thus rare because it is a pragmatically dispreferable situation. The treebank does not involve pragmatic annotation of the type necessary, so I must leave the question unanswered here.

3.6.3 Object-Conditioned Pre-Verbal Subject Tendency

For the final set of interactions, I look at the effects the presence of one argument can have on the position of another. Specifically, I investigate the interaction of subject position and object expression, object position and subject expression, and subject or object position with complement expression. I finish by discussing the interactions of transitivity and subject position. I find that there is a pressure exerted by the presence of an object on the position of the subject that pushes subjects to be pre-verbal when an object is present. I term this the Object-Conditioned Pre-Verbal Subject Tendency.

(12) Object-Conditioned Pre-Verbal Subject Tendency: If a sentence has an expressed subject and object, the subject is more likely to be pre-verbal than if the object was not expressed.

Table 3.32 summarizes the data relating to the interaction of the position of the subject and whether there is an object expressed. In this case, all sentences with subjects expressed,

regardless of whether they also express other arguments, are examined, and split into subsets on the basis of whether the subject occurs before or after the predicate and whether or not an object is expressed. The null hypothesis here is that the presence or absence of the direct object would have no effect on the position of the subject, and as such we should see a similar distribution of pre- and post-predicate subjects across sentences with an object expressed and not expressed. However, a chi square test of independence performed using this data indicates there is a significant relationship and is not consistent with the null hypothesis. $X^2(1,N=1496) = 9.36$, $p = .002$. If there were no relationship, it is expected there would be a higher number of post-predicate S with an object expressed; specifically, roughly 32 examples are expected, versus the observed 17. In other words, post-predicate S make up a higher percentage of cases with no object expressed than they do of cases with an object expressed. Expressing an object thus makes it even less likely for S to be expressed post-verbally.

		Object expressed	No object expressed
Pre-predicate S		125	1034
	% of total pre-pred. (n=1159)	10.79%	89.21%
Post-predicate S		17	320
	% of total post-pred. (n=337)	5.04%	94.96%
All		142	1354
	% of total clauses (n=1496)	9.49%	90.51%

Table 3.32: Interaction of pre- or post-predicate position of S and whether an object is expressed

However, there does not appear to be an analogous relationship between object position and subject expression. Table 3.33 summarizes the data relating to a the interaction of the position of the object and whether the subject is expressed. All sentences with objects expressed are examined, including those which include other arguments like complements. These sentences are subgrouped based on the pre- or post-verbal position of the object and whether a subject is expressed. The null hypothesis here, that there is no relationship between these two variables, would mean that the distribution of pre and post-verbal objects was not significantly different in sentences with a subject from sentences without a subject. A chi square test of independence performed using this data does not indicate a significant relationship (which would require a p value of .05), though it is close. $X^2(1,N=875) = 2.55$, $p = .109$. There is a slight skew visible in the data: post-predicate objects occur with expressed subjects 18.7% of the time, as opposed to pre-predicate Os only appearing 14.5% of the time. As such, objects in the treebank are very slightly more likely to be post-verbal if a subject is present, but this slight likelihood does not raise to the level of statistical significance.

Perhaps this slight likelihood, taken together with the statistically significant likelihood of subjects being pre-predicate if an object is present, may be what is pushing sentences with subjects and objects present to be SVO slightly more often than SOV, however.

		Subject expressed	No subject expressed
Pre-predicate O		74	438
	% of total pre-pred. (n=512)	14.45%	85.55%
Post-predicate O		68	295
	% of total post-pred. (n=363)	18.73%	81.27%
All		142	733
	% of total clauses (n=875)	16.23%	83.77%

Table 3.33: Interaction of pre- or post-predicate position of O and whether S is expressed

Is the effect on the subject's position specific to the presence of the object, or do complements also get involved? I find that complements do not interact significantly with subject and objects in this way, and the presence of a complement seems to have no effect on the distribution of subject or object position. Table 3.34 summarizes the data relating to the potential interaction between subject position and whether a complement is expressed or not. Again, all sentences with subjects expressed are considered here, including those with more than one argument, and these are subgrouped on the basis of the pre- or post-predicate position of S and whether a complement is expressed or not. The null hypothesis of no interaction here would indicate that the distribution of pre- and post-predicate subjects would not be significantly different in the sentences with complements compared to sentences without complements. In other words, expressing a complement would not add any extra bias for a particular subject position. A chi square test of independence performed on the data indicated the data is consistent with the null hypothesis of no significant relationship. $X^2(1,N=1496) = 0.58$, $p = .44$. As such, it seems that, unlike with objects, complements do not influence subject position.

Likewise, complements do not seem to influence object position. Table 3.35 summarizes the data relating to the potential interaction between object position and whether a complement is expressed or not. Again, any sentence with a direct object expressed is considered here, and subgrouped on the basis of the pre- or post-verbal position of the object and the presence or absence of a complement. The null hypothesis is that there is no relationship, and would mean that the distribution of object position is not significantly different in sentences with an expressed complement versus sentences without an expressed complement. A chi square test of independence performed on the data indicated the data is consistent with the null hypothesis of no significant relationship. $X^2(1,N=875) = 0.0006$, $p = .98$. As

		Complement expressed	No complement expressed
Pre-predicate S		122	1037
	% of total pre-pred. (n=1159)	10.53%	89.47%
Post-predicate S		30	307
	% of total post-pred. (n=337)	8.9%	91.10%
All		152	1344
	% of total clauses (n=1496)	10.16%	89.84%

Table 3.34: Interaction of pre- or post-predicate position of S and whether a complement is expressed

such, the presence of a complement does not seem to influence the position of the object, much as it did not influence the position of the subject. As such, we know that the Object-Conditioned Pre-Verbal Subject Tendency seems really to be about subjects and objects as written above. Taken all together, it seems that the object can influence the subject, but complements are independent of the two. As in the other discussions of interactions here, I will leave the explanation of these discovered facts to future work.

		Complement expressed	No complement expressed
Pre-predicate O		34	479
	% of total pre-pred. (n=513)	6.63%	93.37%
Post-predicate O		25	337
	% of total post-pred. (n=362)	6.91%	93.09%
All		59	816
	% of total clauses (n=875)	6.74%	93.26%

Table 3.35: Interaction of pre- or post-predicate position of O and whether a complement is expressed

An alternative explanation for the Object-Conditioned Pre-Verbal Subject Tendency may be that it is not the presence of the object that pushes the subject to be pre-verbal, but rather just transitivity itself that does. Because we want to know if it is the verb's transitivity

independent of the object or any other arguments' expression that is pushing the subject to be pre-verbal, we need to look at the subset of sentences that: 1) are transitive, 2) do not have any other expressed arguments, and 3) have an expressed S. Table 3.36 summarizes this data, focusing only on sentences where S is the only argument, subdivided by whether S is pre- or post-verbal, and whether the predicate is intransitive or transitive. The key part is that sentences with expressed objects, known to influence the subject, are thus excluded. A chi square test of independence performed on the data is consistent with the null hypothesis. $X^2(1,N=1206) = 1.72$, $p = .19$. With the expressed O sentences excluded, there is no interaction where transitivity pushes S to be pre-verbal more often than expected. Thus, the Object-Conditioned Pre-Verbal Subject Tendency does appear to be specifically about the object's presence exerting pressure on the subject, rather than something more abstract like the transitivity of the verb exerting this pressure.

		Pre-pred. S	Post-pred. S
Intransitive		827	271
	% of total intrans. (n=1098)	75.32%	24.68%
Transitive		88	20
(monotrans and ditrans)	% of total trans. (n=108)	81.48%	18.52%
All		915	291
	% of total clauses (n=1206)	75.87%	24.13%

Table 3.36: Interaction of transitivity and S position

3.6.4 Further interactions

There is one further interaction worth discussing that does not fit into the three generalizations discussed above. Is there any relationship between the verbal or non-verbal status of a predicate and its likelihood to have an expressed argument? As subjects are the only possible arguments of non-verbal clauses, they are the only arguments considered here. The data in Table 3.37 looks at the percentage of clauses with S as the only argument that are verbal and non-verbal, versus the rest of the corpus. A chi-square test of independence indicates there is a significant relationship. $X^2(1,N=5352) = 116.82$, $p < .0001$. There are more non-verbal cases of S only than expected given the overall prevalence of non-verbal clauses; specifically, the expected amount is around 112, opposed to the observed 202.

However, comparing the S-only cases to the entire corpus may give a skewed result, since only S-only and sentences with no expressed arguments can ever have non-verbal predicates. As such, non-verbal predicates are not evenly distributed across the corpus, and a significant portion of them are likely within the S-only category.

		Clauses with S as only arg.	Other clauses
Verbal		997	3856
	% of total Verbal (n=4853)	20.54%	79.46%
Non-verbal		209	290
	% of total Non-verbal (n=499)	41.88%	58.12%
All		1206	4146
	% of total clauses (n=5352)	22.53%	77.47%

Table 3.37: Interaction of expression of S as only argument and verbal or non-verbal predicate status

3.7 Conclusion

In this chapter, I have established core syntactic facts about the Karuk language relating to the order of arguments and predicates. In so doing, I have confirmed the impressions of earlier work and pushed past what earlier work has been able to do. Specifically, I provided a solid quantitative footing for these observations and explored several interactions between word order and syntactic properties such as transitivity and subordination that help to structure the discovered distributions of orders.

The primary goals of this chapter were to provide such facts, and also by doing so to showcase the utility of treebanks, and as such I provide some remarks as to that utility here. One of the key features of the Karuk treebank is the wide variety of syntactic properties that are directly annotated or can be derived from the attributes that are directly annotated. This is shown by the ability to, as described earlier, subgroup the corpus on the basis of transitivity even though transitivity was not directly encoded.

Earlier work which commented on Karuk word order has relied on doing counts of various word orders by hand. These counts can, of course, find the argument and predicate orders and allow for estimations to be made that, as discussed, have been proven mostly correct. However, because they are focused only on those orders, they do not at the same time make counts or annotations of other properties (such as transitivity) that may be relevant to the distribution of those orders. Adding those properties in an ad hoc manner after the fact would then double the annotation time, in that the entire corpus would have to be annotated again for those properties. In distinction to this, the concerted effort to annotate these properties systematically, as required in making a treebank, produces a tool which can then be used to pursue a large variety of questions without the need for further annotation, and can be used to pursue questions which were not actually considered at the time of the treebank's creation. Indeed, this was the case with the various interactions discussed above regarding transitivity. The treebank was not designed to have annotation of every verb for its transitivity, but because it included properties like the annotation of object person and number, it allowed transitivity to be derived.

Chapter 4

Agreement

In the last chapter, I looked at large-scale properties of the Karuk corpus related to argument order, showcasing the treebank's propensity for collecting data relevant to large, statistical questions. In this chapter, I turn to a much smaller phenomenon, measured not in hundreds of sentences but, at most, handfuls: sentences where the agreement prefix used on the verb does not match the features of the relevant arguments in the given sentences. Specifically, this chapter will focus on two particular phenomena, rare but nonetheless robust: unexpected plural agreement, where third-person plural subject agreement markers are used with third-person singular subjects; and the inverse: unexpected singular agreement, where third-person singular subject agreement markers are used with known plural subjects. These sentences have, for the most part, eluded previous work on Karuk's agreement system. To put it idiomatically, these sentences are 'needles in a haystack,' and the treebank is just as useful in finding such sentences as in uncovering larger, corpus-wide properties of sentences as showcased previously. As the primary point of this chapter, as with the previous, is to showcase the utility of the treebank, the discussion of the unexpected agreement phenomena will not delve deeply into theoretical issues, but seek rather to establish the basic empirical facts and potential avenues for further research.

The findings in this chapter arose out of a project to use the treebank to confirm that Bright's description of the Karuk agreement system matched the reality of the agreement system as seen in the corpus that Bright himself collected. The two mismatch phenomena (and some inconsistency with the 3pl>3pl agreement marker) which this chapter describes are the only systematic exceptions that this study found to Bright's description of the agreement system, and as such it can be confirmed that Bright's description of the agreement system was overall accurate.

Before discussing these phenomena, a brief overview of the Karuk agreement system and a discussion of the search methodology are in order.

4.1 The Karuk agreement system

At the basic level, the Karuk system primarily involves the use of prefixes attached only to verbs, though one suffix is also involved. These markers appear to be portmanteaus marking the person and number of both the subject and object. The agreement marking also varies with respect to sentence type, with optative/imperative sentences and negative sentences involving different paradigms from the positive indicative sentences. The earliest published work to attempt a systematic description of the agreement paradigms is de Angulo and Freeland (1931, p.195) in a half-page list in the introduction to a number of texts. This work, however, leaves out the optative series and much of the negative series. Their presentation of the data is faithfully reproduced below in 4.1. There are a number of differences between their description and Bright's later description, reproduced in 4.2. Some of these are easily reconciliable: De Angulo and Freeland mark the presence of the negative suffix *-ara* in their paradigms, while Bright treats it separately from the agreement system, for example. Others are more intractable: De Angulo and Freeland say that the thou-them (2sg>3pl) marker is *kani-*, a prefix not found in Bright's corpus at all, versus Bright's *i-*. I will not go more in depth to the differences here, though the data from Bright's corpus as assessed through the treebank mostly confirm his description, with the exception of the marker for 3pl subject on 3pl object in negative sentences, discussed below, and the mismatch cases which make up the bulk of this chapter's discussion.

The first full description of the system is in Bright (1957) grammar of the language, which has served as the foundation for study of the agreement system, and which this very description is based on. As is typical of structuralist grammatical descriptions, Bright does not provide actual forms of verbs, example sentences, or contexts to exemplify his description of the agreement system; instead, the underlying forms of the agreement morphemes are presented, and rules for how they combine with verbs to create surface forms are provided.

Further theoretical work has utilized Bright's descriptive work and attempted to model aspects of the system in a variety of agreement frameworks (Macaulay 1992; Béjar 2003; Béjar and Rezac 2009; Campbell 2012; Kumaran 2018). None of this more recent work has questioned Bright's descriptive accuracy and have assumed that his description matches the surface facts of the language, even if they argue for a reinterpretation of the conditions under which a particular agreement marker can be used. The present work was borne out of a desire to assess the accuracy of Bright's description against the data that he himself collected in the texts (which form the corpus of the treebank). This assessment is necessitated both by the fact that Bright's work has served as the source of data for influential theoretical work (and thus should be confirmed) and due to the discrepancies found between Bright's description and De Angulo and Freeland's earlier description. As stated previously, Bright's description was mostly consistent with the treebank corpus, save some inconsistency with the 3pl>3pl

[1]I reproduce an apparent typo in de Angulo and Freeland's description here. The *i* should be before the -, or not there at all. Which of these possibilities was intended is unknown, though the latter would match Bright's description.

Subject-Object	Positive	Negative
he-him	*u*-Stem	-Stem-*ara*
he-them	*kin*-Stem	*kin*-Stem-*ap*
he-thee	*i*-Stem-*ap*	
he-you	*kin*-Stem-*ap*	
he-me	*na*-Stem	*na*-Stem-*ara*
he-us	*kin*-Stem	
they-him	*kun*-Stem	Stem-*ap*
they-them	*kin*-Stem	Stem-*ap*
they-thee	*e*-Stem-*ap*	
they-you	*kik*-Stem-*ap*	
they-me	*kana*-Stem	*kana*-Stem-*ap*
they-us	*kin*-Stem	*kin*-Stem-*ap*
thou-him	*i*-Stem	-*i*Stem-*ara*[1]
thou-them	*kani*-Stem	
thou-me	*na*-Stem	
thou-us	*kin*-Stem	Stem-*ap*
ye-him	*ku*-Stem	Stem-*ap*
ye-them	*ku*-Stem	
ye-me	*kana*-Stem	
ye-us	*kin*-Stem	*kin*-Stem-*ap*
I-him	*ni*-Stem	*na*-Stem-*ara*
I-them	*ni*-Stem	*na*-Stem-*ara*
I-thee	*nu*-Stem	
I-you	*kik*-Stem-*ap*	*kin*-Stem-*ara*
we-him	*nu*-Stem	*kin*-Stem-*ara*
we-them	*nu*-Stem	
we-thee	*nu*-Stem	
we-you	*kik*-Stem-*ap*	*kin*-Stem-*ap*

Table 4.1: De Angulo and Freeland (1931)'s description of Karuk agreement

marker (discussed below) and the cases of number mismatch which the bulk of this chapter investigates.

Table 4.2 presents Bright's description of the agreement paradigm. Bright (1957) presents this data in two ways: organized by morpheme (p. 60) and organized by context, i.e. which subject and object the marker indexes (p. 64). There is a slight inconsistency between these two presentations of the data. Namely, the agreement marker for a 3pl subject acting on a 3pl object in a negative sentence is given on page 60 to be *kín-...-ap* but on page 64 it is said to be just *-ap*. de Angulo and Freeland (1931) gives just *-ap* as the marker for this

context. Sandy (2017) also notes this discrepancy, and decides in favor of *kin-...-ap* "based on evidence from the corpus" though she does not provide this evidence directly (p. 168, footnote 1).

SBJ	OBJ	1sg	1pl	2sg	2pl	3sg/none	3pl
1sg	pos.ind	-	-	nú-	kiik-́...-ap	ni-	ni-
	neg.ind	-	-	kín-	kiik-...-ap	ná-	ná-
	imper.	-	-	nú-	kiik-́...-ap	kán-	kán-
1pl	pos.ind	-	-	nú-	kiik-́...-ap	nú-	nú-
	neg.ind	-	-	kín-	kiik-...-ap	kín-	kín-
	imper.	-	-	nú-	kiik-́...-ap	nú-	nú-
2sg	pos.ind	ná-	kín-	-	-	i-	i-
	neg.ind	ná-	kín-...-ap	-	-	∅-	∅-
	imper.	ná-	kín-	-	-	∅-	∅-
2pl	pos.ind	kaná-́	kín-́	-	-	ku-	ku-
	neg.ind	kaná-...-ap	kín-...-ap	-	-	-ap	-ap
	imper.	kaná-́	kín-́	-	-	kiik-́	kiik-́
3sg	pos.ind	ná-	kín-́	i-...-ap	kiik-́...-ap	u-	u-
	neg.ind	ná-	kín-...-ap	-ap	kiik-...-ap	∅-	-ap
	imper.	ná-	kín-́	i-...-ap	kiik-́...-ap	kám-	kám-
3pl	pos.ind	kaná-́	kín-́	i-...-ap	kiik-́...-ap	kun-́	kín-́
	neg.ind	kaná-...-ap	kín-...-ap	-ap	kiik-...-ap	-ap	kín-...-ap / -ap ?
	imper.	kaná-́	kín-́	i-...-ap	kiik-́...-ap	kun-́	kín-́

Table 4.2: Karuk verbal agreement

It turns out this is a difficult inconsistency to resolve. On close inspection, there are no particularly clear examples in the online corpus showing either to be correct, but let us examine a few relevant examples. First, (1) has a ditransitive verb, *ikvarish* 'buy' with a plural subject, plural inanimate direct object (baskets), and singular indirect object (her). (1) exhibits the *-ap* agreement marker, without *kin-*. Thus, if the plural direct object is what matters for agreement here, we have *-ap* being used for 3pl>3pl. However, the fact that the direct object is inanimate makes it less likely that it is treated as plural for agreement, and from other cases with this verb we know that the indirect object can be what matters for agreement instead of the direct object. (2) shows this later possibility, with a 1st person indirect object causing the agreement to show up as the 3pl>1sg marker *kana-*. As such, it is unclear in (1) what is actually acting as the relevant object for agreement. If it is the 3sg indirect object, the agreement is as expected: just *-ap*. Thus, this example does not help in deciding between the two choices here.

(1) vúra pu-'ikvarish-tih-ap
INTNS NEG-buy-DUR-INV
'People didn't buy (baskets) from her.' (Nettie Ruben, WB_KL-59:3)

(2) kári xás pa-mukun-axvâa kich kanee-kvárish.
and then the-3PL.POSS-head only 3PL>1SG-buy
'Then they bought just the heads from me.' (Benonie Harrie, DAF_KT-05a:51)

There are two other relevant examples with a negative quantifier as subject and third plural object. These two examples, however, have different agreement markers, despite the similarity of context. (3), with *pu-'akara* 'nobody' as the subject and the translation indicating a third plural object is marked with *kín-...-ap*. (4), on the other hand, with the subject *púfaat* 'nothing' as the subject and the translation indicating a third plural object is marked just with *-ap*. One possible explanation for this discrepancy is animacy: the *pu-'akara* 'nobody' of (3) is a quantifier over humans whereas the *púfaat* 'nothing' of (4) quantifies over any object. As discussed below, plural inanimate entities do not necessarily trigger plural marking in Karuk, and so perhaps (4), with an inanimate subject, is utilizing the negative 3sg>3pl agreement (which is just *-ap*) as opposed to (3), with an animate subject, utilizing the negative 3pl>3pl agreement. This would suggest that *kin-...-ap* is the correct 3pl>3pl agreement for the negative series. However, a confounding aspect to these sentences is the presence of the obviative postposition *îin* in both (3) and (4). Macaulay (2000) investigated this postposition in depth, and identified it as being used in situations where a non-protagonist (called 'obviative') is the subject of a verb and the protagonist (called 'proximate') is the object. This is roughly the same function that obviative marking accomplishes in Algonquian languages, hence the use of the term 'obviative' for this postposition. As will be discussed later in this chapter, the presence of *îin* triggers a change in agreement, causing plural agreement with third person singular subjects. But, if that is true, why does triggering plural agreement then result in two different outcomes for these two sentences? These examples, then, are also not ideal for determining the correct agreement for negative 3pl>3pl.

(3) xás pu-'akara-'îin **kín-**maah-tih**-ap**
and NEG-anyone-OBV 3PL>3PL-see-DUR-INV
'And nobody had seen them.' (Julia Starritt, WB_KL-06:6)

(4) púfaat vúra îin áam-tih**-ap**
nothing INTNS OBV eat-DUR-INV
'Nothing eats them.' (Phoebe Maddux, JPH_TKIC-III.3:3)

One last example will be instructive. In (5), the subject is third plural, though the object is translated as third singular. The expected agreement is such a case, with negation, is just *-ap*. However, what is found in this example is actually *kin-...-ap*. Neither Bright nor De Angulo and Freeland mention this as a possibility for 3pl>3sg negative agreement. One possibility is that the translation here is inaccurate, and that it is really a case of 3pl>3pl. (5) comes from 'The Pool at Big Rock,' told by Nettie Ruben, which focuses on the miraculous

effects which the eponymous pool produces. In this story, a man drowns in the pool, and his bones are found by an old woman. She brings his bones to her house, and then (5) is said. After a year, his flesh grows back and he comes back to life. Prior to (5), there is a stretch of sentences with the old woman as the subject, and immediately prior is the sentence translated in (5-a). Perhaps it is possible, in this context, that (5) actually would be more accurately translated as 'people did not see them for a year,' meaning both the old woman and the man are not seen, but there is no further evidence in the story one way or another. If this retranslation is correct, then (5) is evidence that *kin-...-ap* is the correct agreement marker for negative 3pl>3pl. However, the validity of the retranslation is not assured. It is also possible that this sentence is showing us that negative 3sg>3pl is more variable than Bright or De Angulo and Freeland stated, and is seemingly variable between the two possibilities also found for negative 3pl>3pl: *-ap* and *kin-...-ap*.

(5) a. And she carried him back into the house.

b. víri-va itha-hárinay pu-kín-maah-tih-ap
so-thus all-year NEG-3PL>3PL-see-DUR-INV
'People didn't see him for a year.' (Nettie Ruben, WB_KL-59:32)

As it is, I will not be able to resolve the inconsistency for the negative 3pl>3pl agreement marker. It is possible that this supposed inconsistency is actually variability, and that both options are valid but used in different contexts which have as of yet not been identified. I leave the issue to future research. Apart from this inconsistency and the two case studies on unexpected plural and singular agreement below, Bright's description of the agreement system was found to be accurate to the treebank corpus. The case studies below form the only systematic exceptions to Bright's description, and even then, the cases discussed below will not show that any of Bright's description is wrong per se, only that it did not cover all possible uses of each agreement marker.

Before turning to the discussion on unexpected plural agreement, I describe the methodology used to identity such cases below.

4.2 Methodology

The key to identifying when a given agreement prefix is unexpected relies on comparing the expected context of each agreement marker, as explicated above, with the actual context, the person and number of the arguments in a given sentence. Bright's description gives us the expected context, and the annotation of the treebank with the person and number information for each verb's arguments gives us the actual context. For a description of how the latter annotation was completed, see 2.8. A script looked at each instance of each agreement prefix, and compared the actual context of that instance to what was expected.

Those instances that matched the expected context were excluded from the data output, leaving only those that did not match at the end of the search.

The way that the agreement annotation proceeded in the treebank depends quite heavily on the English translations of Karuk sentences, which is a potential confound. As we are trying to uncover cases where the agreement prefix is used when it is not expected, we cannot rely on the agreement markers themselves to tell us what the arguments' person and number information is. In some cases, the expressed arguments themselves will reveal this information, but as plural marking is optional and restricted to only animates and noun-adjective compounds in Karuk this is a rare case when considering number. Translations and close attention to the texts and how each entity is referenced in them are the only consistent sources of information. Translations into a different language are not in general a good way to investigate the meaning of a language, as translations are going to reflect the demands of whatever the language translated into are, not the demands of the source language (see Matthewson 2004 for further discussion). And that difficulty exists even when the translations are generated by bilingual speakers themselves, translating from one language they speak to another they speak. When the translation is possibly the product of another person altogether, the chance for inaccuracy must of course be greater. The exact conditions of the translations in Bright's grammar are unknown. The ideal situation for the current work would be if Bright's translation is merely a standardized version of his consultant's own English translation; the consistency of the translations across several consultants and their adherence to the grammar of formal American English are a sure sign that Bright did some editing, even if the core meanings of the translations themselves originate with his consultants. It is, of course, possible that there was less consultation than this ideal scenario, but it is impossible to know at this point without finding the actual field notes that Bright made during the elicitations of these texts. Unfortunately, the relevant field notes have not, to my knowledge, been located or made accessible. Even if those notes were accessible, there would of course still be uncertainty as to the accuracy of the translations. Like with every other aspect of the treebank, the agreement annotation process is not necessarily going to reflect the truth, but is merely a tool to isolate sentences that may be of interest. As such, there is a slight bias to mark any sentences which could have a mismatch as indeed having one, as not to lose any potential examples. For the case of the data presented below, they were all returned as possible mismatches by the search script, but I then examined each one in context to find further evidence, aside from the translation, that they did indeed involve a mismatch. This evidence is discussed for each individual sentence below.

4.3 Unexpected plural agreement

In this section, we look at the aforementioned cases of unexpected plural agreement in Karuk. These cases involve the use of either the *kun-* or *kin-* prefix on the relevant verb, both of which are taken in Bright's description to indicate a third-person plural subject, but in sentences in which the subject is known to be singular rather than plural. Some aspects of

this pattern have been briefly noted by both William Bright and Monica Macaulay, but this chapter is the first to describe this pattern in depth.

4.3.1 Previous work

The first mention of the unexpected plural agreement under investigation here unsurprisingly comes from Bright (1957), Bright's grammar of Karuk. In his introduction to his presentation of the agreement system, Bright makes the following observation: "Transpersonal[2] themes admit of a special passive construction, formed with the personal morphemes which indicate third person plural subject and third person singular or plural object... The formal distinctiveness of the passive construction is shown by the fact that that it may be used in a sentence where the performer of the action is clearly singular." (p. 59) He illustrates this latter point with the example reproduced in (6).

(6) pa-mu-taat-'îin kun-mah
the-3SG.POSS-mother-OBV 3PL>3-see
'His mother sees him.' (Bright 1957, p. 59)

The subject, *pamutaat'îin*, 'his mother' is singular, and yet the agreement prefix is *kun-*, which is more typically used for plural third person subjects. Thus, this plural agreement is unexpected. Bright explicitly ties this plural agreement to a 'special passive construction' whose only apparent marking is the unexpected use of plural agreement.

However, the presence of the obviative marker *îin* presents a potential confound. Macaulay (2000) identifies this postpositiion as an obviative marker: it is used to mark when an obviative (non-protagonist or non-main character) noun phrase is the subject when a proximate (protagonist or main character) is the object. As Macaulay (2000) describes, the use of the obviative marker *îin* can trigger the use of the agreement markers *kun-* or *kin-* even when the subject is third-person singular. To explain this fact, Macaulay argues that the use of those agreement markers is not a sign of a 'passive construction.' Instead, she argues that, in sentences with *îin*, those agreement markers index the proximate argument, not the subject: *kun-* indicates agreement with a third-person singular proximate argument and *kin-* agreement with a third-person plural proximate argument. Because *îin* is used to mark obviative subjects, its presence indicates that the object is what is proximate in those sentences. For Macaulay, this is what explains examples like (6); The apparent unexpected plural agreement ends up being just a different usage of those agreement markers, one that is expected if you assume those markers are concerned in certain contexts with the obviative vs. proximate distinction instead of number.

A key part of Macaulay's argument is that these cases of unexpected plural agreement only happen in the presence of *îin*, as she states explicitly: "the second key is that these apparently anomalous cases occur... in sentences with *îin*"(p. 490). Indeed, Bright's only provided example of this phenomenon, (6) above, includes *îin*. Bright himself states, however,

[2]Transpersonal is Bright's term for verbs that can have local person objects (Bright 1957, p.59)

that "the special passive construction cannot be considered as conditioned by the presence of *îin* Agentive, since this morpheme also occurs with non-passive transpersonal forms" (p. 59, fn. 2). Bright says nothing more of the construction nor provides any other examples to illustrate it. Though Bright does not tie the passive construction to *îin*, his explanation for why does not say that passive examples are found without *îin*; instead he suggests *îin* is found in non-passive examples. This, then, does not necessarily conflict with Macaulay's understanding of the facts.

Macaulay notes, however, that she found one exception, a sentence with unexpected plural agreement that lacks *îin*. Macaulay does not provide this example, but describes in a footnote the problem: "In fact, I have found one other example with a similarly anomalous use of *kín-*, but which does not contain *'îin*. It is a sentence which does not supply the proper context for use of *'îin*... Yet it does contain an unambiguously singular object marked by *kín-* which normally marks a third-person plural object" (p. 493, fn. 35). In other words, the issue with the sentence lacking *îin* isn't in the anomaly of using a plural subject marker with a singular subject, but with using *kin-* when it doesn't match the object's number.

With the treebank, we are empowered to check the corpus to see if the facts do indeed line up with Macaulay's argument. As I will show, they do not; specifically, Bright's 'special passive construction' is indeed found in sentences that lack *îin*, and as such *îin* cannot be the reason for the unexpected agreement associated with it. Because of the data uncovered using the treebank, we can surmise that Bright's original characterization of these sentences as a 'passive construction' was close to the truth. I will not argue specifically that these sentences are syntactic passives, with the concomitant restructuring of verbal argument structure and syntactic demotion of the agent argument. However, it appears that the contexts of unexpected plural agreement can be subsumed under a general rubric of 'subject demotion,' the details of which will be explored below.

The discovery of the examples of unexpected plural agreement without *îin* is a great example of the utility of the treebank over traditional methods in which linguists have utilized corpora (i.e. by-hand data collection.) Macaulay (2000) reports on data found in the same corpus that serves as the basis for the treebank, plus texts from sources currently not included in the treebank. Macaulay completed this work before the current online dictionary and corpus existed, and so searching for relevant examples was entirely manual. Despite this difficulty, Macaulay collected a comprehensive amount of sentences including the particle *îin*, as the purpose of her paper was to describe and analyze *îin*. However, the necessity of focusing on sentences with *îin* in a manual search meant that relevant sentences which lacked *îin* were likely to be missed. If Macaulay had decided to search back through the corpus in search of sentences with unexpected plural agreement and no *îin*, it would have doubled the already time-consuming work of sifting through the corpus by hand. In addition, identifying sentences is which the agreement is not expected requires more deliberation per sentence than just finding sentences in which the sequence *îin* occurs. One can find sentences with *îin* just by running one's eyes over the sentence. Noticing unexpected agreement would require a longer process of figuring out the subject and object and their features, and checking it against the expected agreement from Bright and actually attested agreement. With a

treebank, the time-consuming work is done up front, and supports continual asking of new questions and finding of new patterns in the data without needing to repeat the time-consuming work of sifting through data again.

4.3.2 Unexpected plural agreement indicates subject demotion

In this section I go through each extant example of unexpected plural agreement identified through use of the treebank. The reason for going through every example, as opposed to just picking a few representative ones as is typical practice, is that this construction has not been previously identified in a rigorous way. I will not claim to explain every facet of this construction, but I do seek to provide an exhaustive examination of its use in the treebank's corpus. I identify four contexts of use, each of which can be subsumed under a general rubric of 'subject demotion.'

Demoted known referent

The most common context for unexpected plural agreement involves a known, singular referent which is apparently demoted. These sentences have the following common properties:

- The subject is a known, singular entity already introduced in the text.
- The subject is unexpressed.

In many of these sentences, the object is the protagonist of the story, but the obviative marker *îin* is not used, and the patterning here differs from the cases where *îin* is used. Namely, in the demoted known referent sentences, the obviative (i.e. non-protagonist) subject is unexpressed and the proximal (i.e. protagonist) object is typically expressed, whereas in the cases where the obviative marker *îin* is found the obviative subject is expressed and the proximal object is typically unexpressed.

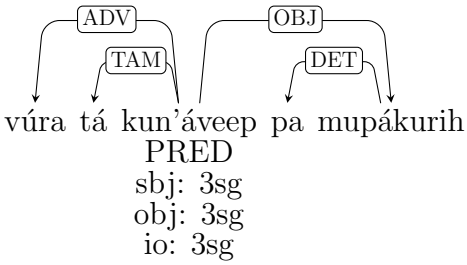
Another typical feature is that these are often translated using English passives. This serves as a clue towards the agent demotion function of this construction. However, the exact conditions of the translation are unknown, as discussed previously, so in the end we must merely assume that the translation's systematic use of passive is reflecting something true about the Karuk sentences themselves.

I now turn to examining each sentence of this type in turn, explaining the context of each and how it is known that there is a mismatch between the agreement and the actual number of the referents.

The example in (7) comes from the story 'Coyote Trades Songs,' told by Nettie Ruben. The essential plot is that Coyote is traveling, comes across another traveller singing a different song, and trades his own song for it. When he tires of the new song, he tries to sing his old one and finds he is unable to. At that point, (7) is spoken. There is no expressed subject in this sentence, but the object *pamupákurih* 'his song' is expressed. The verb, *kun-'áveep*, has the third-person plural subject prefix *kun-*. The translation is passive, but essentially

this is expressing that the other traveller took Coyote's song from Coyote. We know that the other traveller is singular and not a plural entity because, earlier in the text, when the traveller meets Coyote, he triggers singular agreement on the verb: for example, *u'árihvarak* 'he came down from upriver' (WB_KL-07:23).

(7) vúra tá kun-'áveep pa-mu-pákurih
INTENS PER 3PL>3-take.away.from the-3SG.POSS-song
'His song had been taken away from him.', (Nettie Ruben, WB_KL-07:60)

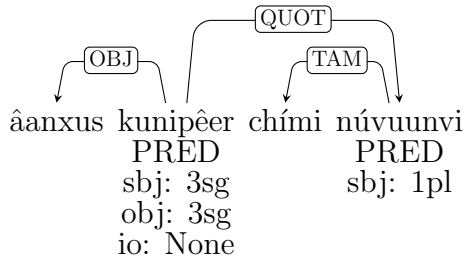


The next series of examples come from stories with Weasel as the protagonist, each called 'The Perils of Weasel.' There are three instances of this story, each told by a different person: WB_KL-18 by Lottie Beck, WB_KL-19 by Mamie Offield, and WB_KL-20 by Daisy Jones. Unexpected plural agreement is found in each of these texts, though not always in the same context. Generally, the plot of each story is that Weasel is given challenges, such as defeating a monster, by a devious old man or woman, and succeeds in these, eventually defeating the old man or woman at the end.

(8) comes from the first Weasel story in Bright (1957), WB_KL-18 by Lottie Beck. In (8), the agreement prefix on the verb *kunipêer* is third-person plural *kun-*, but the translation is passive and does not indicate any plural subject. Again, there is no expressed subject, but the subject is a known entity: the subject is in fact the antagonist of the story, an old widow - a singular entity, not plural. This is clear from the context, as in the previous sentences Weasel has gone to the widow's house to challenge her and then is confronted with this quote; in (9), the sentence following (8), the quote continues with the widow speaking in first-person, using first-person singular possessives and agreement prefixes: she says *nápaathripaahaak* 'if you throw me uphill', with na- indicating first-person singular object, and *naní'aramah* 'my child' with *nani-* indicating a first-person singular possessor. So, the speaker of the quote is treated as singular, despite the quote verb itself having third-plural agreement with *kun-*. Furthermore, in WB_KL-18:62, two sentences after (8), it says the old woman sang, with the verb, *u'árihish*, having third singular agreement *u-*. From this evidence, we can see that the use of plural agreement in (8) is indeed unexpected.

(8) âanxus kun-ipêer chími nú-vuunv-i
weasel 3PL>3S-say.to soon 1>2S-wrestle-IMPER

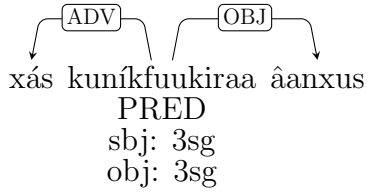
'Weasel was told, "Let's wrestle!",' (Lottie Beck, WB_KL-18:60)



(9) pa=mâaka ná-paathripaa-haak xáat vaa kári naní-'aramah ihrôoha
NOMZ-uphill 2SG>1SG-throw.inland-IRR may so then 1SG.POSS-child wife
'If you throw me into (the corner) uphill, let my child be (your) wife.' (Lottie Beck, WB_KL-18:61)

(10) occurs a few sentences after (9), and the subject is the same: the old woman. Note again, the verb has *kun-* agreement, and there is no expressed subject. We know the old woman is the subject here, because the sentence after (10) states that Weasel threw the widow into the uphill corner, indicating that he and the widow were the ones wrestling (as the widow had challenged him to do in (9)). This example is particularly interesting since the known subject had been mentioned and triggered singular agreement merely a few sentences prior to this one.

(10) xás kun-íkfuukiraa âanxus
then 3pl>3-grab weasel
'Then Weasel was grabbed.', (Lottie Beck, WB_KL-18:65)

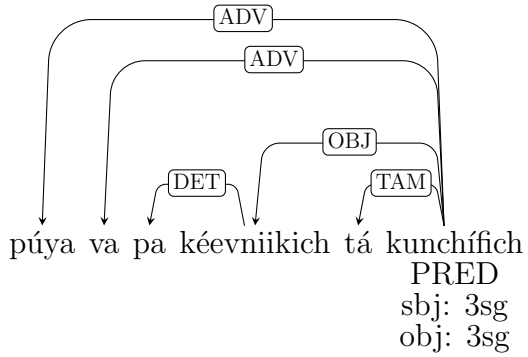


(11) comes from the second Weasel story, WB_KL-19, told by Mamie Offield. The verb, again, has the *kun-* prefix and lacks an expressed subject. The object, the old woman, is expressed. As is typical it is translated as a passive, though there is no passive morphology present. In this case, the subject is actually the protagonist of the story, Weasel. In context, this sentence expresses that the old woman was beaten at gambling.[3] The preceding sentences

[3]Karuk 'gambling' is a traditional Karuk adversarial bluffing game, where two gamblers each have a bundle of small sticks, with one stick marked. The gamblers split up their sticks between their two hands,

set up that the old woman and Weasel are gambling against each other, with the widow's daughter as the prize if Weasel wins. We can tell only these two are gambling, since no other participant is mentioned, and the preceding two sentences showcase the old woman and Weasel's respective gambling songs, and no other. This sentence is particularly interesting for the fact that the protagonist is the demoted subject, rather than another character as is typically the case.

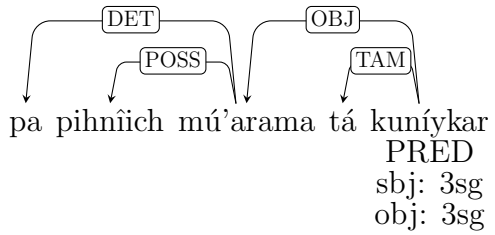
(11) púya-va pa-kéevniikich tá kun-chífich
and-so the-old.woman PER 3PL>3S-beat
'Then the old woman was beaten.', (Mamie Offield, WB_KL-19:15)



(12) comes from the last Weasel story, WB_KL-20, told by Daisy Jones. The verb has the *kun-* prefix, and the subject is unexpressed. The subject is, like (11), the protagonist, Weasel, a singular entity. We know this because in the preceding sentences, Weasel finds and kills the antagonist old man's 'pet' bird (which turns out to be the old man's child). In fact, in the preceding sentence, stating 'he killed the bird,' the verb has singular agreement: *tóo ykar* 'he killed it' with *u-* indicating a third-person singular subject. So, it is just in (12) where the agreement switches to plural, from a previously observed singular. Like (11), the subject of this one is the protagonist of the story. Interestingly, this one is not translated as passive by Bright.

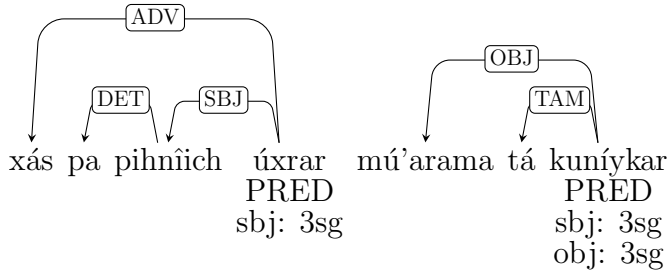
(12) pa-pihnîich mú-'arama tá kun-íykar
the-old.man 3SG.POSS-child PER 3PL>3-beat
'He killed the old man's child.', (Daisy Jones, WB_KL-20:35)

and their opponent tries to guess which hand the marked stick is in, with the one who has split their sticks singing their gambling song.



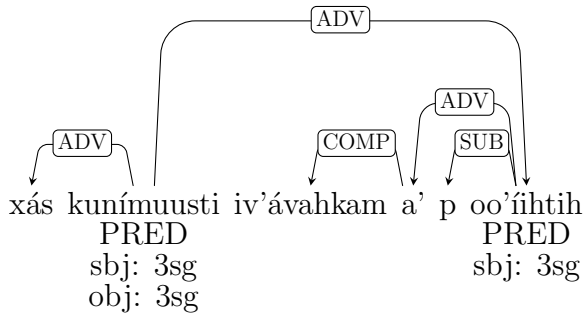
(13) comes also from Daisy Jones' telling of the story, and is similar to (12). The verb in question is the final word in the sentence, and has *kun-* agreement and no expressed subject. The subject is Weasel, as before, and in this case in the preceding sentences Weasel shot and killed a monster, who again turns out to be the old man's child. We know the killing was done by a singular entity because, again, the previous mention of the killing in the previous sentence has singular agreement on the verb: *ukúniihka* 'he shot him,' with *u-* indicating a third person singular subject.

(13) xás pa-pihnîich ú-xrar mú-'arama tá kun-íykar
then the-old.man 3S>3-weep 3SG.POSS-child PER 3PL>3-beat
'And the old man cried, his child had been killed.', (Daisy Jones, WB_KL-20:70)



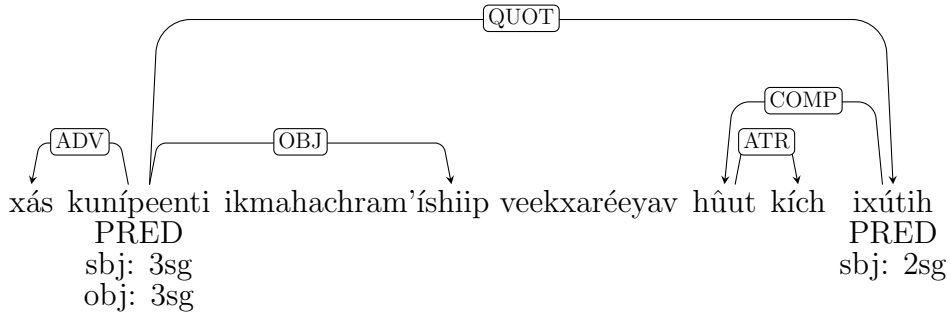
(14) comes from 'Lizard and Grizzly Bear,' WB_KL-34, told by Nettie Ruben. In this sentence, neither subject nor object are expressed, but we know that the subject isn't Lizard, since the previous sentences established that he was dancing on the roof. We also know that the subject is Grizzly, a singular entity, because the previous sentences establish that Grizzly came close to the house where Lizard was, using the verb *tu'uum*, 'she got there', with the *u-* indicating a third person singular subject.

(14) xás kun-ímuusti iv'ávahkam a' p-oo-'íih-tih
then 3PL>3-look.at roof above SUB-3S>3-dance-DUR
'And (Lizard) was looked at as he danced, up on the roof.', (Nettie Ruben, WB_KL-34:35)



(15) comes from 'Medicine for the Return of Wives,' WB_KL-52, told by Chester Pepper. The second word of the sentence is the relevant verb, and has *kun-* as its prefix. The subject is unexpressed and the expressed object is the protagonist of the story, Sacred Sweathouse Spirit. The preceding sentences establish that one person approached Sacred Sweathouse Spirit and began chatting with him, using the verb *u'uum* 'he got there', with *u-* indicating third person singular agreement. As such, we know the subject is singular, and it is known in the sense of having been mentioned in previous sentences. However, there is not much information about the subject, which makes this example somewhat similar to the unspecified subject examples below.

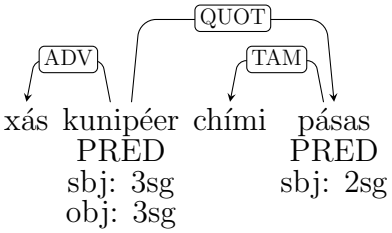
(15) xás kun-ípeen-ti ikmahachram'íshiip v-eekxaréeyav hûut kích
then 3PL>3S-say.to-DUR sacred.sweathouse 3.POSS-god how only
i-xú-tih
2S>3-think-DUR
'And (the person) said to Sacred Sweathouse Spirit, "How are you feeling?"', (Chester Pepper, WB_KL-52:49)



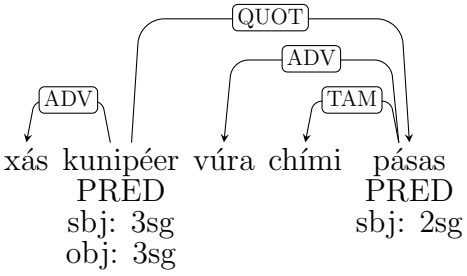
(16) and (17) are nearly identical sentences which both come from 'The Boy from Itúkuk', WB_KL-57, told by Nettie Ruben. The relevant verb in both sentences is *kunipéer*, with the *kun-* prefix. The context for both of these sentences is that the protagonist, who is the object of these sentences, is talking with *pataprih'ifápiit*, the Pataprihak girl, who exhorts him to dress in regalia for a coming dance. The sentence prior to (16) makes it clear that

the subject must be a singular entity, the Pataprihak girl, as it states that the Pataprihak girl was there, using the verb *ukrii* with *u-* indicating a third person singular subject.[4]

(16) xás kun-ipéer chími pásas
then 3PL>3S-say.to soon get.dressed
'And she told him, "Dress up (in dance regalia)!"', (Julia Starritt, WB_KL-57:73)



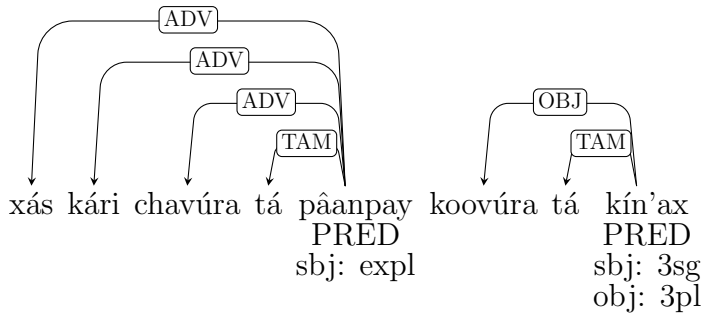
(17) xás kun-ipéer vúra chími pásas
then 3PL>3-say.to INTENS soon get.dressed
'And she told him, "Do dress up!"', (Julia Starritt, WB_KL-57:76)



(18) comes from 'Lizard and Grizzly Bear', WB_KL-34, told by Nettie Ruben. We had an example from this text above as well, but this particular one uses *kin-* as the agreement prefix rather than *kun-*; *kin-* indicates a third person plural subject and third person plural object, as opposed to *kun-* which has a plural subject and singular object. Apart from this difference, the construction is the same: a plural agreement prefix, the subject is unexpressed but is a known singular entity. In this case, the subject is Grizzly. Previous sentences in the story establish that Grizzly has been killing each of a group of ten brothers, using, for example, the verb *u'ax* 'she bit him' (WB_KL-34:10). As such, the killer has already been identified and is singular, despite the use of *kin-* in (18).

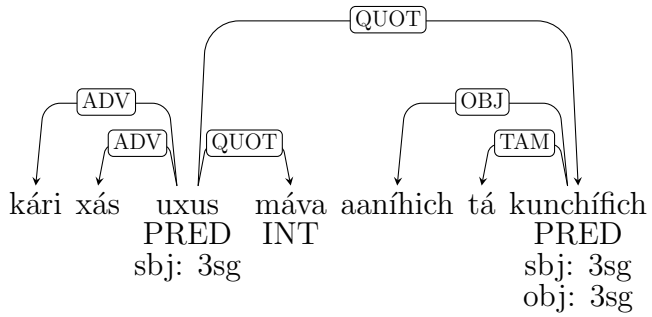
[4]Two sentences prior to this one, it notes that an old woman and the Pataprihak girl are there, but the succeeding sentence focuses specifically on the girl as a singular entity. Furthermore, the translation actually indicates the subject is the singular 'she' in this case.

(18) xás kári chavúra tá pâanpay koovúra tá kín-'ax
then then finally PERF after.a.while all PERF 3PL>3PL-bite
'And finally after a while all of them were killed.', (Nettie Ruben, WB_KL-34:14)



(19) comes from 'Shinny Game Medicine' (WB_KL-54), told by Mamie Offield. The relevant verb is the last word in the sentence, *kunchífich* 'they beat him.' It has the prefix *kun-* and the translation is passive. The thinker of this quote is the protagonist, the littlest of Burrill Peak Spirit's children. In the story, the protagonist's nine brothers all go to play the shinny game with the son of Baldy Peak Spirit and are all defeated. (19) is the first mention of one of these defeats. The key here is that the protagonist's big brother is not defeated by a plurality, but by a singular entity, the son of Baldy Peak Spirit, as (20), two sentences prior, lays out. This fact is what shows us that the plural agreement marker *kun-* is not expected here. In (20), the evidence for the singularity comes not from the verb *kun'îimasar*, which has plural agreement with *kun-*, but from the use of the subject postposition *xákaan*, which is used only when the subject numbers two individuals and no more. The postposition *koovan* is used if there are more than two individuals making up the subject. As such, we know that the subject of (20) is two individuals, and those two individuals are the protagonist's brother, not expressed in the sentence, and the son of Baldy Peak Spirit. Those are the only two who are engaged in the game, and as such when the brother is defeated in (19), it can only be by his singular opponent.

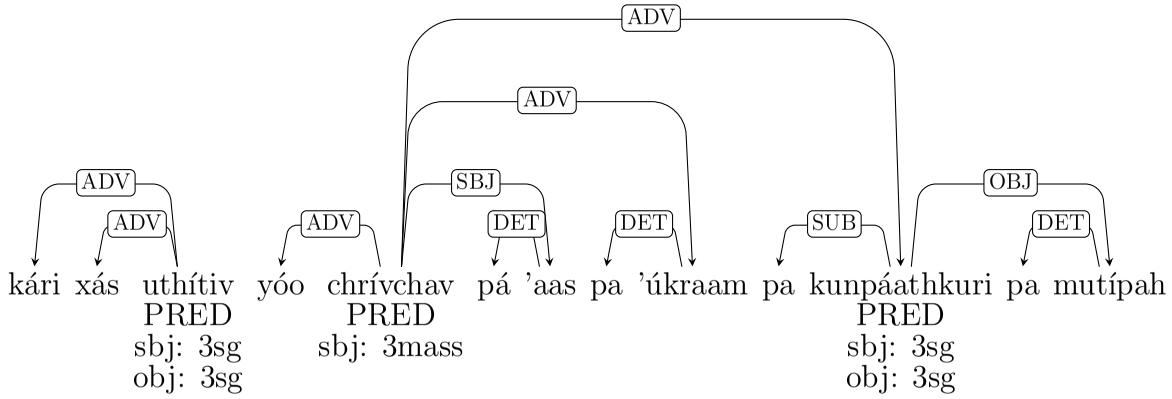
(19) kári xás u-xus máva aaníhich tá kun-chífich
and then 3SG>3-think look! my.older.brother PER 3PL>3S-beat
'And he thought, "Look, big brother's getting beaten."', (Mamie Offield, WB_KL-54:11)



(20) asaxêevar v-eekxaréeyam mú-'arama xákaan tá kun-'îimasar
Baldy.Peak 3.POSS-god 3SG.POSS-child together PER 3PL-grapple
He and Baldy Peak Spirit's child grabbed each other (preparing to play). (Mamie Offield, WB_KL-54:9)

(21) comes from 'Wrestling Medicine' (WB_KL-55), also told by Mamie Offield. The relevant part is the final clause, with the verb *pakunpáathkuri* 'when they threw him in water,' which has the plural subject agreement marker *kun-*. The context for (21) is a group of ten brothers, including the protagonist named *kunâach'aa*, are one-by-one going off to wrestle a giant and, until *kunâach'aa* goes, being defeated as (21) says. This is a case of unexpected plural agreement because we know that the subject of *pakunpáathkuri* is not a plural entity; it is the singular giant. Earlier in the story, this giant is even indexed with singular agreement, as shown in (22). In (22), the brothers discuss the giant, saying he lives at the edge of a lake using the verb *ú-krii*, where the *u-* indicates a third singular subject. The giant has already been established as singular, but then is unexpectedly indexed with a plural agreement marker in (21).

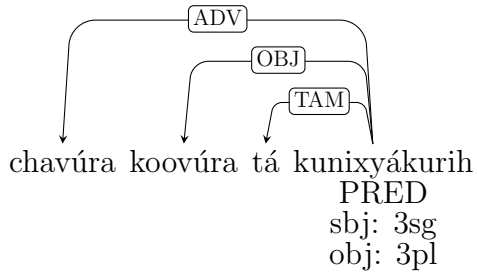
(21) kári xás u-thítiv y-óo chrívchav pá-'aas pa-'úkraam pa-kun-páathkuri
and then 3S>3-hear visible-3S>3 splash the-water the-lake SUB-3PL>3-throw.into
pa-mu-típah
the-3SG.POSS-brother
'Then he heard it, he saw the water splash in the lake, when (the giant) threw his brother in.', (Mamie Offield, WB_KL-55:12)



(22) kári xás kun-piip "maruk'áraar vaa káan úknam-tiimich ú-krii."
and then 3PL-say giant so there lake-edge 3SG-live
'And they said, "A giant is staying there at the edge of the lake."' (Mamie Offield, WB_KL-55:3)

(23) likewise comes from Mamie Offield's 'Wrestling Medicine,' and the argumentation from above applies as well. The verb, *kunixyákurih* 'they threw him in' has the *kun-* prefix indicating a plural subject, but we know the subject, the giant, is singular.[5]

(23) chavúra koovúra tá kun-ixyákurih
finally all PER 3PL>3-throw.into.water.(several)
'Finally (the giant) threw all (the brothers) in.', (Mamie Offield, WB_KL-55:13)



We've now seen a number of examples of this pattern, from multiple Karuk speakers, and it is safe to say that that reanalysis of these as somehow having a plural subject after all is impossible or unlikely; there is something else going on here, and the pattern is robust enough that it seems unlikely to be some kind of speech error. The question that remains is what this construction is being used for. I mentioned above that this construction appears to be used for agent or subject demotion. In the sentences discussed above under the rubric of 'known demoted referent' however, there is not much direct evidence in the Karuk sentences themselves to defend that hypothesis; rather we can detect agent demotion in these only subtly, through the consistent use of the English passive in translation. Other subtypes of this construction, however, lend themselves slightly more readily to that hypothesis, and as such we can hypothesize that the same explanation holds for all the subtypes.

[5]Interestingly, there is a different sort of mismatch here, with *kun-* being used instead of *kin-*. The object here, all the brothers, should be plural, but *kun-* is used despite it indicating a singular object. This sort of mismatch is not as robust as the unexpected plural and as such will not be discussed much, but it seems this kind of mismatch is facilitated by the fact that the verb root itself, *ixya-*, inherently means 'to throw (plural objects).' The plurality of the objects is an inherent part of the verb root. Perhaps future research can explore the relationship between such lexical plurality and the plurality in agreement markers.

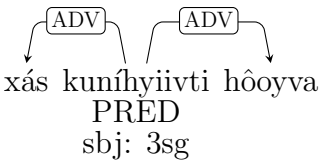
Unspecified subjects

The context of the unspecified subject type is closely related to the demoted known referent type above. The difference between this context and the known referent type is that, instead of the subject having already been introduced in the narrative, in these sentences the subject has not yet introduced. As in the known referent cases, the subject is still not explicitly mentioned.

These cases of unspecified subjects can be detected to involve agent or subject demotion because they seem to have a 'suspense-building' function in the narratives. The use of *kun-* gives a sense of withholding information about the subject until an eventual reveal, much as the passive can be used in English to achieve the same function. The following examples make this function clear.

(24) comes from 'The Creation of Eels' (WB_KL-41) told by Nettie Ruben. In the sentences before (24), the protagonist *ithyarukpíhriiv*, Across-the-Water Widower, is heading upriver. Then he hears a shout, which is described as in (24): *kun-* is on the verb, and there is no expressed subject. The only character who has been mentioned is Across-the-Water Widower, and there is no reason to expect one way or the other that the shouting is coming from a plural or a singular entity at this point. Later on, we find that the shouting is indeed coming from a singular entity, the character Tick. As such, we know that the *kun-* here is not being used to indicate a plural subject, since that is later contradicted.

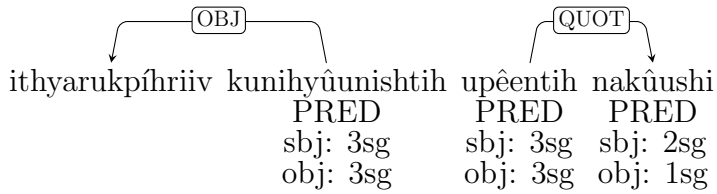
(24) xás kun-íhyiiv-ti hôoyva
then 3PL>3-shout-DUR somewhere
'And there was a shout somewhere.', (Nettie Ruben, WB_KL-41:3)



(25) comes from the same text, 'The Creation of Eels,' a few sentences after (24). The first part is essentially the same as (24), with the verb being *ihyûunish* 'to shout at' (with the same root as the verb in (24)) and the prefix being *kun-*. The subject is still unexpressed, though the object, *ithyarukpíhriiv*, is expressed. This sentence is quite revealing because of the second half, however. The verb in the second half, *upêentih*, has the same subject as the shouting verb; in fact these verbs seem to be describing essentially the same event. And yet, the second verb has *u-*, indicating a singular subject, as its prefix, rather than *kun-*, despite following immediately after the verb with *kun-*.

(25) ithyarukpíhriiv kun-ihyûunish-tih u-pêen-tih
across.water.widower 3PL>3-shout.to-DUR 3S>3-say.to-DUR

na-kûush-i
2SG>1SG-copulate-IMPER
'Across-the-Water Widower was shouted at, (the person) said to him, "Copulate with me!"', (Nettie Ruben, WB_KL-41:7)

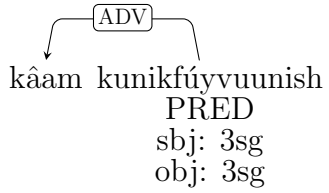


A few sentences later, in (26), it is explicitly revealed that the shouter is Tick, and thus could not possibly be a plural entity. We also see that the identify of the shouter is withheld for 13 sentences, and that when it is finally revealed, the adverb *hínupa* is used; *hínupa* is glossed as 'surprise' but seems here to be used when the suspense is finally resolved.

(26) kári xás hínupa chantírih
and then surprise tick
'And there it was Tick.' (Nettie Ruben, WB_KL-41:13)

(27) comes from the story 'The Kidnapped Child' (WB_KL-61), told by Lottie Beck. In this story, a child is kidnapped while very young from outside his mother's house, and grows up living with his kidnappers. Eventually, the child grows up and is told by his kidnappers not to shoot his bow over the hills. Curious, the child does so, and when he goes to retrieve his arrows from over the hills, an unknown person or persons gets his attention and tells him that he was kidnapped. (27) has the first case of unexpected plural agreement in the text. The relevant verb, *kunikfúyvuunish* 'they whistled at him' has the *kun-* prefix, indicating a plural subject, but at this point the identity or number of the whistlers is unknown. Two sentences later, (28) reveals that a singular voice rings out, with the verb *u'aramsîiprin* using the *u-* prefix indicating a singular subject. From this mismatch between the plural agreement of the whistle and the singular agreement of the voice ringing out, we cannot really surmise the number of the whistlers.

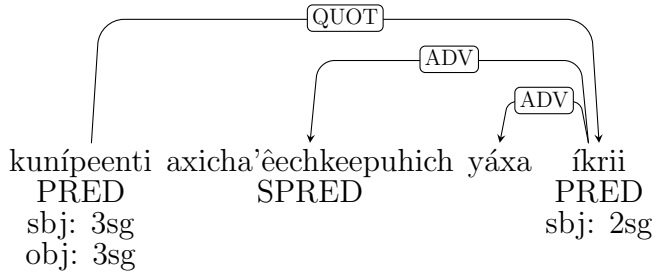
(27) kâam kun-ikfúyvuunish
little.upriver 3PL>3-whistle.at
'He was whistled at, a little ways upriver.', (Lottie Beck, WB_KL-61:25)



(28) axmáy xás pa-'úuhyan hôoyva u-'aramsîiprin
suddenly then the-voice somewhere 3SG-start.out
'Suddenly a voice came from somewhere.' (Lottie Beck, WB_KL-61:27)

Immediately after (28), we get another example of unexpected plural agreement, where the speaking switches from singular back to plural agreement in (29). The verb is *kunípeenti* 'they are saying to him,' with the plural subject prefix *kun-*. The translation here is passive, like as to the demoted known referent cases discussed above. In these examples, unlike the others, I cannot say for sure that there is a singular subject being indexed as plural. However, the point here is that *kun-* is being used when there is not clearly a plural subject, and importantly, when there are no details about the subject being put forward at all. In English, we may use a passive to achieve the same effect, of not providing any information about the subject when it is either irrelevant or being withheld for a narrative purpose, as it seems is true in this set of examples; the identity of the characters who reveal to the child that he is kidnapped is not important, given that the story is focused on the child and his eventual return to his mother.

(29) kun-ípeen-ti axicha-'êechkee-puh-ich yáxa í-krii
3PL>3-say.to-DUR child-kidnap-having.been-DIM look! 2SG>3-live
'He was told, "Look, you are a kidnapped child!"', (Lottie Beck, WB_KL-61:28)



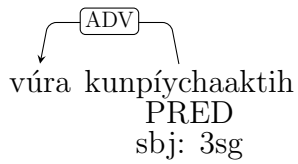
Impersonal

The least common context is with impersonals, where there is no specific subject. Only two examples of this kind are found. Impersonals appear to be more readily used with the *-ahi* suffix, discussed below.

The examples below fall into two distinct subtypes of impersonal construction: (30) is an example of a corporate impersonal. Corporate impersonals are those in which the vague or unspecified subject is 'culturally designated' (Malamud 2013, p.5), such as in the English sentence 'They've raised the taxes again' where the cultural understanding that the government raises taxes allows one to infer that *they* refers to the government. (30) comes from 'The Story of Madrone' (WB_KL-35), told by Lottie Beck. In (30), the verb is marked with the plural subject marker *kun-*: *kunpíychaaktih* 'they inflicted bad luck on him.' The translation renders the Karuk expression into the more common English expression for bad luck, 'had bad luck', in which the subject is the one who experiences bad luck, but Bright in his lexicon (Bright 1957, p.375) notes that this Karuk verb is transitive. As 'bad luck' is not an expressed nominal in the Karuk sentence, it is not the object as in the English expression; the object is instead the one who is experiencing bad luck, as the glossing with 'inflict' represents. In the context of the story, there is no obvious antecedent for this plural; up to that point, the story has been establishing that the protagonist (the object in (30)) is poor and had poor luck at hunting. In (31) we can see that the sentences immediately prior to (30) in the story reference the protagonist using *u-*, the singular 3rd person subject marker, so even if *kun-* in (30) was meant to reference the protagonist, we would have a mismatch.

However, the use of *kun-* here appears to not be an unexpected way to reference the singular protagonist. The Karuk dictionary offers up an explanation for the unexpected use of *kun-* with this verb, stating that *píychaak* is 'generally used with implied supernatural subject, as in... 'they (the spirits) inflicted bad luck on me.' With this information, we can surmise that the *kun-* here is used as in a corporate impersonal; the subject is never explicitly identified, but cultural understanding would lead a listener to infer that the subject is 'the spirits.'

(30) vúra kun-píychaak-tih
INTENS 3PL>3S-inflict.bad.luck-DUR
'He had bad luck.', (Lottie Beck, WB_KL-35:10)

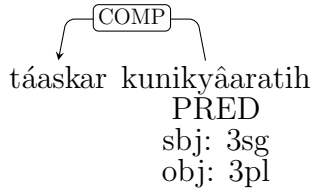


(31) imáan-kam kúkuum t-u-vâaram
next.day-side again PER-3SG-go
'The next day he would go again.' (Lottie Beck, WB_KL-35:9)

(32) comes from 'Salmon Fishing,' a procedural text told by Julia Starritt describing how, in

general, Karuk people traditionally fished for salmon. (33) is the sentence immediately prior to (32) and provides the context that people made 'fishing platforms' *imvír* to fish from; (32) begins to describe how these platforms were made. The verb in (32), *kunikyâaratih*, is marked with the plural subject marker *kun-*, but this is not reflected in the translation, which opts for a passive instead. In (33), the subject is also indexed with *kun-*, but there the translation is rendered with English 'they.' Perhaps the subject of (32) must be this same 'they' from the previous sentence; why is this unexpected?

(32) táaskar kun-ikyâara-tih
pole 3PL>3-make.with-DUR
'It was made of poles.', (Julia Starritt, WB_KL-69:3)



(33) píshiip imvír tá kun-íkyav
first fishing.platform PER 3PL>3-make
'First they made fishing platforms.' (Julia Starritt, WB_KL-69:2)

(34) will make this clear. (34) is the first sentence of this text, introducing that the text is about how Karuk people fished for salmon. However, the verb is not marked with a plural subject marker; rather, it is marked with *u-*, indicating the subject is singular. In fact, this is the typical practice for generic sentences such as this one: this text is not describing any particular episode where Karuk people fished, but a general process for how it was done, without referring to any specific person or event. The marker *u-* is typically used in such contexts; as such, the *kun-* in (33) and (32) is unexpected. (33) and (32) seem much like the so-called 'universal' impersonals, like the English sentence 'They speak Spanish in Spain;' such sentences are used to describe generalities without particular reference to any specific people (Malamud 2013), much as (33) and (32) do. The use of *kun-* in these examples, then, parallels the use of 3rd person plurals in impersonal constructions in languages like English and Russian (see Malamud (2013) for discussion of these latter two languages).

(34) pa-'áraar uum pa-'áama u-kup-éekriihv-ahi-tih
the-indian 3SG.PRO the-salmon 3SG>3-MOD-fish-MOD-DUR
'The Indians fished for salmon in a certain way.' (Julia Starritt, WB_KL-69:1)

So why mention these impersonals in this discussion at all? They are not unexpected in the same way as the previous types of unexpected plural were, in that one can generally

come up with a conceivable plural entity who the *kun-* could be indexing. The reason for mentioning them is that impersonals are well-understood as being constructions that demote the subject (Malamud 2013), and the usage of plural agreement in these cases is the only otherwise exceptional thing about these sentences which could be the source of this demotion. The other types of unexpected plural agreement discussed above also had some subtle hints that subject was demoted, in having passive translations. The fact that apparently the same construction, the unexpected use of *kun-*, is also used in impersonals which are recognized as involving subject demotion, is further evidence that subject demotion is an effect of the use of this construction.

Obviative

The last type I will discuss is actually the most common context for unexpected plural agreement, and that is contexts where the obviative marker *îin* is present. Macaulay (2000) treats this phenomenon in depth, and as such I will only review some of the surface properties most relevant to distinguishing it here. These sentences all involve the use of *îin*, which Bright (1957) describes as an 'agentive postposition' (p. 129). This postposition *îin* takes the subject noun phrase as its complement,[6] and whether or not that subject is singular or plural, the verb gets plural marking.[7]The cases where the subject is singular and thus does not match the use of plural agreement marking constitute the cases of unexpected plural agreement of interest for this discussion. Macaulay (2000) argues that *îin* sentences exhibit the following properties (p. 466, adapted from (3)):

- The verb must be transitive
- The subject is typically expressed by a full NP
- The object is typically not expressed
- The object must be the main character/protagonist in the narrative (and thus, the subject must not be.)

(35) is a typical example of this type of sentence. It comes from 'The Story of Madrone,' told by Lottie Beck. In the events before (35) is said, Madrone (the protagonist) has been wandering daily and meeting a girl from upriver, but one day he stays away from home too long talking to this girl. Then (35) is uttered. The verb *kunpapivar* is marked with the plural agreement marker *kun-*, but the subject here is Madrone's father, who is a singular entity.

[6]This class of postposition is described in the Subject section of the annotation guidelines, found in 2.7.1.

[7]The discussion of the obviative in this section is relevant mainly for uses where the obviative marked subject is third-person. Obviative marking can be found with first and second person subjects as well, but these do not exhibit the same agreement behavior and typically have the expected agreement marker given the person/number of the subject and object. The mismatches described here are only in cases where third person subjects are involved.

The expressed subject, *pamu'akah'îin*, 'his father,' is itself marked by *îin*[8] The protagonist, Madrone, is the object of this verb, and as such, the subject is not the protagonist.

(35) pa-mu-'akah-'îin kun-p-apiv-ar
the-3SG.POSS-father-OBV 3PL>3-ITER-look.for-AS.MOT
'His father went to look for him.' (Lottie Beck, WB_KL-35:22)

There are a few key differences between the obviative type and the other types of unexpected plural agreement discussed above. For one, in the other cases and particularly the demoted known referent cases, the subject can be either the protagonist or another participant in the story. Interestingly, even in cases where the protagonist is the object (and all other conditions on *îin*'s use are met), occasionally *îin* is not used. In (36), repeated from (12) above, Weasel is the protagonist and the (unexpressed) subject. Conversely, in (37), repeated from (8), the (unexpressed) subject is the antagonist, the old woman, and Weasel, the protagonist, is the object.

(36) pa-pihnîich mú-'arama tá kun-íykar
the-old.man 3SG.POSS-child PER 3PL>3SG-beat
'He killed the old man's child.', (Daisy Jones, WB_KL-20:35)

(37) âanxus kun-ipêer chími nú-vuunv-i
weasel 3PL>3SG-tell soon 1SG>2SG-wrestle-IMPER
'Weasel was told, "Let's wrestle!', (Lottie Beck, WB_KL-18:60)

Another key difference is that, in the non-obviative cases, it is more typical for the object to be expressed, as in both (37) and (36), than for the subject to be expressed. With *îin*, it is more likely for the subject to be expressed.

These differences, I surmise, are really about the presence and function of *îin*, rather than a difference in the function of the unexpected plural agreement itself. Namely, the use of *îin* specifically marks the subject noun phrase as being an unexpected subject. in the sense that non-protagonists are not as expected to be subjects while the protagonist is the object. This phenomenon appears to be an example of what Haspelmath (2021) formulates as "the grammatical form-frequency correspondence hypothesis," reproduced below:

(38) *The grammatical form-frequency correspondence hypothesis* When two grammatical construction types that differe minimally (i.e. that form a semantic opposition) occur with significantly different frequencies, the less frequent construction tends to be overtly coded (or coded with more segments), while the more frequent construction tends to be zero-coded (or coded with fewer segments), if the coding is asymmetric. (Haspelmath 2021, p.606)

[8]In some contexts, îin cliticizes onto its complement, forming a single phonological word, and in others it is a separate word. Syntactically, the treebank treats these cases the same.

Regarding *îin* in light of (38), its presence indicates a less frequent case: that the subject is not a protagonist, and thus involves not only the expression of 'more segments' (the particle *îin* itself) but also frequently the full expression of that less expected type of subject. Less frequent occurrence leads to more overt expression. One could argue about whether the distinction between proximate/protagonist and obviative/non-protagonist is a 'semantic opposition,' but perhaps a general principle like (38) should apply equally to more pragmatic or discourse-structural properties like protagonist-hood as well.

Though the function of *îin* is tied up in the variable expression of the subject, the use of unexpected plural agreement in these cases is for the function of subject/agent demotion, just as with the other cases described above. The differences in subject expression do not signal a change in the function of unexpected plural agreement, in other words. In the obviative cases, the 'subject demotion' is found in the fact that subject is obviative, not a protagonist, and thus ranked lower than the proximate, protagonist object, assuming that, in the case of the Karuk obviatives, the relevant scale is one in which a protagonist is ranked higher than other characters (Macaulay 2000, p.5). Taken all together, the function of unexpected plural agreement is to mark, in a general sense, that the subject is demoted. The particular way this demotion is handled differs by context, however, in the ways described above.

It is worth attempting to locate this Karuk construction within the cross-linguistic typology of passive and passive-like constructions. Legate (2021) proposes three criteria of canonical passives, reproduced below (from p.375). Legate states that languages 'may exhibit a voice construction with any subset of these properties," and gives an example of a language for each logical possibility.

(39)
a. Agent demotion. The agent is semantically present but is not syntactically present as a noun phrase in its characteristic thematic position. Instead, the agent is either interpreted as existential ('someone') or associated with a by-phrase.
b. Theme promotion. The theme raises from its low syntactic position associated with the interpretation as a theme to the grammatical subject position.
c. Morphological marking. The verbal morphology is distinct from the active voice. (Legate 2021, p.375)

The unexpected plural agreement of Karuk fulfills the morphological marking criteria, though it does so by an unexpected use of agreement morphology whose sole purpose isn't to mark 'passivization.' The other two criteria are harder to judge. I have argued that unexpected plural agreement involves subject/agent demotion, but in a different sense from how Legate uses the term. For Legate, agent demotion is syntactic: as she states, the demoted subject is "not syntactically present in its characteristic thematic position." For the Karuk unexpected plural cases, subjects are optionally present like in any typical declarative Karuk sentence, and only sometimes are part of a postpositional phrase (with the obviative *îin*) that is, however, not conclusively a by-phrase. Whether the subject is in its "characteristic thematic position" is a hard question to answer for Karuk, given the freedom of position noun phrases

enjoy in Karuk. The only discernable effect is the mismatch in agreement, which perhaps could be caused by the subject not being in its normal position and thus being unavailable for agreement.[9] There also is not clear evidence for the promotion of the theme in these sentences: the objects can be optionally expressed with the apparent free word order of any other declarative transitive sentence, and the object does not control subject agreement (as in these cases the subject agreement is always plural, irrespective of the object's number.) As such, the unexpected plural case only clearly exhibits Legate's morphological marking criteria, with the other two requiring further investigation to determine conclusively.

4.3.3 The *-ahi* passive

So far, I've argued that the use of unexpected plural agreement is a sign of subject demotion, and that it appears to be used in several distinct contexts of subject demotion. There are, however, other contexts which can be described as subject demotion that do not utilize *kun-* or *kin-* agreement; these contexts instead make use of the verbal suffix *-ahi*, which Bright called the 'essive' suffix (Bright 1957, p.92), and have *u-*, the third-person singular subject prefix, as their agreement prefix. (40) presents a typical example. The verb, *unhíkahitih*, is marked with *u-*, the 3rd person singular subject marker. The subject here is *pa-'urípi*, 'the net,' by virtue of the fact that *-ahi* has reduced the valence of the transitive verb *inhi* 'to tie,' which without the addition of *-ahi* would have the agent who performed the tying as the subject, as (41) shows, as opposed to a theme like the net.

(40) xás vaa káan pa-'urípi u-nhí-k-ahi-tih
and thus there the-net 3SG-tie-onto-ESS-DUR
'And the net was tied on there.' (Julia Starritt, WB_KL-69:7)

(41) kári xás u-hní-shriih-va koovúra pa-'ûumukich pa-'áthiith...
and then 3SG>3-tie-down-PL.ACT all the-near the-hazel.branches
'And she tied all the hazel branches nearby...' (Mamie Offield, WB_KL-64:7)

As such, cases of subject demotion with *-ahi* do not involve unexpected agreement, which seems to complicate my claim that the third-person plural prefixes are involved in subject demotion. Why are they not used in these clear subject demotion cases involving *-ahi*?

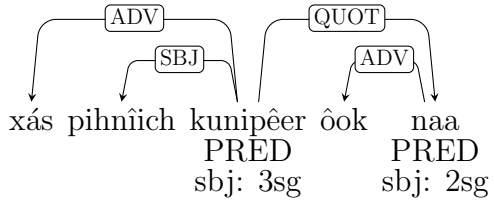
[9] Peter Jenks (p.c.) suggested that the unexpected plural agreement could involve transitive expletive constructions. Transitive expletive constructions involve the use an expletive pronoun (like 'there' in the English sentence 'There are a lot of cats outside') combined with a transitive verb (and its two arguments, subject and object). This sort of construction is not typically licit in English, but an ungrammatical example can suffice to show what such a construction would entail: *There has someone eaten an apple. (This example is taken from (Bobaljik and Jonas 1996, p.208) ex. 15a). For the unexpected plural cases in Karuk to be examples of transitive expletive constructions, one could posit that each example of unexpected plural agreement actually contains a null expletive pronoun; perhaps this pronoun triggers the plural agreement that otherwise has no obvious source. Delving into the details of such an analysis would take us far afield here of the goal of exemplifying the utility of the treebank, however, and so I leave it for future exploration.

Essentially, it boils down to there being two different constructions which can lead to subject/agent demotion, with some differences between them.This sort of system is not unattested: Legate et al. (2020) describes how Turkish has both an impersonal construction and passive construction, both previously identified as passives and which on a surface level appear to involve agent demotion (though Legate et al. (2020) in fact argue that the Turkish impersonal does not involve syntactic agent demotion). I will not be able, with only the corpus data at hand, to elucidate the differences between the Karuk constructions in all their detail, nor to propose any formal derivations for these. Some differences between the two constructions are apparent, however, and described below:

- The *-ahi* construction can be used with both intransitive and transitive verbs. Only the transitive cases appear to involve agent demotion. Unexpected plural agreement only appears with transitive verbs.
- Agents are never expressed with the *-ahi* construction, whereas agents are sometimes expressed with the unexpected plural agreement construction, typically co-occuring with the *îin* obviative particle, though not always.

The main important difference between the *-ahi* cases and unexpected plural cases is that, as shown above and reproduced below, the agent can still be expressed in sentences with unexpected plural agreement, whereas the agent is never expressed with *-ahi*. Namely, the obviative sentences described above generally include an expressed subject/agent, and even some of the unexpected plural cases without the obviative *îin* have an expressed subject as well. (42) comes from the Story of Weasel (WB_KL-20), and has the singular subject/agent, *pihnîich* 'old man', expressed along with plural agreement *kun-* on the verb.

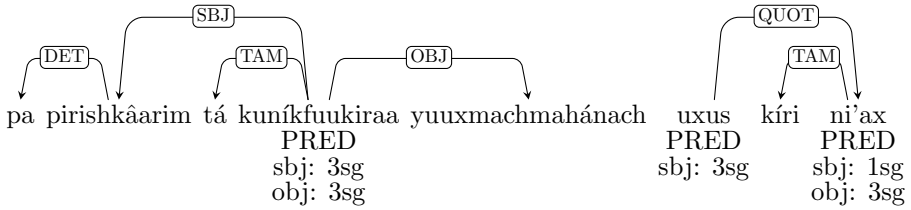
(42) xás pihnîich kun-ipêer ôok naa
then old.man 3PL>3-say.to here come
'And the old man said, "Come here!"', (Daisy Jones, WB_KL-20:42)



(43) likewise is a sentence with unexpected plural agreement and the singular subject expressed. The subject/agent is *pa-pirishkâarim*, Grizzly Bear, and the verb is marked with plural agreement *kun-*.

(43) pa-pirishkâarim tá kun-íkfuukiraa yuuxmachmahánach u-xus kíri ni-'ax
the-grizzly PER 3PL>3-grab lizard 3SG-think let 1S>3-bite

'Grizzly (in her death throes) grabbed at Lizard, she thought, "Let me kill him!"', (Nettie Ruben, WB_KL-34:58)



Though these two sentences above are clearly a minority of examples of unexpected plural agreement that lack the obviative marker, their existence nonetheless shows that it is possible for the agent to be expressed in these cases. And that is what matters for us.

In contrast, no example involving *-ahi* involves an expressed agent, out of 65 examples on the online corpus.[10] Subjects are expressed in these sentences, though these are not agents, as the following examples will illustrate.

In (44), the verb is *ukyâarahitih*, which includes *-ahi* and the stem *ikyâar*, built of *ikyav* 'to make' and *-ara*, an applicative which adds an instrument (in (44), the added instrument argument is the little poles, *taskanatunvêech*.). Typically, *ikyav* has an agentive subject, as (45) shows, where the subject *pa'avansáxiich* 'the boy' is the agent, the maker of the fire. In (44), however, the maker of the fish-trap is not mentioned. In this case, the treebank treats *pamukun'ikríhar* as the subject, but it is a patient or theme, the object which results from the making event, not the agent of that event.

(44) xás pa-mukun-'ikríhar uum taskana-tunvêech-as u-kyâar-ahi-tih
and the-3PL.POSS-fish.trap 3SG.PRO pole-little-PL 3SG-make-ESS-DUR
'And their fish-trap was made of little poles.' (Julia Starritt, WB_KL-69:6)

(45) xás kári pa-'avansáxiich aah ú-kyav ikmaháchraam
and then the-boy fire 3SG-make sweathouse
'Then the boy made a fire in the sweathouse.' (Julia Starritt, WB_KL-32:40)

If the theme is a plural in *-ahi* sentences, the subject agreement prefix on the verb reflects that plurality, which I take as evidence that the patient is in the subject in these cases. For example, in (46), the verb is *kuniyxôorarivahitih*, with the plural subject prefix *kun-*. This one is not an unexpected plural, because the subject is Karuk women, though the subject is unexpressed; this text is about the kind of clothing Karuk people wore traditionally. The

[10]The online corpus includes texts not included in the treebank; 8 of the sentences including -ahi are from such texts, but they likewise have the property of no expressed agents.

root of the verb is *iyxôorariv*, meaning 'to cover with' and is a transitive verb with the agent being a person who covers something with something else. In the case of (46), with *-ahi* on the verb, the women are the people being covered, not necessarily those doing the covering, and as such are a theme. Nonetheless, their plurality is reflected by the agreement marker. (47) shows an example of this verb without *-ahi*, where the subject is instead the coverer, the agent.

(46) xás vaa vúra kun-iyxôorariv-ahi-tih
and that INTNS 3PL-cover.with-ESS-DUR
'And they (Karuk women) were just wrapped in it (deerskin dress).' (Julia Starritt, WB_KL-86:3)

(47) kári xás pa-mú-vaas xás vaa u-yxôorariv.
and then the-3SG.POSS-blanket and thus 3SG>3-cover.with
'And she covered him with her blanket.' (Nettie Ruben, WB_KL-50:18)

Counterexamples

The *-ahi* examples of the preceding section are typical cases of *-ahi* exhibiting agent demotion. These cases all involve transitive verbs, but some sentences appear to be counterexamples to this agent demotion use. Typically, they involve *-ahi* being added to intransitive verbs, and their function is not understood. A few such examples are included here, though I will not be able to explain the use of *-ahi* with them.

In (48), the verb with *-ahi* is *kunkitaxríhahitih* 'they have wings,' the root of which is a verbalized form of *kitáxrih*, the noun meaning 'wing.' As such, it is not transitive and it appears to be a stative predicate which would not have an agent. It is unclear what role *-ahi* plays in this sentence, since there is no agent at any point which could have been demoted.

(48) kári xás asvúut kun-ivyiih-rishuk, kun-kitaxríh-ahi-tih
and then ant 3PL-come-out 3PL-have.wing-ESS-DUR
'And ants come out, they had wings.' (Mamie Offield, WB_KL-05:80)

Likewise, in (49), the verb *upuhyîimahitih* has *-ahi* added to the root *puhyiim*, meaning 'to rise.' This verb is only ever used to describe the rising of water, and there are only 2 examples of it in the corpus. The verb of the other example, shown in (50), does not have *-ahi*, but seems roughly to have the same sort of meaning as (49).

(49) xás ta'ítam p-eeshkêesh u-puhyîim-ahi-tih
and then the-river 3SG-rise.high-ESS-DUR
'The river was at the high-water mark.' (Chester Pepper, WB_KL-51:12)

(50) vaa uum pay'ôok p-eeshkêesh mit u-púhyiim.
thiss 3SG.PRO here the-river PST 3SG-rise.high
'This is where the water rose up to.' (KV)

In these cases the verb is clearly intransitive, and water rising is not obviously a case of an event with an agent. As such, the function of *-ahi* in this case is as mysterious.[11]

Despite these open questions about the use of *-ahi* with intransitives, the point of this section remains. *-ahi* is used for agent demotion at least in certain cases, but its properties are different from the use of unexpected plural agreement, specifically in that the cases of *-ahi* lack any expression of agents, whereas the unexpected plural cases can, albeit rarely in non-obviative cases, express the agent.

To summarize, let us return to the three criteria for canonical passives laid out by Legate (2021): agent demotion, theme promotion, and morphological marking. The *-ahi* construction exhibits all three: agents do not appear as subjects with *-ahi* marked verbs, themes act as subjects with them (controlling agreement) and there is specific morphological marking (the suffix *-ahi*. The unexpected plural agreement construction, in contrast, only conclusively exhibits one of the criteria, that of morphological marking.

We could speculate further on what this difference means for the syntactic structure of these sentences. Perhaps *-ahi* affects the valence of the verb, like a true passive, and grammatically demotes the agent (even disallowing for by-phrases), as might be expected since it exhibits all three of Legate (2021)'s canonical passive criteria. Perhaps the use of the unexpected plural does not involve actual syntactic demotion of the agent, rather doing something more pragmatic. But, on the basis of the evidence gathered here, it is difficult to say conclusively. The appearance of subjects with the unexpected plural cases is not necessarily an argument against their grammatical demotion. Their rare appearance could be due to their being the equivalent of an English by-phrase, and perhaps the obviative *îin* is even acting much like *by* does. There does not seem to be the sort of evidence necessary in this corpus data to adjudicate between the two possibilities, and as such I will cease the speculations here.

4.4 Unexpected singular agreement

In this section, I examine the phenomenon of unexpected singular agreement. These cases involve the use of the agreement prefix *u-*, described as being used for third person singular subjects, with subjects known to be plural. Knowing if a given subject is plural is complicated in Karuk by the morphological "optionality" of plural marking on nouns and its restriction to only certain types of nouns, and as such some of the reasoning behind identifying a subject as plural is similar to the reasoning used in the unexpected plural section in identifying that a subject was singular. Namely, understanding driven by the context of the story the

[11] Peter Jenks (p.c.) notes that the pair in (49) and (50) feel similar to cases where middle voice marking can be used to indicate the inchoative sentence in a causative alternation as in Romance languages. Labelle (1992) describes how in French, some inchoatives have a reflexive or middle marking (i.e. *Le vase se casse* 'The vase breaks') and some lack this marking (i.e. *Le vase casse* 'The vase breaks') (p.375), though they seem to have similar meaning. Perhaps the alternation between (49) and (50) is a similar sort of alternation as the French inchoatives, but I leave this possibility to further research.

example is found in, or a change in agreement where a subject previously agreed with as plural is agreed with as singular. However, there are also examples with explicit plural marking that nonetheless have singular agreement. We can be certain in these cases that there is a mismatch, and they will form the core of the examples discussed here.

Animacy is implicated in nominal plurality. Bright (1957, p.81) notes that plural marking on nouns is restricted to animates (which he calls "personal nouns") and, surprisingly, to adjectives. For the case of unexpected singular agreement, both animates and inanimate nominals are found, with a skew towards inanimates. The only absolutely certain cases, those with explicit plural marking of some form, are only found with inanimates, however; cases with nominal plural marking involve a compound of an adjective and the inanimate noun. Most cases of animates, on the other hand, are potentially amenable to a reanalysis whereby there is no mismatch. The focus in this section will be those sentences with some explicit plural marking on subjects which nonetheless display singular subject agreement. Like in the section above about unexpected plural agreement, the point here is to showcase the ability of the treebank to find rare examples of an interesting phenomenon, examples which had until the treebank remained undiscovered. It is hoped the characterization of this phenomenon provided here can serve as the basis for future analysis.

Unlike the phenomenon of unexpected plural agreement, to my knowledge there has never been a mention of this phenomenon of unexpected singular marking in previous studies of Karuk. This is a further showcase of the utility of the treebank to find rare phenomena even in corpora as well-studied as the Karuk one.

4.4.1 Explicitly plural inanimates

Plurality as a category is exponed in several distinct ways in Karuk. Nominal plural marking involves the use of a suffix *-as*[12] Explicit number words and quantifiers can combine with nouns that have plural marking and nouns which do not. The latter is more common, owing to the rarity of plural marking in general; Bright (1957) notes that the plural marker is only "usually only optionally present in those environments where it occurs" (p. 81). There is, of course, plurality indicated by the agreement prefixes on verbs. Additionally, certain verb roots will indicate the plurality of their subjects; the verb *ikrii* 'to live' is reportedly used for singular subjects, while its semantic counterpart *iin* is used for dual subjects, and *araarahi* is used for plurals above two.

Before getting into the details of examples where singular verbal agreement is combined with explicit plural marking elsewhere, the following data will show the more typical case: the types of plural marking mentioned above matching the agreement on the verb. (51) showcases a typical example of a plural-marked noun triggering plural verbal agreement. The subject, *pa'ávansas*, is marked with the plural suffix *-as* and the verb *kun'ípak* bears the third person plural subject agreement marker *kun-*.

[12]This *-as* suffix has the allomorphs *-sas* and *-sa* as well.

(51) xás kun-'ípak pa-'ávansa-s
and 3PL-return the-man-PL
'Then the men returned.' (Julia Starritt, WB_KL-21:16)

(52) shows that a plural marked noun and adjective compound can trigger plural agreement on the verb. Given that such compounds are common cases of unexpected singular agreement, it is important to thus show that plural agreement is actually possible in such cases. In (52), the subject is *pa'axchaytunvêechas*, 'little ground squirrels,' comprised partially of the noun *axchay* 'squirrel' compounded with the adjective *tunvêech* 'small' and plural marked. The verb *kuniktírish* is marked with the third person plural subject agreement marker *kun-*.

(52) kári xás tá kun-iktírish pa-'axchay-tunvêech-as
and then PER 3PL-faint the-squirrel-small-PL
'And the little ground squirrels fainted.' (Nettie Ruben, WB_KL-46:9)

(53) showcases a typical case of an explicit plural number word being used as a quantifier, *áxak* 'two.' The verb *kun'iruvêehriv* bears the third person plural subject agreement marker *kun-*. As was mentioned to be possible above, the noun of the subject, *avansa*, does not bear plural marking.

(53) ishkêesh-ak xákarari áxak ávansa kun-'iruvêehriv
river-LOC on.both.sides two man 3PL-stand
'Two men are standing on each side of a river.' (Julia Starritt, WB_KL-92:90)

(54) exhibits the universal quantifier *koovura* in a typical case, where the verb displays plural agreement. The subject, *koovúra pá'aah* is quantified by *koovura*, and the verb *kunímshiipva* bears the third person plural subject agreement marker *kun-*.

(54) xás tá'itam koovúra pá-'aah kun-imshiip-va ôokninay
and then all the-fire 3PL-cool.off-PL.ACT around.here
'And so all the fire went out around here.' (Julia Starritt, WB_KL-10:9)

(55) exhibits the use of an inherently plural verb root, matched with plural agreement. The inherently plural verb root is *ithvirip* 'to run (used of two animates)' and it is indeed marked with the expected third person plural subject agreement marker *kun-*.

(55) xás pa-mú-chaas xákaan sáruk kun-ithvírip-fak.
and the-3SG.POSS-younger.brother with downhill 3PL-run.(two)-downhill
'And he and his younger brother ran downhill.' (Julia Starritt, WB_KL-32:44)

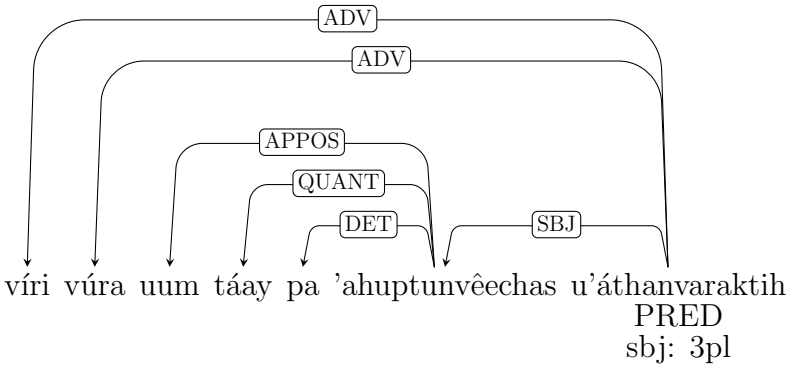
The preceding examples show that these various types of plurality can and do align with each other. However, as the examples below make clear, these types of plurality do not necessarily align. For example, a plural suffix on a noun does not guarantee that said noun triggers plural agreement on the verb. This is the case for the first class of inanimate examples I present here.

Explicitly plural nouns

The four examples below involve the following properties: the subject noun is marked with the plural suffix *-as*, and the verb of which said noun is the subject is marked with the agreement prefix *u-*, indicating a third-person **singular** subject. Interestingly, these examples are also all inanimate. Earlier, I mentioned that inanimates are not able to be plural marked in Karuk, but there is a way around this restriction. Adjectives are able to be plural marked, and so an inanimate noun compounded with an adjective is able to bear plural marking. This is the case for each of the sentences below.

In (56), the subject is *pa'ahuptunvêechas*, 'little sticks.' The noun part of the compound is *ahup*, stick, which is compounded with the postpound adjective *-tunveech*[13] which bears the plural suffix *-as*. As such, we know the subject is plural, at least in whatever sense the plural suffix cares about. The verb in this sentence is *u'áthanvaraktih* ' it is floating downriver,' with the *u-* prefix that is expected to indicate a singular subject. So, this amounts to a plural subject triggering singular agreement.

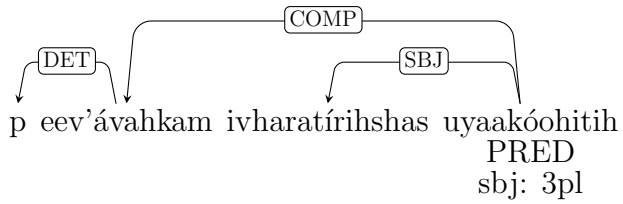
(56) víri vúra uum táay pa-'ahup-tunvêech-as u-'áthanvarak-tih
so INTENS 3SG.PRO much the-wood-small-PL 3SG-float.downriver-DUR
'There were a lot of little sticks floating down from upriver.', (Chester Pepper, WB_KL-03:92)



(57) exhibits the same phenomenon. The subject, *ivharatírihshas* 'wide boards,' is built out of a compound of *iivhar* 'board' with the adjective *tirih* 'wide, flat.' This compound is marked with the plural *-shas* (with palatalization of the initial *s* by phonological rule). Despite the explicit plural marking, the verb's agreement prefix is *u-*, indicating a singular subject.

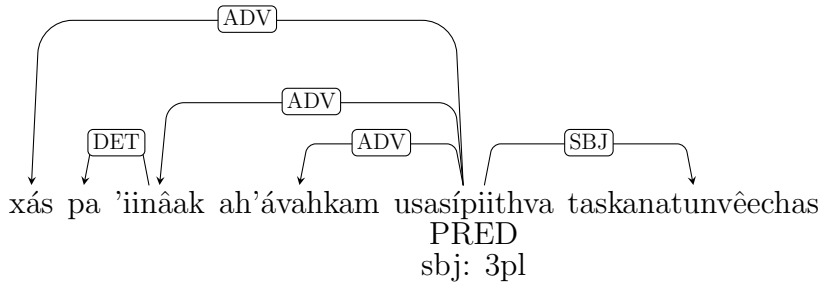
[13]Technically, the form *-tunveech* is only found in plurals. The singular form is *-tunviiv*.

(57) p-eev'ávahkam ivhara-tírih-shas u-yaakóo-hi-tih
the-roof board-wide-PL 3SG-put.(several).on-ESS-DUR
'And broad boards were put on the roof.', (Julia Starritt, WB_KL-77:9)



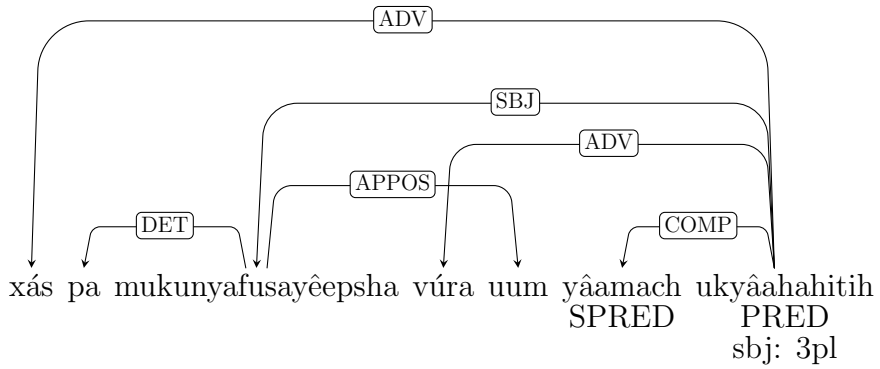
In (58), the subject is *taskanatunvêechas*, 'little poles,' made up of *táaskar* 'pole' compounded with *tunvêech* 'little.' It is marked with the plural suffix *-as*, but the verbal agreement prefix is nonetheless *u-*, for a singular subject.

(58) xás pa-'iinâak ah-'ávahkam u-sasípiithva taskana-tunvêech-as
then the-indoors fire-over 3SG-spiral.around pole-small-PL
'And on the inside, above the fire, little poles were stretched around.', (Julia Starritt, WB_KL-77:13)



(59) is the last example found in the treebank corpus with an explicit plural subject and singular agreement. The subject is *pamukunyafusayêepsha*, built from a compound of *yafus* 'dress' and the adjective *yêep* 'good' plus the plural *-sha* (with the initial *s* palatalized). Nonetheless, the verb has the singular agreement prefix *u-*.

(59) xás pa-mukun-yafusa-yêep-sha vúra uum yâamach u-kyâah-ahi-tih
then the-3PL.POSS-dress-good-PL INTENS 3SG.PRO pretty 3SG-make-ESS-DUR
'And their good dresses were made pretty.', (Julia Starritt, WB_KL-86:5)



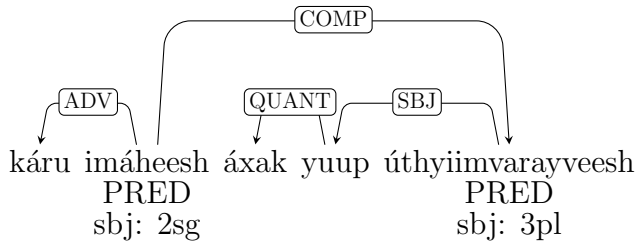
Unmarked noun with explicit number

There are also cases where a noun, which lacks plural marking, co-occurs with an explicit number word like *áxak* 'two' and nonetheless triggers the singular agreement prefix *u-*. Interestingly, there are no cases of a number word co-occuring with a plural-marked noun and singular agreement. There are cases of plural marking co-occuring with an explicit number word, as in the sentence in (60); so there is not a restriction on using a number word with a plural-marked noun. It is more likely to be an artifact of the rarity of the phenomena involved; perhaps a larger corpus would contain such examples.

(60) kári xás ú-kmar áxak ifápiit-shas
and then 3SG-meet two girl-PL
'And he met two young women.' (Mamie Offield, WB_KL-09:2)

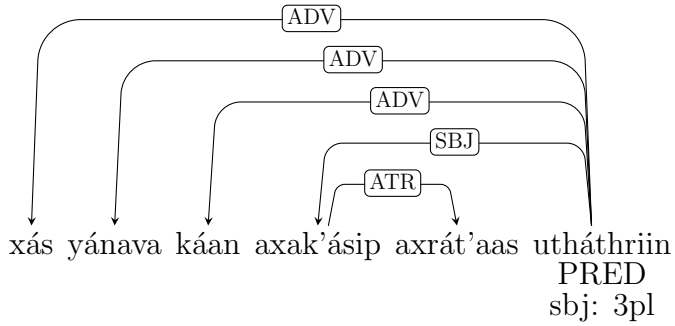
There are only two examples of the unmarked noun with explicit number and singular agreement phenomenon in the corpus. The first is in (61) below; the subject is *áxak yuup* 'two eyes.' The noun *yuup* lacks plural marking, as expected given that it is inanimate. Despite the explicit number word *áxak* 'two,' the verb bears the *u-* agreement marker for singular subjects.

(61) káru i-máh-eesh áxak yuup ú-thyiimvarayv-eesh
also 2SG>3-see-PROSP two eye 3SG-float.around.(two)-PROSP
'And you will see two eyes float around.', (Mamie Offield, WB_KL-58:14)



(62) is the second and final example of this phenomenon in the corpus: the subject is *axak'ásip* 'two bowls,' with no plural marking and the explicit number word *áxak*, but the verb nonetheless is marked with the singular subject agreement prefix *u-*. The verb is also an inherently plural verb root, *thathriin*, 'sit' used only of dual inanimate subjects.

(62) xás yánava káan axak-'ásip axrát-'aas u-tháthriin
then visible there two-bowl gooseberry-water 3SG-sit.(two.things)
'And he saw two baskets of berry juice sitting there.', (Julia Starritt, WB_KL-04:57)



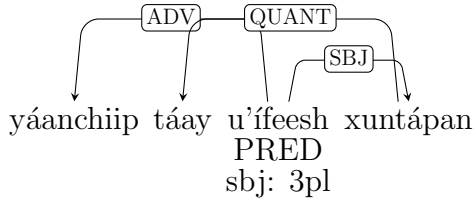
Unmarked noun with quantifier

There are also several examples in which a quantifier co-occurs with a subject which lacks plural marking, and yet the verb bears the singular agreement prefix *u-*. The two quantifiers found in these examples are *taay* 'many' and *koovura* 'all,' with two examples of each.

(63) and (64) are examples of quantified nouns with *taay* 'many' triggering singular agreement. In (63), the subject is *táay páxaath* 'lots of grasshoppers' and in (64) it is *taay ... paxuntápan* 'lots of acorns.' (64) involves a split noun phrase; see the Quantifier section in the annotation guidelines 2.7.5 for more discussion. Both of these examples the verb is marked with the 3rd person singular agreement prefix *u-*, though in (63) the agreement prefix shows up as *oo* due to coalescence with the vowel of the preceding particle *ta*.

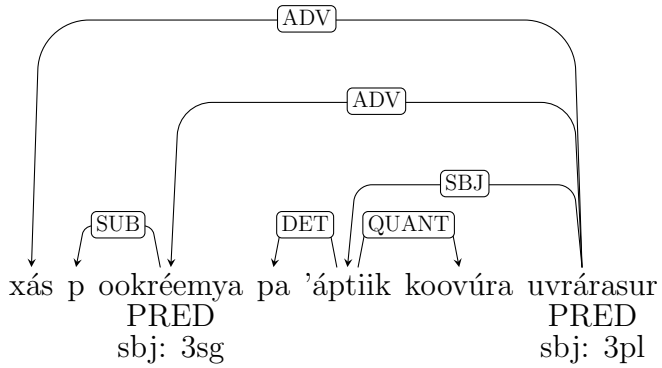
(63) yánava vúra táay pá-xaath t-óo mtupíshriih-va xás vúra amáyav kumish
visible INTENS much the-grasshopper PER-3SG be.cooked-PL.ACT then INTENS good-tasting sort.of
'He saw lots of grasshoppers cooked, and they were sort of good-tasting.', (Julia Starritt, WB_KL-04:31)

(64) yáanchiip táay u-'íf-eesh xuntápan
next.year much 3SG-rise-PROSP acorn
'The next year many acorns will grow.', (Chester Pepper, WB_KL-47:18)

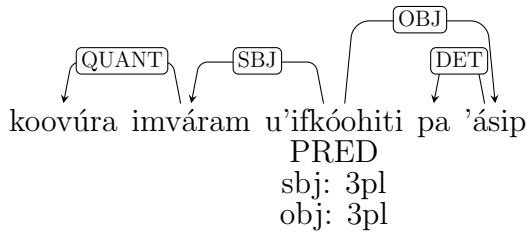


(65) and (66) involve the quantifier *koovúra* 'all.' In (65) the subject is *pa'áptiik koovúra* 'all the branches' and in (66) it is *koovúra imváram* 'all plate-baskets.' In both cases, the verb bears the *u-* agreement prefix for 3rd person singular subjects.

(65) xás p-oo-kréemya pa-'áptiik koovúra u-vrárasur
then SUB-3SG-blow the-branch all 3SG-fall.off.(several)
'And when it blew, the branches all fell off.', (Daisy Jones, WB_KL-20:32)



(66) koovúra imváram u-'ifkóo-hi-ti pa-'ásip
all plate 3SG-fit.on-ESS-DUR the-bowl
'And the plate-baskets (for the salmon) fit into the soup baskets.', (Nettie Ruben, WB_KL-74:22)

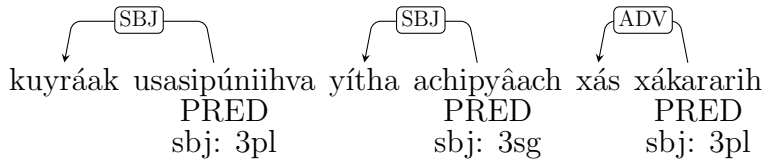


Plural number word subject

One example is found where the subject is itself a plural number word, but nonetheless triggers the singular agreement prefix *u-* on the verb. In (67), the subject of interest is *kuyraak* 'three', and the verb *usasipúniihva* 'to be in a line running down' is marked with the *u-* prefix.

(67) kuyráak u-sasip-úniih-va yítha achipyâach xás xákararih
three 3SG-be.in.line.(several)-down-PL.ACT one very.center then on.both.sides

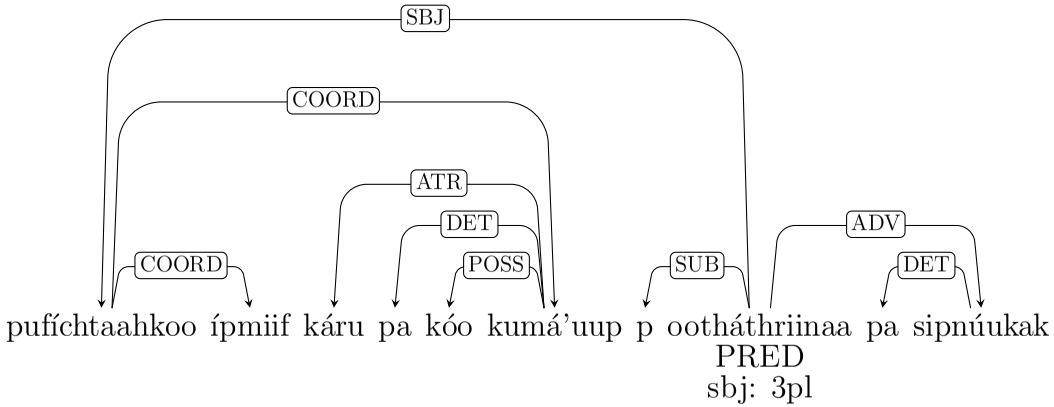
'There were three stripes running down, one right in the middle and (two) on each side.', (Julia Starritt, WB_KL-87:2)



Plural verb root

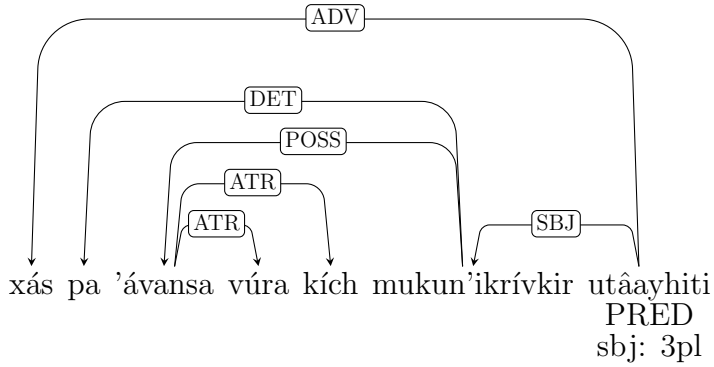
As mentioned above, some verbs in Karuk have plurality baked into the meanings of their roots, and exist in a paradigmatic relationship with singular verbs of the same meaning. One such verb is *thathriinaa* 'to be, sit (used of several filled containers)', found in (68). The singular counterpart is to this verb is *iithri*, and the dual counterpart is *thathriin*. Despite the inherent plurality of this verb, it is only ever found with the singular agreement prefix *u-* as in (68).

(68) pufíchtaahkoo ípmiif káru pa-kóo kumá-'uup p-oo-tháthriinaa pa-sipnúuk-ak
white.deer black.deer also SUB-all 3SG.POSS-possession SUB-3SG-sit.(two.things) the-storage.basket-LOC
'White deerskins, black deerskins, and every kind of treasure sat in the storage baskets.', (Nettie Ruben, WB_KL-57:118)



There are other verbs which implicate plurality in their meanings which do not necessarily exist in a paradigmatic relationship with other verbs of similar meaning. The verb *taayhi* means 'to be many' and is built from the quantifier *taay*, mentioned above, with the verbalizing suffix *-hi*. This verb occurs with the 3rd person singular agreement marker *u-*, even though its meaning should necessarily assert the plurality of its subject.

(69) xás pa-'ávansa vúra kích mukun-'ikrívkir u-tâayhi-ti
then the-man INTENS only 3PL.POSS-disk.seat 3SG-be.many-DUR
'And only the men's seats were there.', (Julia Starritt, WB_KL-77:16)



Remarks

I have described in the above sections those examples where explicit plurality of various kinds is found with singular agreement. There are many more examples where the plurality of an inanimate argument can be deduced from context and co-occurs with singular agreement, but I have opted to focus on the above cases due to the impossibility of reanalyzing them as (semantically) singular.

Why should the agreement prefix *u-* be used so readily with subjects that are explicitly plural? The possibility of *u-* being some sort of general underspecified agreement, perhaps which does not require or expone any features, seems a promising angle, though one would need to make sure that the system allows for the correct underspecification of both *u-* and *kun-*, since both the 3rd person singular and plural agreement prefixes are able to be used with subjects that bear the opposite number value. I will not work out the details of such an analysis here; the point here is to showcase that these examples, which eluded manual searches of the Karuk corpus, are discoverable by the treebank. Any theory of Karuk agreement should have to account for the patterns discussed here, patterns not discovered by either manual corpus searches or prior elicitation. It may also be that the various ways to mark plurality are not actually concerned with the same features or meaning; there may be multiple 'pluralities' at play. Garrett and Maier (2022) hypothesize, for instance, that

inherently plural verb roots are not concerned with nominal number per se, but with how events are conceptualized.

The inanimacy of the subjects in this section must be of some importance, since the patterns found above do not hold for animate subjects with unexpected singular agreement. It is suggestive that inanimates both are unable to directly combine with plural marking and often do not trigger plural agreement on verbs. Yet, as we have seen, plural marking and plural agreement do not always match and probably are not concerned with exactly the same features. It should also be mentioned that these cases of unexpected singular agreement seem to involve 'non-agentive' contexts, involving verbs that for one reason or another do not have agents, whether through being unaccusative verbs or through the use of the *-ahi* passive.[14] Given that inanimates are unlikely to be agents, if the unexpected singular agreement somehow comes about because of the inanimacy of the subject, that could also explain the seeming restriction to non-agentive contexts, as those non-agentive contexts are the only ones in which inanimates could be subjects at all.

In the next section, I briefly discuss some examples of potentially unexpected singular agreement with animate subjects.

4.4.2 Animates

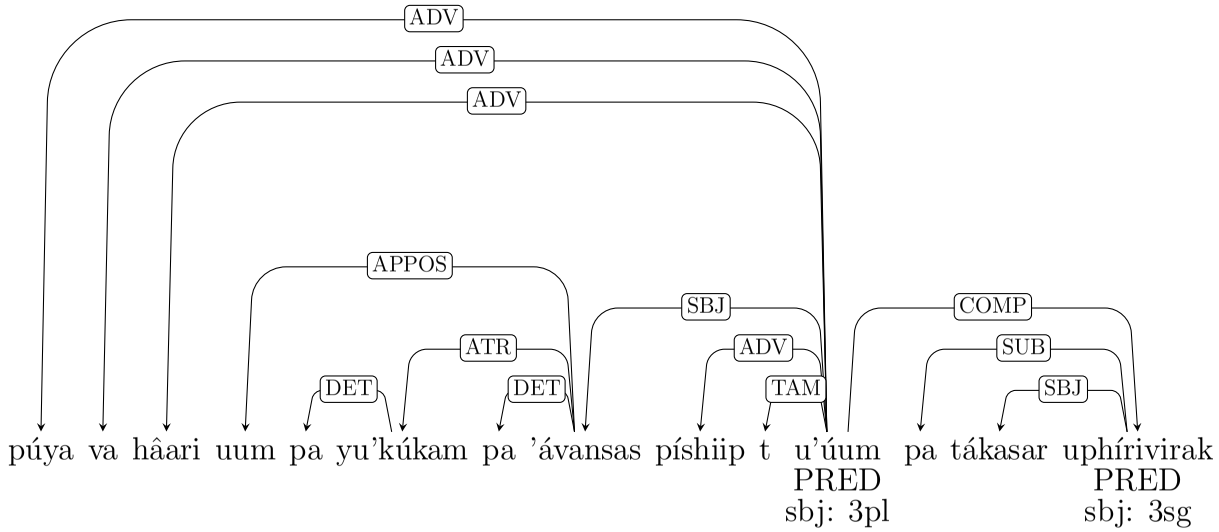
The treebank returned a decent amount of sentences where a putatively plural animate subject was paired with a verb bearing singular subject agreement. However, most of these cases are amenable to some other analysis on which the subject is either not actually plural, or the subject is not actually what was annotated. The examples are too many to discuss individually here, though interested readers can find lists of all these examples in Appendix B. One thing can be said for certain: there are no cases of animate subjects with singular agreement where the plurality of the subject is explicit or impossible to reanalyze, unlike the cases of plural inanimates discussed above. I discuss one example of a mismatched plural animate below, meant to show the kinds of sentences found in this search.

In (70), the noun phrase which was assigned the subject relation to the verb *tu'úum* 'arrived' is *pa'ávansas* 'the men.' It is explicitly marked as plural, yet the verb *tu'úum* bears the singular agreement prefix *u-*. This would be the only case where an explicitly plural animate subject is paired with singular agreement, but there is a confound. The whole noun phrase of the subject includes another nominal: *payu'kúkam* 'the downriver side.' Cases where nouns are directly modified by other nouns in this way are rare in the corpus, and the annotation of *pa'ávansas* as the subject is based on the principle of following the English translation: if there are two alternatives the one which fits the translation will be chosen. However, that is no real argument that the head of the subject noun phrase is not actually *payu'kúkam*, a singular noun, and *pa'ávansas* is a modifier or appositive. An analysis that posited *payu'kúkam* 'the downriver side' (meaning something like 'the ones on the downriver side') as the subject, with *pa'ávansas*as an appositive phrase, seems in principle possible.

[14]Thanks to Peter Jenks for pointing this out to me.

At the given moment I am not sure what evidence would militate against either analysis in favor of the other. Given that there are no other cases where an unambiguously plural animate subject co-occurs with singular agreement on the verb, that may be an argument in favor of *payu'kúkam* being the subject and this sentence not involving any unexpected agreement at all.

(70) púya-va hâari uum pa-yu'-kúkam pa-'ávans-as píshiip t-u-'úum pa-tákasar
and-so sometime 3SG.PRO the-downriver-side the-man-PL first PER-3SG-arrive SUB-shinny.tossel
u-phíriv-irak
3SG.-lie.(two)-where
'Sometimes the men on the downriver end arrived first where the tossel lay.', (Julia Starritt, WB_KL-78:25)



Chapter 5

Conclusion

In this dissertation, I have introduced the Karuk treebank and described in detail the nature of its annotations, and in two sets of case studies, I showcased some of the utility the treebank can offer. In short, the treebank offers, for an upfront time and effort cost, the ability to more quickly access both large-scale quantitative data, as showcased in Chapter 3: Argument and Predicate Order, and much smaller scale rare data for phenomena that might only occur a handful of times in an entire corpus, as showcased in Chapter 4: Agreement. In both cases, the treebank enabled new information to be uncovered: in chapter 3, I described in detail the prevalance of a variety of clause types, and in chapter 4 I detailed a heretofore undescribed subject demotion phenomenon signalled by a mismatch in agreement. The treebank is not built on a new corpus; the texts within have been accessible since the publication of William Bright's grammar (Bright 1957) and have been rendered even more accessible by the morphological annotation added to these texts in *Ararahih'urípih*, the online dictionary and text corpus. The information uncovered by using the treebank could perhaps in principle have been uncovered without the creation of a treebank, but would have taken a difficult commitment to combing through the entire corpus and nevertheless be susceptible to overlooking data, as was discussed in Chapter 4. With a treebank, a concerted and systematic effort to annotate the corpus allows you to save time and ensure comprehensiveness for any future study which bears on questions the treebank's annotation are relevant for.

The treebank, nonetheless, could doubtlessly yet be improved, and there are two major directions for the treebank's future. First, though the treebank's corpus is considerable for a small language like Karuk, it does not cover the entire documentary corpus of Karuk; indeed, even exempting unpublished documentary material, of which there is plenty, there are many published Karuk texts that are not yet part of the treebank's corpus. Particularly, the texts recorded by J.P. Harrington (Harrington 1930; Harrington 1932b; Harrington 1932a) and the texts recorded by Jaime de Angulo and L.S. Freeland (de Angulo and Freeland 1931) would prove a useful addition, and of course, the vast unpublished documentary corpus described in Chapter 1 would no doubt increase the size of the treebank considerably. Such an increase would enable, of course, increased confidence in results from the treebank and the ability to find as-of-yet undescribed phenomena in the corpus; further, the addition of material

from earlier (and later) periods of time may allow the treebank to be useful for uncovering syntactic change.

Second, as of the current moment the ability to search the treebank relies on the ability to write python scripts. One of the most important benefits of the morphological annotation in *Ararahih'urípih* is that it is accessible and searchable through an easy-to-understand online interface, thus enabling members of the Karuk community and outside researchers to quickly search a particular word or morpheme and see many examples of its use. Though abstract syntactic information is inherently less accessible than words or even morphemes, I believe there could be a real benefit for Karuk teachers or advanced learners to be able to easily find examples of, say, sentences where the subject is post-verbal. Thus, the integration of the treebank's annotation into *Ararahih'urípih*'s search interface is a high priority. Unfortunately, the scope of such an integration is quite large and as such was not possible within the scope of this dissertation, which already had taken on the large task of creating the treebank itself. The amount of queries the treebank makes possible is very large, and it is not particularly obvious how one could recreate the ease of a text-based search box for something like searching for, as an arbitrary possible example, a sentence with a subject omitted and complement present and pre-verbal. There may be other ways to integrate information gleaned from the treebank into the existing framework of *Ararahih'urípih*, such as including the most common subjects and objects of a verb into that verb's dictionary entry. Whatever the method, it is hoped that the treebank can help enrich understanding of the Karuk language for the Karuk community, and allow more people to access the rich and beautiful stories in the Karuk corpus.

Bibliography

Bamman, David and Gregory Crane (2011). "The Ancient Greek and Latin Dependency Treebanks". In: *Language Technology for Cultural Heritage*. Ed. by Caroline Sporleder, Antal van den Bosch, and Kalliopi Zervanou.

Béjar, Susana (2003). "Phi-syntax: A theory of agreement". PhD thesis. University of Toronto.

Béjar, Susana and Milan Rezac (2009). "Cyclic Agree". In: *Linguistic Inquiry* 40.1, pp. 35–73.

Bobaljik, Jonathan David and Dianne Jonas (1996). "Subject positions and the roles of TP". In: *Linguistic Inquiry* 27.2, pp. 195–236.

Bright, William (1952). "Linguistic Innovations in Karok". In: *International Journal of American Linguistics* 43, pp. 53–62.

— (1957). *The Karok Language*. University of California Press.

— (1958). "Karok Names". In: *Names* 6, pp. 172–179.

— (1977). "Coyote Steals Fire (Karok)". In: *International Journal of American Linguistics Native American Texts Series* 2.2, pp. 3–9.

— (1978). "Karok". In: *Handbook of the North American Indians*. Ed. by Robert F. Heizer. Vol. 8.

— (1979a). "A Karok Myth in "Measured Verse":The Translation of a Performance". In: *JCGBA* 1, pp. 117–123.

— (1979b). "Toward a Typology of Verbal Abuse: Naming Dead Kin in Northwestern California". In: *Maledicta* 3, pp. 177–180.

— (1980a). "Coyote Gives Acorns and Salmon to Humans". In: *Coyote Stories II*. Ed. by Martha B. Kendall. Vol. 6. IJAL Native American Texts Series. University of Chicago Press, pp. 46–52.

— (1980b). "Coyote's Journey". In: *American Indian Culture and Research Journal* 4, pp. 21–48.

— (1984). *American Indian Linguistics and Culture*. Berlin: Mouton de Gruyter.

Bright, William and Jane O. Bright (1965). "Semantic Structures in Northwestern California and the Sapir-Whorf Hypothesis". In: *American Anthropologist* 67, pp. 249–258.

Bright, William and Susan Gehr (2004). *Karuk Dictionary*. Karuk Tribe of California.

Brugman, Claudia and Monica Macaulay (2009). "Relevance, cohesion, and the storyline: The discourse function of the Karuk particle káruma". In: *Journal of Pragmatics* 41, pp. 1189–1208.

Brugman, Claudia and Monica Macaulay (2015). "Characterizing evidentiality". In: *Linguistic Typology* 19.2, pp. 201–237.

Burga, Alicia, Alp Öktem, and Leo Wanner (2017). "Revising the METU-Sabanci Turkish Treebank: An Exercise in Surface-Syntactic Annotation of Agglutinative Languages". In: *Proceedings of the Fourth International Conference on Dependency Linguistics*, pp. 32–41.

Campbell, Amy M. (2012). "The Morphosyntax of Discontinuous Exponence". PhD thesis. University of California, Berkeley.

Campbell, Lyle (1997). *American Indian languages: The historical linguistics of Native America*. Oxford University Press.

Carpenter, Kayla (2014). "Aspectual Future in Karuk". Unpublished manuscript, UC Berkeley.

Conathan, Lisa (2004). "The Linguistic Ecology of Northwestern California: Contact, functional convergence and dialectology". PhD thesis. University of California, Berkeley.

— (2006). "Recovering sociolinguistic context from early sources: The case of northwestern California". In: *Anthropological Linguistics* 48.3, pp. 209–232.

Conathan, Lisa and Esther Wood (2003). "Repetitive reduplication in Yurok and Karuk: Semantic effects of contact". In: *Papers of the 34th Algonquian conference*. Ed. by H.C. Wolfart.

Cook, Sherburne F. (1956). "The aboriginal population of the north coast of California". In: *University of California Anthropological Records* 16, pp. 81–129.

Curtin, Jeremiah (1889a). *Ehnikan (Karok) vocabulary, June-July 1889, NAA MS 847*. National Anthropological Archives, Smithsonian Institution.

— (1889b). *Fifteen untitled Karok stories, May 1889, NAA MS 269*. National Anthropological Archives, Smithsonian Institution.

— (1889c). *Karok vocabulary, June and July 1889, NAA MS 1450*. National Anthropological Archives, Smithsonian Institution.

Davis, Charron (Sonny), Vina Smith, Nancy Super (née Jerry), Peter Super Sr., Charlie Thom Sr., and Line Mikkelsen (2020). "Forms and functions of backward resumption: The case of Karuk". In: *Language* 96.4 (December), pp. 841–873.

de Angulo, Jaime and L. S. Freeland (1931). "Karok Texts". In: *International Journal of American Linguistics* 6, pp. 194–226.

Dixon, Roland B. and A.L. Kroeber (1913). "New linguistic families in California". In: *American Anthropologist* 15.4, pp. 647–655.

Dryer, Matthew S. (2013). "Order of Subject, Object, and Verb". In: *World Atlas of Linguistic Structure Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: `http://wals.info/chapter/81`.

Du Bois, John W. (1987). "The discourse basis of ergativity". In: *Language* 63.4, pp. 805–855.

Garrett, Andrew, Susan Gehr, Erik Hans Maier, Line Mikkelsen, Crystal Richardson, and Clare S. Sandy (in press). "Karuk". In: *Handbook of North American Languages*. De Gruyter.

Garrett, Andrew and Erik Hans Maier (2022). "How to do things with birds: The pragmatics of lexical and syntactic choice in early Karuk narrative texts". Talk given at University of California, Berkeley's Syntax and Semantics Circle, February 11th, 2022.

Gibbs, George (1853). "Vocabularies of Indian languages in northwest California". In: *Historical and statistical information respecting the history, condition, and prospects of the Indian tribes of the United States*. Ed. by Henry R. Schoolcraft. Vol. 3, pp. 428–445.

Givón, T., ed. (1983). *Topic Continuity in Discourse: Quantitative Cross-Language Studies*. Amsterdam: Benjamins.

Golla, Victor (2011). *California Indian Languages*. University of California Press.

Haas, Mary R. (1964). "California Hokan". In: *Studies in Californian Linguistics*. Ed. by William Bright. Vol. 34. University of California Publications in Linguistics, pp. 73–87.

Hajic, Jan and Barbora Hladka (2001). "The Prague Dependency Treebank: Annotation Structure and Support". In: *Proceedings of the IRCS Workshop on Linguistic Databases*.

Hale, Kenneth L. (1983). "Warlpiri and the Grammar of Non-Configurational Languages". In: *Natural Language and Linguistic Theory* 1, pp. 5–47.

Harrington, John P. (1930). "Karuk Texts". In: *International Journal of American Linguistics* 6, pp. 121–161.

— (1932a). *Karuk Indian Myths*. Bulletin 107. Bureau of American Ethnology.

— (1932b). *Tobacco Among the Karuk Indians of California*. Bultetin 94. Bureau of American Ethnology.

Haspelmath, Martin (2021). "Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability". In: *Journal of Linguistics* 57.3, pp. 605–633.

Kempen, Gerard and Karin Harbusch (2019). "Mutual attraction between high-frequency verbs and clause types with finite verbs in early positions: corpus evidence from spoken English, Dutch, and German". In: *Language, Cognition and Neuroscience* 34.9, pp. 1140–1151.

Kroeber, Alfred L. (1911). "The Languages of the Coast of California North of San Francisco". In: *UC-PAAE* 9.3, pp. 273–435.

— (1925). *Handbook of the Indians of California*. Bulletin 8. Bureau of American Ethnology.

Kroeber, Alfred L. and Edward W. Gifford (1980). *Karok Myths*. Ed. by Grace Buzaljko. University of California Press.

Kumaran, Elango (2018). "Karuk agreement". In: *Proceedings of the Twenty-Ninth Western Conference on Linguistics (WECOL)*.

Labelle, Marie (1992). "Change of state and valency". In: *Journal of Linguistics* 28, pp. 375–414.

Lang, Julian (1994). *Ararapíkva: Creation Stories of the People*. Heyday Books.

Legate, Julie Anne (2021). "Noncanonical passives: A typology of voices in an impoverished Universal Grammar". In: *Annual Review of Linguistics* 7, pp. 157–76.

Legate, Julie Anne, Faruk Akkuş, Milena Šereikaitė, and Don Ringe (2020). "On passives of passives". In: *Language* 96.4, pp. 771–818.

Macaulay, Monica (1989). "A Suffixal Analysis of the Karok "Endoclitic"". In: *Lingua* 78, pp. 159–180.

Macaulay, Monica (1990). "A preliminary look at Karuk pitch accent". In: *Papers from the 1989 Hokan-Penutian Languages Workshop. University of Oregon Papers in Linguistics*. Vol. 2, pp. 41–61.

— (1992). "Inverse Marking in Karuk: the Function of the Suffix -ap". In: *International Journal of American Linguistics* 58, pp. 182–201.

— (1993). "Reduplication and the Structure of the Karuk Verb Stem". In: *International Journal of American Linguistics* 59, pp. 64–81.

— (2000). "Obviative Marking in Ergative Contexts: The Case of Karuk iin". In: *International Journal of American Linguistics* 66, pp. 464–498.

— (2004). "On the Karuk Directional Suffixes". In: *BLS 30*, pp. 81–101.

Maier, Erik Hans (2016a). "Above or below: Modeling a telicity restriction on Karuk directional applicatives". In: *NELS 46: Proceedings of the Forty-Sixth Annual Meeting of the North East Linguistic Society*. Ed. by Christopher Hammerly and Brandon Prickett. Vol. 2. Amherst: University of Massachusetts, GLSA, pp. 317–330.

— (2016b). "Patterns of discontinuity in Karuk noun phrases". Unpublished UC Berkeley ms.

Malamud, Sophia A. (2013). "(In)definiteness-driven typology of arbitrary items". In: *Lingua* 126, pp. 1–31.

Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz (1993). "Building a Large Annotated Corpus of English: The Penn Treebank". In: *Computational Linguistics* 19.2, pp. 313–330.

Meyer, Karen Sundberg (1992). "Word Order in Klamath". In: *Pragmatics of Word Order Flexibility*. Ed. by Doris L. Payne, pp. 167–191.

Mikkelsen, Line (2014). "What goes postverbal in a verb-final language? On the interplay of prosody, information structure, and word order in Karuk". UC Santa Cruz Syntax Circle Handout.

Moorman, Karie (2015). "Polyfunctionality of *-ara* in Karuk". Undergraduate thesis.

Nivre, Joakim, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret (2007). "The CoNLL 2007 Shared Task on Dependency Parsing". In: *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*.

Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Manning. Christopher D., Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Rent Tsarfaty, and Daniel Zeman (2016). "Universal Dependencies v1: A multilingual treebank collection". In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation*.

Poser, William J. (1995). "Binary comparison and the history of Hokan comparative studies". In: *International Journal of American Linguistics* 61, pp. 135–144.

Powers, Stephen (1877). "Tribes of California". In: *Contributions to North American Ethnology* 3.

Roland, Douglas, Frederic Dick, and Jeffrey L. Elman (2007). "Frequency of Basic English Grammatical Structures: A Corpus Analysis". In: *Journal of Memory and Language* 57.3, pp. 348–379.

Sandy, Clare S. (2017). "Prosodic prominence in Karuk". PhD thesis. University of California, Berkeley.

Silver, Shirley (1974). "Some Northern Hokan plant-tree-bush forms". In: *Journal of California Anthropology* 1, pp. 102–109.

— (1976). "Comparative Hokan and the Northern Hokan languages". In: *Hokan studies: Papers from the First Conference on Hokan Languages*. Mouton, pp. 193–202.

Soegaard, Anders (2017). "What I think when I think about treebanks". In: *Proceedings of the 16th International workshop on treebanks and linguistic theories*.

Taylor, Ann (2020). "Treebanks in Historical Syntax". In: *Annual Review of Linguistics*. Online Review in Advance.

Thomas, Gullaume (2019). "Annotation Guidelines for the Mbyá Treebank". In.

Tonhauser, Judith and Erika Colijn (2010). "Word Order in Paraguayan Guaraní". In: *International Journal of American Linguistics* 76.2, pp. 255–288.

White, Whitney (2015). "Typological canonicality of Karuk agreement". Honors undergraduate thesis. University of California, Berkeley.

Yu, Kevin (2021). "Pragmatic influences on argument word order in Karuk narrative texts". Honors undergraduate thesis. University of California, Berkeley.

Appendix A

Dataset: Unexpected plural agreement

This appendix includes all sentences which the treebank search related to unexpected plural agreement returned. This appendix is intended as a way to include data that were not relevant or were excluded in the discussions above. In some cases, a treebank search would produce sentences which, though annotated correctly, were judged to not be a good example of the phenomenon under investigation. The data included here contain all the sentences which the treebank search uncovered, including data that were excluded. These data are included for the sake of transparency.

(1) vúra tá kun-'áveep pa-mu-pákurih
INTENS PER 3PL>3-take.away.from the-3SG.POSS-song
'His song had been taken away from him.', (WB_KL-07:60)

(2) âanxus kun-ipêer chími nú-vuunv-i
weasel 3PL>3-say.to soon 1>2SG-wrestle-IMPER
'Weasel was told, "Let's wrestle!', (WB_KL-18:60)

(3) xás kun-íkfuukiraa âanxus
then 3PL>3-grab weasel
'Then Weasel was grabbed.', (WB_KL-18:65)

(4) púya-va pa-kéevniikich tá kun-chífich
and.so-so the-old.woman PER 3PL>3S-beat
'Then the old woman was beaten.', (WB_KL-19:15)

(5) pa-pihnîich mú-'arama tá kun-íykar
the-old.man 3SG.POSS-child PER 3PL>3-beat

'He killed the old man's child.', (WB_KL-20:35)

(6) xás pihnîich kun-ipêer ôok naa
then old.man 3PL>3-say.to here come
'And the old man said, "Come here!"', (WB_KL-20:42)

(7) xás pa-pihnîich ú-xrar mú-'arama tá kun-íykar
then the-old.man 3SG>3-weep 3SG.POSS-child PER 3PL>3-beat
'And the old man cried, his child had been killed.', (WB_KL-20:70)

(8) xás kun-ímuusti iv'ávahkam a' p-oo-'íih-tih
then 3PL>3-look.at roof above SUB-3SG>3-dance-DUR
'And (Lizard) was looked at as he danced, up on the roof.', (WB_KL-34:35)

(9) kári xás kun-ipéer hôoy i-máh-anik
then then 3PL>3-say.to where 2SG>3-see-ANC
'And he was asked, "Where did you find it?', (WB_KL-34:42)

(10) pa-pirishkâarim tá kun-íkfuukiraa yuuxmachmahánach u-xus kíri
the-grizzly PER 3PL>3-grab lizard 3SG>3-think I.wish
ni-'ax
1s(>3)-bite
'Grizzly (in her death throes) grabbed at Lizard, she thought, "Let me kill him!"', (WB_KL-34:58)

(11) vúra kun-píychaak-tih
INTENS 3PL>3-inflict.bad.luck-DUR
'He had bad luck.', (WB_KL-35:10)

(12) xás kun-íhyiiv-ti hôoyva
then 3PL>3-shout-DUR somewhere
'And there was a shout somewhere.', (WB_KL-41:3)

(13) ithyarukpíhriiv kun-ihyûunish-tih u-pêen-tih
Across.the.Water.Widower 3PL>3-shout.to-DUR 3SG>3-say.to-DUR
na-kûush-i
2s/3s>1s-copulate-IMPER
'Across-the-Water Widower was shouted at, (the person) said to him, "Copulate with me!"', (WB_KL-41:7)

(14) kári xás áxvaay chémi xás tá koo tá kun-piikívshiip koovúra
then then Crane all.right then PER all PER 3PL>3-put.necklaces.up.on all
'And Crane (said), "All right," and that was all, he had on all the necklaces.', (WB_KL-42:9)

(15) kári xás u-píip pa-'asiktávaan pa-yaas'ára
then then 3SG>3-say the-woman SUB-rich.person
u-'iiníshrih-aak víri xáat káru tá kun-'íitshur víri-va vúra
3SG>3-come.into.existence-when so may also PER 3PL>3-leave so-so INTENS
u-pmáh-eesh pa-nini-pákuriha mûuk
3SG>3-see.again-PROSP the-1SG.POSS-song with
'And the woman said, "When Mankind comes into existence, (a woman) may also become abandoned, (but) she will find (her sweetheart) again by means of my song.', (WB_KL-49:33)

(16) xás kun-ípeen-ti ikmahachram'íshiip veekxaréeyav hûut
then 3PL>3-say.to-DUR Sacred.Sweathouse.Spirit how only
kích i-xú-tih
2SG>3-think-DUR
'And (the person) said to Sacred Sweathouse Spirit, "How are you feeling?"', (WB_KL-52:49)

(17) kári xás kun-piip asaxêevar v-eekxaréeyav xákaan chími
then then 3PL>3-say mossy 3SG.POSS-spirit.person both soon
kun-ímthaatv-eesh
3PL>3-play."stick.game"-PROSP
'And they said, Baldy Peak Spirit (said), "Let's play shinny together!"', (WB_KL-54:2)

(18) kári xás u-xus máva aaníhich tá kun-chífich
then then 3SG>3-think here! my.older.brother PER 3PL>3-beat
'And he thought, "Look, big brother's getting beaten."', (WB_KL-54:11)

(19) kári xás u-thítiv y-óo chrívchav pá-'aas pa-'úkraam
then then 3SG>3-hear visible-3SG>3 splash the-water the-lake
pa-kun-páathkuri pa-mu-típah
SUB-3PL>3-throw.into the-3SG.POSS-brother
'Then he heard it, he saw the water splash in the lake, when (the giant) threw his

brother in.', (WB_KL-55:12)

(20) chavúra koovúra tá kun-ixyákurih
finally all PER 3PL>3-throw.(pl).into.water
'Finally (the giant) threw all (the brothers) in.', (WB_KL-55:13)

(21) xás kun-ipéer chími pásas
then 3PL>3-say.to soon get.dressed
'And she told him, "Dress up (in dance regalia)!"', (WB_KL-57:73)

(22) xás kun-ipéer vúra chími pásas
then 3PL>3-say.to INTENS soon get.dressed
'And she told him, "Do dress up!"', (WB_KL-57:76)

(23) yáas u-xú-ti naa ni-xú-ti tá kun-'ípas
then 3SG>3-think-DUR 1SG.PRO 1s(>3)-think-DUR PER 3PL>3-bring.(person)
'Then she thought, "I think he's been taken."', (WB_KL-61:15)

(24) kâam kun-ikfúyvuunish
little.upriver 3PL>3-whistle.at
'He was whistled at, a little ways upriver.', (WB_KL-61:25)

(25) kun-ípeen-ti axicha-'êechkee-puh-ich yáxa í-krii
3PL>3-say.to-DUR child-kidnap-having.been.-ed-DIM look! 2SG>3-live
'He was told, "Look, you are a kidnapped child!"', (WB_KL-61:28)

(26) xás kun-ipêer hãã
then 3PL>3-say.to yes
'And he was told (by his kidnappers), "Yes.', (WB_KL-61:31)

(27) xás u-thvuyâana-ti ípa kóo kun-ípeer-at ikvan
then 3SG>3-call.by.name-DUR PAST all 3PL>3-say.to-PAST buy
'And he was naming all that they had told him to buy.', (WB_KL-66:6)

(28) xás koovúra t-u-p-ipshinvárihva p-éethvuy ípa kun-ípeer-at ikvan
then all PER-3SG>3-ITER-forget the-name PAST 3PL>3-say.to-PAST buy
'And he forgot all the names that they had told him to buy.', (WB_KL-66:11)

(29) táaskar kun-ikyâara-tih
pole 3PL>3-make.with-DUR
'It was made of poles.', (WB_KL-69:3)

(30) xás vúra puxích tá kun-ímchax xás tá kun-ástuukha
then INTENS very.much PER 3PL>3-get.hot then PER 3PL>3-sweat
'And it got very hot, and they sweated.', (WB_KL-76:15)

(31) xás páy nanu'ávahkam kúuk kin-p-ôonva
then sky to.there 3PL>3PL-ITER-take.(people)
'And they were taken to the sky.', (WB_KL-08:2)

(32) xás kin-ipêer chími kii-vyíhish
then 3PL>3PL-say.to soon 2pl(>3)-get.there.(pl.)
'And they were told, "Gather together."', (WB_KL-30:16)

(33) xás kári chavúra tá pâanpay koovúra tá kín-'ax
then then finally PER after.while all PER 3PL>3PL-bite
'And finally after a while all of them were killed.', (WB_KL-34:14)

(34) kári xás kin-'ákih amveeváxrah
then then 3PL>3PL-give.to dried.salmon
'And they were given dried salmon.', (WB_KL-58:47)

(35) víri-va itha-hárinay pu-kín-maah-tih-ap
so-so one-year NEG-3PL>3PL-see-DUR-NEG
'People didn't see him for a year.', (WB_KL-59:32)

(36) kin-ípeen-ti kêemish pa-'apxantínihich-as
3PL>3PL-say.to-DUR something.dangerous the-white.man-PL
'They were told that the white men were devils.', (WB_KL-65:4)

Appendix B

Dataset: Unexpected singular agreement

This appendix includes all sentences which the treebank search related to unexpected singular agreement returned. This appendix is intended as a way to include data that were not relevant or were excluded in the discussions above. In some cases, a treebank search would produce sentences which, though annotated correctly, were judged to not be a good example of the phenomenon under investigation. The data included here contain all the sentences which the treebank search uncovered, including data that were excluded. These data are included for the sake of transparency.

(1) kári xás kári ta'ítam u-pakurîihv-aheen
and then and so 3SG-sing.songs-ANT
'And so they sang.', (WB_KL-02a:14)

(2) koovúra pa-'áraar u-máah-vunaa-tih iimkun vúra pufáat-sa-hara
all the-human 3SG>3-see-PL-DUR 2PL.PRO INTENS nothing-PL-NEG
'He saw all the people (and said), "You-all are just nothings.', (WB_KL-03:10)

(3) t-u-'invá-kaam-ha
PER-3SG>3-forest.fire-large-DENOM
'There was a big forest fire.', (WB_KL-03:30)

(4) chavúra yiimúsich t-u-'uum
finally little.ways.off PER-3SG>3-arrive
'Finally he went a little ways.', (WB_KL-03:33)

(5) vúra t-u-'invá-kaam-ha
INTENS PER-3SG>3-forest.fire-large-DENOM

'There was a big forest fire.', (WB_KL-03:44)

(6) vúra t-u-'invá-kaam-ha
INTENS PER-3SG>3-forest.fire-large-DENOM
'There was a big forest fire.', (WB_KL-03:80)

(7) víri úuth ishkêesh-ak t-u-'ahirímkaanva
so out.to.water river-LOC PER-3SG>3-lie.across.stream
'There were trees falling out into the river.', (WB_KL-03:81)

(8) víri vúra uum táay pa-'ahup-tunvêech-as
so INTENS 3SG.PRO much the-wood-small.(pl.)-PL
u-'áthanvarak-tih
3SG>3-float.down.from.upstream-DUR
'There were a lot of little sticks floating down from upriver.', (WB_KL-03:92)

(9) víri u-vuunôovu-tih
so 3SG>3-flow.upstream.from.here-DUR
'There was an eddy.', (WB_KL-03:100)

(10) xás p-oo-pakátkat amayaa-'íshara
then SUB-3SG>3-take.a.taste good-tasting-extremely
'And when he tasted them, they were very good-tasting.', (WB_KL-03:132)

(11) xás yánava káan uxráa t-óo mtup
then visible there berry PER-3SG>3 be.ripe
'And he saw berries ripe there.', (WB_KL-04:28)

(12) yánava káan u-'íinva-hi-tih
visible there 3SG>3-forest.fire-DENOM-DUR
'He saw there was a forest fire there.', (WB_KL-04:30)

(13) yánava vúra táay pá-xaath t-óo mtupíshriih-va xás vúra
visible INTENS much the-grasshopper PER-3SG>3 be.cooked-PL.ACT then INTENS
amáyav kunish
good-tasting sort.of
'He saw lots of grasshoppers cooked, and they were sort of good-tasting.', (WB_KL-04:31)

(14) xás vúra hûutva t-u-'iin púxay
then INTENS somehow PER-3SG>3-experience.something.unpleasant not.yet
ta'ítam yâavahi-tih-ara
so get.enough-DUR-NEG
'But what was the matter with him? he wasn't getting full.', (WB_KL-04:34)

(15) yánava pá-xaath vaa vúra u-kupa-'íshipithun-ahi-ti
visible the-grasshopper so INTENS 3SG>3-MOD-run.like.string-Modal-DUR
p-oo-kupa-vúrayv-ahi-tih-een
SUB-3SG>3-MOD-go.around-Modal-DUR-ANT
'He saw the grasshoppers strung around where he had been wandering.', (WB_KL-04:36)

(16) xás yánava káan axak-'ásip axrát-'aas u-tháthriin
then visible there two-bowl gooseberry-water 3SG>3-sit.(two.things)
'And he saw two baskets of berry juice sitting there.', (WB_KL-04:57)

(17) xás káan yánava pa-mukun-patúmkir káru pa-mukun-'ikrívkir athkúrit
then there visible the-3PL.POSS-head-rest also the-3PL.POSS-disk-seat fat
u-kyâar-ahi-tih
3SG>3-make.with-ESS-DUR
'And he saw there that their pillows and their chairs were made of fat.', (WB_KL-04:127)

(18) káru yítha u-píip hôoy pa-nani-patúmkir
also one 3SG>3-say where the-1SG.POSS-head-rest
'and one said, "Where's my pillow?"', (WB_KL-04:135)

(19) púyava kúth uum p-oo-tíshraam-hi-ti panámniik
you.see because.of 3SG.PRO SUB-3SG>3-valley-DENOM-DUR Orleans
'That's why there is a flat at Orleans.', (WB_KL-04:177)

(20) kári xás u-mah yánava u-tháthriinaa axraat
then then 3SG>3-see visible 3SG>3-sit.(two.things) gooseberry
'Then he saw it, he saw bowls of gooseberries sitting.', (WB_KL-05:32)

(21) kári xás u-piip ip-nîinamich-p-i ip-nîinamich-p-i
then then 3SG>3-say ITER-little-DENOM-IMPER ITER-little-DENOM-IMPER
'And he said, "Get little, get little!"', (WB_KL-05:91)

(22) ta'ítam u-'arankúrih-een ayâach pá-siit tá kun-thárupriin
so 3SG>3-sink-ANT it.was.because the-mouse PER 3PL>3-gnaw.through
'But they sank, because the mice had gnawed holes in them.', (WB_KL-05:109)

(23) ta'ítam yée naa hínupa páy uum vúra pihnêefich pa-yûum
so well 1sg. surprise this 3SG.PRO INTENS coyote SUB-uphill.downriver
u-thívtaap-tih
3SG>3-do.war.dance-DUR
'So (they said), "Well, that's Coyote who is dancing downriver!', (WB_KL-06:40)

(24) xás yurúkthuuf p-oo-kvíripma xás nani-'ífuth thúf-kaam
then Bluff.Creek SUB-3SG>3-run.to then 1SG.POSS-behind creek-large
kam-'árihish
3SG>3-become
'And when he ran to Bluff Creek, then (he said) "Let it become a big creek behind me!"', (WB_KL-06:55)

(25) vaa páy pihnêefich u-kúpha-anik pa-káruk kahyúras u-vâaramu-tih
so this coyote 3SG>3-do-ANC SUB-upriver Klamath.Lakes 3SG>3-go-DUR
kupánakanakana
the.end
'Coyote did that, when he went upriver to Klamath Lakes. kupánakanakana.', (WB_KL-07:61)

(26) xás pihnêefich vúra uum sípnu-kaam t-óo thárish
then coyote INTENS 3SG.PRO storage.basket-large PER-3SG>3 put.down
pa-káan u-máhyaan-eesh
SUB-there 3SG>3-put.in-PROSP
'And Coyote put a big storage basket down where they were to put it in.', (WB_KL-08:7)

(27) kári xás u-píip chími man kúna vúra xáyfaat ík
then then 3SG>3-say soon INT in.addition INTENS don't! must
i-'ûurih
2SG>3-be.unwilling
'And they said, "All right, but you mustn't get tired."', (WB_KL-09:27)

(28) xás ta'ítam vaa vúra káru t-u-páaxkiv
then so that INTENS also PER-3SG>3-win.(game)
'And so (the upriver people) won that too.', (WB_KL-10:7)

(29) xás p-aaxíich u-patánviish-vunaa hôoy uumkun pa-'ávans-as
then the-child 3SG>3-ask.question-PL where they the-man-PL
'And he asked the children, "Where are the men?"', (WB_KL-10:24)

(30) xás vúra pa-t-u-'iink-áyaachha xás tée imnaká-kaam
then INTENS SUB-PER-3SG>3-be.on.fire-well then PER coal-large
'And when (the bark) had burned well, then there was a big coal.', (WB_KL-10:41)

(31) xás pa-t-óo kfuuyshur xás kári pa-yítha u-'êe pá-'aah
then SUB-PER-3SG>3 be.tired then then the-one 3SG>3-give the-fire
'And when he got tired, then he gave the fire to the (next) one.', (WB_KL-10:45)

(32) xás kári uum pa-t-óo kfuuyshur yítha kúna t-u-'éeh
then then 3SG.PRO SUB-PER-3SG>3 be.tired one in.addition PER-3SG>3-give
'And when he got tired, he gave it to another one.', (WB_KL-10:46)

(33) ta'ítam súva t-u-pakúriihva
so listen! PER-3SG>3-sing.songs
'So he heard them singing.', (WB_KL-11:16)

(34) vaa uum vúra pa-yúruk tá kun-víitrup
so 3SG.PRO INTENS SUB-downriver PER 3PL>3-paddle.downstream.from.here
t-u-thívruuhrup yúruk
PER-3SG>3-float.downstream.from.here downriver
'When they traveled downstream by boat, they floated downstream.', (WB_KL-15:4)

(35) ithyáruk kúna ú-p-viitroov-eesh
across in.addition 3SG>3-ITER-paddle.upstream.from.here-PROSP
u-thívruuhroov-eesh káru káruk u-vuunôov-ahi-ti
3SG>3-float.upriver-PROSP also upriver 3SG>3-flow.upstream.from.here-ESS-DUR
pa-'íshaha
the-water
'They would travel back upstream on the other side, they would float upstream also, the water was flowing upstream.', (WB_KL-15:5)

(36) kúna vúra pa-mu-krívraam vaa vúra u-músahi-ti
in.addition INTENS the-3SG.POSS-house that INTENS 3SG>3-look-DUR
pa-nunu-krívraam koovúra p-oo-tâayhi-ti iinâak vaa vúra
the-1plPOSS-house all SUB-3SG>3-be.many-DUR indoors that INTENS
u-músahi-ti ôok iinâak p-oo-tâayhi-tih
3SG>3-look-DUR here indoors SUB-3SG>3-be.many-DUR
'"But his house looks just like our house, everything that is inside looks just like what is inside here.', (WB_KL-16:13)

(37) vúra pu-sakeemvárihv-eesh-ara koovúra vaa u-músahi-ti
INTENS NEG-be.homesick-PROSP-NEG all that 3SG>3-look-DUR
pa-nunú-'uup
the-1PL.POSS-possession
'You won't be homesick, everything looks like our things.', (WB_KL-16:16)

(38) xás u-pí-ti vúra if koovúra vaa u-músahi-ti ôok
then 3SG>3-say-DUR INTENS true all that 3SG>3-look-DUR here
pa-nanú-'uup pa-nini-'ávan mu-krívraam
the-1PL.POSS-possession the-1SG.POSS-husband 3SG.POSS-house
'And she said, "It's true, everything looks like our things here, in my husband's house."', (WB_KL-16:29)

(39) xás pâanpay xás u-xús naa ni-xú-ti
then after.while then 3SG>3-think 1sg. 1s(>3)-think-DUR
na-pikshayvûunish-ti yukún koovúra vaa u-músahi-ti
2s/3s>1s-tell.lies.to-DUR you.see all that 3SG>3-look-DUR
pa-nunú-'uup káru uum vúra vaa u-músahi-ti
the-1PL.POSS-possession also 3SG.PRO INTENS that 3SG>3-look-DUR
pa-nini-'áka
the-1SG.POSS-father
'Then after a while she thought, "I think he's deceiving me, everything looks like our things, and he looks just like my father."', (WB_KL-16:36)

(40) kári xás áama ú-kyiim-nishuk
then then salmon 3SG>3-fall-out.of
'And salmon fell out.', (WB_KL-17:22)

(41) víri vaa kúth payêem pa-xuntápan kôokaninay vúra u-'íif-tih
so that because.of now the-acorn everywhere INTENS 3SG>3-rise-DUR
'That's why the acorns grow everywhere now.', (WB_KL-17:38)

(42) víri vaa kúth sâam u-saamnúpu-tih káru
so that because.of little.downhill 3SG>3-flow.downstream.from.here-DUR also
vaa kúth áama u-kvíripraa-tih
that because.of salmon 3SG>3-run.up.from.downhill-DUR
'That's why (the water) flows downstream, and that's why salmon run up the river.', (WB_KL-17:42)

(43) hínu páy ikreemyaha'úru p-óo-peen-ti iktûunih-i
surprise eggs.of.wind SUB-3SG>3-say.to-DUR carry.down-IMPER
'There it was eggs of the wind that she told him to take down.', (WB_KL-18:34)

(44) simsimvôo kích u-veehríshuk-va
sword only 3SG>3-stick.out-PL.ACT
'Nothing but swords were sticking out.', (WB_KL-18:59)

(45) kári xás u-xus tîi kan-imús-an
then then 3SG>3-think let 1SG>3-look.at-go.to
'And he thought, "Let me go see her!"', (WB_KL-19:10)

(46) xás p-oo-kréemya pa-'áptiik koovúra u-vrárasur
then SUB-3SG>3-blow the-branch all 3SG>3-fall.off.(pl.)
'And when it blew, the branches all fell off.', (WB_KL-20:32)

(47) xás u-kúniihka u-kyívunih
then 3SG>3-shoot.at 3SG>3-fall.downward
'Then they shot (a squirrel), (and) it fell down.', (WB_KL-20:60)

(48) xás u-piip hôoy uumkun pa-'ávans-as
then 3SG>3-say where they the-man-PL
'And they said, "Where are the men?"', (WB_KL-21:9)

(49) káruma uum pa-'ifápiit áxak pa-mu-'ífuni
in.fact 3SG.PRO the-young.unmarried.woman two the-3SG.POSS-hair
u-paathrámni pa-pátarav-ak
3SG>3-throw.into the-soup.basket-LOC
'The fact was, the young women had thrown two of their hairs into the soup-baskets.', (WB_KL-21:35)

(50) víri kún siit kích u-ksah-ár-ahi-tih
so meaning.unknown mouse only 3SG>3-laugh-go.to-ESS-DUR
'There only mice were squeaking.', (WB_KL-23:49)

(51) víri vaa yaas'ararée-thvaaykam u-vúrayvu-tih-eesh
so so rich.person-front 3SG>3-go.around-DUR-PROSP
'They will be around in front of rich people.', (WB_KL-24:47)

(52) tishravará'iivreer yanéekva pa-mu-'îin u-thivnúru-tih
Etna.Mountain visible; the-3SG.POSS-falls 3SG>3-roar-DUR
'On Etna Mountain he heard his falls thundering (at Katimin).', (WB_KL-26:5)

(53) yítha mú-'arama ú-krii káru mu-hrôoha
one 3SG.POSS-child 3SG>3-live also 3SG.POSS-wife
'His one child and his wife lived there.', (WB_KL-26:7)

(54) yukún yíiv á' u-tásunih-tih-anik pa-'íin
you.see far above 3SG>3-be.vertical-DUR-ANC the-falls
'You see, the falls were (like) a barrier (reaching) a long ways up.', (WB_KL-27:17)

(55) púyava p-óo-p-vaavruk á'iknêechhan tishravará'iivreen
you.see SUB-3SG>3-ITER-go.down.over falcon Etna.Mountain
u-xus hûut áta u-'íina-ti
3SG>3-think how maybe 3SG>3-experience.something.unpleasant-DUR
pa-nani-'íin
the-1SG.POSS-falls
'So when Duck Hawk looked down over Etna Mountain, he thought, "I wonder what's wrong with my falls?', (WB_KL-27:18)

(56) ôok íp ni-thítiim-tih-at p-óo-xaak-tih
here PAST 1s(>3)-hear-DUR-PAST SUB-3SG>3-make.noise-DUR
'Formerly I heard them sounding from here.', (WB_KL-27:19)

(57) púyava p-oo-'ípak yánava pa-nani-'îin t-óo
you.see SUB-3SG>3-come.back visible the-1SG.POSS-falls PER-3SG>3
p-vuunup
ITER-flow.downstream.from.here
'So when he got back, he saw it, "My falls have flowed downriver."', (WB_KL-27:21)

(58) kári xás ta'ítam xúus u-'uum-áheen
then then so thought 3SG>3-arrive-ANT
'Then they doctored her.', (WB_KL-29:3)

(59) kári ithívthaaneen t-óo thárish xás pa-xuntápan tá kun-íf-ar
then land PER-3SG>3 put.down then the-acorn PER 3PL>3-rise-go.to

'They were creating (lit., laying down) the world, and the acorns came to grow.', (WB_KL-30:1)

(60) xás xuntápan kun-ipêer hûut iim
then acorn 3PL>3-say.to how 2sg.
u-'íina-ti kúth
3SG>3-experience.something.unpleasant-DUR because.of
pa-pu-'ipthíth-aheen pa-mí-pxaan
SUB-NEG-finish.weaving-ANT the-2sPOSS-cap
'And they said to Tan Oak Acorn, "What's the matter with you that you didn't finish weaving your cap?"', (WB_KL-30:18)

(61) xás pa-'avansáxiich u-píktar pa-mú-taat
then the-boy 3SG>3-miss the-3SG.POSS-mother
'And the boys missed their mother.', (WB_KL-32:16)

(62) xás u-píip hôoy uum tátach
then 3SG>3-say where 3SG.PRO mama
'And they said, "Where's mama?"', (WB_KL-32:17)

(63) xasík p-aaxvahara-xárah-sas ku-vêehkurih-eesh yúux-ak
then.(future) the-ptich-wood-long-PL 2pl(>3)-stick.into-PROSP dirt-LOC
u-'ahí-tih-eesh
3SG>3-burn-DUR-PROSP
'You will stick the long pieces of pitch-wood in the sand, they will burn.', (WB_KL-32:38)

(64) yánava káan ára ú-kriihvu-tih
visible there person 3SG>3-fish.with.set-net-DUR
'They saw a man fishing there.', (WB_KL-32:75)

(65) ta'ítam u-pêethkee-heen
so 3SG>3-take.back.out-ANT
'So they took her away.', (WB_KL-33:72)

(66) púya-va pa-t-óo kxáramha pa-'asiktávaan t-u-'ípak
and.so-so SUB-PER-3SG>3 be.night the-woman PER-3SG>3-come.back
'And when it got dark, the woman returned home.', (WB_KL-39:8)

(67) kári xás chîimich sáruk ník u-'ákichnim-ach
then then little.bit downhill a.little 3SG>3-lie.coiled-DIM.(verbs)
pa-mú-'iikiv
the-3SG.POSS-necklace
'And his necklaces were just a little dab down at the bottom (of his neck)', (WB_KL-42:11)

(68) xás áchkuun kun-ipéer túus u-piip
then swamp.robin 3PL>3-say.to mockingbird 3SG>3-say
pa-mi-pakuhíram t-u-'ífikaraha
the-2sPOSS-acorn.gathering.ground PER-3SG>3-be.many.people.picking
pa-xuntápan
the-acorn
'And Swamp Robin was told, Mockingbird said, "They're picking the acorns at your acorn-picking grounds."', (WB_KL-43:7)

(69) xás kári áchkuun u-piip úma pa-mí-mvir
then then swamp.robin 3SG>3-say emphatic(?) the-2sPOSS-fishery
ikriróov t-óo páx pa-'áama
fishing.spot.name PER-3SG>3 catch.(fish) the-salmon
'And Swamp Robin said, "They've caught the salmon at your fishery, Ikrirôov."', (WB_KL-43:12)

(70) púyava payêem pa-t-óo snur pa-'ípaha t-óo kfuukiraa
you.see now SUB-PER-3SG>3 thunder the-tree PER-3SG>3 grab
t-óo sxáxaar
PER-3SG>3 tear.open
'So now when it thunders, (Lightning) grabs the trees, he rips them open.', (WB_KL-44:16)

(71) kári xás u-páthrih xás vúra u-páthrih
then then 3SG>3-rain then INTENS 3SG>3-rain

'But then it rained and it rained.', (WB_KL-45:19)

(72) kári xás chikus! t-u-vur
then then sound.of.poking PER-3SG>3-jab
'And wham! They jabbed him.', (WB_KL-46:21)

(73) p-eekxaréeyav vaa u-kúph-aanik
the-spirit.person so 3SG>3-do-ANC
'The gods did this.', (WB_KL-47:1)

(74) víri-va yítha u-sáam kéevniikich káru axiich
so-so one 3SG>3-remain old.woman.(dimin.) also child
'(Finally) one old woman and a child were left.', (WB_KL-47:9)

(75) yáanchiip táay u-'íf-eesh xuntápan
next.year much 3SG>3-rise-PROSP acorn
'The next year many acorns will grow.', (WB_KL-47:18)

(76) kári xás kun-piip vaa páy hínupa uum p-oo-kupí-tih-eesh
then then 3PL>3-say so this surprise 3SG.PRO SUB-3SG>3-do-DUR-PROSP
'And they said, "That's the way (Mankind) will do it."', (WB_KL-48:20)

(77) víri p-oo-ksah-ár-ahi-tih kun-tákaam-tih pa-'asiktávaan
so SUB-3SG>3-laugh-go.to-ESS-DUR 3PL>3-ridicule-DUR the-woman
pa-kâanimich p-oo-'ûupvu-tih
the-poor SUB-3SG>3-dig.roots-DUR
'So they laughed, they ridiculed her, the woman, the poor one, as she dug roots.', (WB_KL-50:11)

(78) kári xás vaa kun-kúupha kurihkirá-'aachipvari u-thrîish
then then so 3PL>3-do eaves-pole-towards.the.middle 3SG>3-set.(liquid).down
pa-mukun-yupastáran
the-3PL.POSS-tear.(from.weeping)
'They did this, their tears collected halfway up the roof-beam.', (WB_KL-51:48)

(79) xás p-oo-fúmtaapsur víri yûuth
then SUB-3SG>3-blow so downriver.and.across-stream
p-oo-trûupu-tih xánahich axmáy vaa u-kuupha
the-3SG>3-look.downriver.from.here-DUR after.while suddenly so 3SG>3-do

pa-mukun-furax-pikshipíkmath axmáy u-'áapuchur
the-3PL.POSS-woodpecker.head-sun-shade suddenly 3SG>3-collapse
'And when he blew, as he looked down across, in a little while suddenly they did this, (the women's) woodpecker-head sun-shades suddenly collapsed.', (WB_KL-52:74)

(80) víri vaa u-kuupha chaka'îich kúnish kun-'ixipúniih-va
so so 3SG>3-do slowly sort.of 3PL>3-drift.down-PL.ACT
'They did that, they sort of floated slowly down.', (WB_KL-52:75)

(81) kári xás pa-niinamichtâapas iknûumin veekxaréeyav
then then the-smallest Burrill.Peak.spirit the-3SG.POSS-child
pa-mú-'arama vaa vúra u-kupa-'íf-aha
so INTENS 3SG>3-MOD-rise-Modal the-3SG.POSS-hand
pa-mú-tiik u-'ákchaak-tih xákarari
3SG>3-close.hands-DUR on.both.sides the-3SG.POSS-hand
pa-mú-tiik u-'ákchaak-tih
3SG>3-close.hands-DUR
'And Burrill Peak Spirit's littlest child grew up this way, his hands were closed, both his hands were closed.', (WB_KL-54:4)

(82) víri kún pa-kéevniikich ú-krii káru
so meaning.unknown the-old.woman.(dimin.) 3SG>3-live also
patapriha-'ifápiit
placename-young.unmarried.woman
'There lived the old woman, and the young woman of patapríhak.', (WB_KL-57:8)

(83) vúrava u-'áharamu-naa-tih
just 3SG>3-follow-PL-DUR
'He was following them that way.', (WB_KL-57:27)

(84) káru pa-mu-krívraam u-pikchákiroopithva pa-sípnuuk axyará-va
also the-3SG.POSS-house 3SG>3-be.lined.up the-storage.basket full-Distributive
'And they were lined up around (the inside of) his house, the storage baskets were all full.', (WB_KL-57:117)

(85) pufíchtaahkoo ípmiif káru pa-kóo kumá-'uup
white.deer black.deer also SUB-all 3SG.POSS-possession
p-oo-tháthriinaa pa-sipnúuk-ak
SUB-3SG>3-sit.(two.things) the-storage.basket-LOC

'White deerskins, black deerskins, and every kind of treasure sat in the storage baskets.', (WB_KL-57:118)

(86) itaharatápas pa-mu-sípnuuk p-oo-pikchákiroopithva
whole.lot the-3SG.POSS-storage.basket SUB-3SG>3-be.lined.up
'There were a whole lot of storage baskets lined up around.', (WB_KL-57:119)

(87) káru i-máh-eesh áxak yuup ú-thyiimvarayv-eesh
also 2s(>3)-see-PROSP two eye 3SG>3-float.around.(two)-PROSP
'And you will see two eyes float around.', (WB_KL-58:14)

(88) kári xás hâari vúra pirísh-riik pa-tá kun-'áhoo pa-mukun-yáfus
then then sometime INTENS plant;-place SUB-PER 3PL>3-go the-3PL.POSS-dress
t-u-tatitítit
PER-3SG>3-be.tattered
'And sometimes it was a brushy place where they traveled, their dresses got torn.', (WB_KL-58:35)

(89) víri chavúra pu-'áraar iim-tih-ara chavúra p-eethívthaaneen u-p-áxyar
so finally NEG-human die-DUR-NEG finally the-land 3SG>3-ITER-fill
pa-'áraar
the-human
'Finally no person died, finally the people filled up the earth.', (WB_KL-58:56)

(90) káan xás mah'íitnihach u-papivan-kôo-ti
there then early.morning 3SG>3-go.search.for-to-DUR
pa-mu-sárum ishkêesh-ak hôoy kích t-óo
the-3SG.POSS-Jeffrey.pine.root river-LOC where only PER-3SG>3
p-thívruuhruprav
ITER-float.out.through
'Then she went early in the morning to look for her pine-roots there in the river, (she wondered) where they had floated out.', (WB_KL-59:7)

(91) víri îifuti u-thívruuh-tih
so sure.enough 3SG>3-float-DUR
'Sure enough, they were floating (there).', (WB_KL-59:8)

(92) itahara-hárinay t-u-taxváh-ahi-tih
ten-year PER-3SG>3-seal.shut-ESS-DUR

'They had been sealed up for ten years.', (WB_KL-61:46)

(93) xás vaa vúra u-p-íthvuuymath rúup
then so INTENS 3SG>3-ITER-name.(someone)
'So they named him Rube.', (WB_KL-66:21)

(94) káakum pa-'éekoons táay vúra tá kun-'ífik xás itahara-'átimnam
some the-acorns much INTENS PER 3PL>3-pick.up then ten-pack-basket
kóo t-óo píishha
as.much.as PER-3SG>3 soak.acorns
'Some people gathered a lot of acorns, and put as many as ten baskets to soak.', (WB_KL-68:11)

(95) pufích'aan t-óo kyav
string.for.snares PER-3SG>3 make
'They made twine for deer-traps.', (WB_KL-68:13)

(96) axak-súpaa t-u-p-músan pa-mu-tátapva
two-day PER-3SG>3-ITER-go.see the-3SG.POSS-trap
'Every two days they would go look at their traps.', (WB_KL-68:15)

(97) pa-xúrish t-óo krav páykuuk ása múuk
the-shelled.acorn PER-3SG>3 grind over.there rock with
'They ground the shelled acorns with that stone over there (pointing to a pestle).', (WB_KL-68:19)

(98) xás t-u-thántap pa-xúrish
then PER-3SG>3-winnow the-shelled.acorn
'And they sifted the shelled acorns.', (WB_KL-68:20)

(99) xás sáruk t-u-tákir astiip
then downhill PER-3SG>3-leach.(acorn.meal) shore
'And they leached them, downhill on the river bank.', (WB_KL-68:21)

(100) xás áh-kaam t-óo kyav
then fire-large PER-3SG>3 make
'And they made a big fire.', (WB_KL-68:22)

(101) pa-'ás t-óo párish
the-rock PER-3SG>3 heat.cooking.stones
'They heated stones (for cooking).', (WB_KL-68:23)

(102) xás asíp-kaam t-u-máhyaan pa-'éekoons
then bowl-large PER-3SG>3-put.in the-acorns
'And they put the acorns in a big soup-basket.', (WB_KL-68:24)

(103) t-u-'arámpuk
PER-3SG>3-cook.acorn.soup
'They cooked acorn soup.', (WB_KL-68:25)

(104) pa-'áraar uum pa-'áama u-kupée-kriihv-ahi-tih
the-human 3SG.PRO the-salmon 3SG>3-MOD-fish.with.set-net-Modal-DUR
'The Indians fished for salmon in a certain way.', (WB_KL-69:1)

(105) xás pa-mukun-'ikríhar uum taskana-tunvêech-as
then the-3PL.POSS-fish.trap 3SG.PRO pole-small.(pl.)-PL
u-kyâar-ahi-tih
3SG>3-make.with-ESS-DUR
'And their fish-trap was made of little poles.', (WB_KL-69:6)

(106) púyava pá-'aas u-kríkurih-va púyava pa-'áama tá
you.see SUB-water 3SG>3-set.net-ESS you.see SUB-salmon PER
kun-ívyiihraa xás urípih-ak tá kun-ihmáravar
3PL>3-come.here.(pl.) then net-LOC PER 3PL>3-run.in.through.(pl.)
'So when they set it into the water, when the salmon came up, then they ran into the net.', (WB_KL-69:14)

(107) púyava pa-'ipanîich pa-'áama t-u-'uum-áhaak púyava pa-'áan
you.see SUB-very.end the-salmon PER-3SG>3-arrive-when you.see the-string
t-óo kéen
PER-3SG>3 move
'When the salmon got to the end, the string quivered.', (WB_KL-69:15)

(108) koovúra u-yvúruk-ahi-ti pa-pufich-'ánav
all 3SG>3-rub-ESS-DUR the-deer-medicine
'They rubbed deer medicine on everything.', (WB_KL-70:13)

(109) víri-va máh'iit pa-tá kun-tátapv-an-va víri vaa u-kupí-tih
so-so morning SUB-PER 3PL>3-trap-go.to-PL.ACT so so 3SG>3-do-DUR
pa-mukun-chíshiih víri koovúra pa-mukun-chíshii ánav tá
the-3PL.POSS-dog so all the-3PL.POSS-dog medicine PER
kun-iyvúruk-va
3PL>3-rub-PL.ACT
'When they went trapping in the morning, their dogs would do this, they would all be rubbed with medicine.', (WB_KL-70:14)

(110) víri-va u-kupi-tih pa-t-u-paxfúr-oo pa-púufich
so-so 3SG>3-do-DUR SUB-PER-3SG>3-catch.in.trap-PL.ACT the-deer
'That's what they did, when they snared deer.', (WB_KL-70:19)

(111) púya-va pa-t-u-súpaah-aak púya-va ukráam kúuk tá
and.so-so SUB-PER-3SG>3-become.day-when and.so-so lake to.there PER
kun-ihmárava tá kun-páatvu-naa pa-'ávans-as
3PL>3-run.to.there.(pl.) PER 3PL>3-bathe-PL the-man-PL
'When day came, they went to a pond, the men bathed.', (WB_KL-72:5)

(112) xás pa-t-óo mtúp-ahaak pa-xuntápan kun-ivrarasúr-oo-tih
then SUB-PER-3SG>3 cooked-when the-acorn 3PL>3-fall.off.(pl.)-PL.ACT-DUR

'And when they were ripe, the acorns fell off.', (WB_KL-73:2)

(113) púyava pa-t-óo mfír-ahaak pá-yaaf tá
you.see SUB-PER-3SG>3 be.hot-when the-acorn.dough PER
kun-'ákith-ramni tharámpuukrav-ak
3PL>3-handle.(soft.mass)-into cooking.basket-LOC
'When they were hot, they put the acorn dough into a cooking basket.', (WB_KL-73:25)

(114) xás t-u-váxrah tá kun-iyvôonih xás t-óo
then PER-3SG>3-be.dry PER 3PL>3-take.(acorns).down then PER-3SG>3
yvax
shell.(acorns)
'And they dried; they took them down, and they hulled them.', (WB_KL-74:3)

(115) xás takiríram t-óo kyav
then acorn-leaching.hole PER-3SG>3 make
'And they made a leaching-hole.', (WB_KL-74:12)

(116) yáas ú-spaas-tih t-u-'akíthkith
then 3SG>3-be.leached-DUR PER-3SG>3-pick.up.(acorn.dough)
'Then (the meal) was leached, they stirred it up.', (WB_KL-74:13)

(117) iváxra t-óo kyav
dry PER-3SG>3 make
'They made it dry.', (WB_KL-74:14)

(118) t-óo thxah pá-yuux t-óo vyiih-shur
PER-3SG>3 wash.acorn.dough the-dirt PER-3SG>3 go.(pl.)-off
'They washed it, the sand came off.', (WB_KL-74:16)

(119) ásip-ak u-snap-ráamnih-va
bowl-LOC 3SG>3-put.on-into-ESS
'They put it in a cooking basket.', (WB_KL-74:17)

(120) tée p ás u-párish-at
PER PAST rock 3SG>3-heat.(cooking.stones)-PAST
'They had already heated rocks.', (WB_KL-74:19)

(121) pa-'ás u-tururáamnih-va
the-rock 3SG>3-put.(hot.stones).into-PL.ACT
'They put the rocks in.', (WB_KL-74:20)

(122) koovúra imváram u-'ifkóo-hi-ti pa-'ásip
all plate 3SG>3-fit.on-ESS-DUR the-bowl
'And the plate-baskets (for the salmon) fit into the soup baskets.', (WB_KL-74:22)

(123) púyava pa-t-u-'amayâa-haak xás kári tá kun-'av
you.see SUB-PER-3SG>3-good-tasting-when then then PER 3PL>3-eat
'And when they became good-tasting, then they ate them.', (WB_KL-75:8)

(124) xás pa-'iinâak íivhar u-thiiv-árayv-ahi-tih
then the-indoors board 3SG>3-lie.(of.one.thing)-here.and.there-ESS-DUR

'And on the inside boards were placed around.', (WB_KL-77:4)

(125) p-eev'ávahkam ivhara-tírih-shas u-yaakóo-hi-tih
the-roof board-wide-PL 3SG>3-put.(pl.).on-ESS-DUR
'And broad boards were put on the roof.', (WB_KL-77:9)

(126) xás ás u-thíiv-ahi-ti ivíthvaaykam
then rock 3SG>3-lie.(of.one.thing)-ESS-DUR front.of.house
'And rocks were laid in front of the house.', (WB_KL-77:12)

(127) xás pa-'iinâak ah-'ávahkam u-sasípiithva taskana-tunvêech-as
then the-indoors fire-over 3SG>3-spiral.around pole-small.(pl.)-PL
'And on the inside, above the fire, little poles were stretched around.', (WB_KL-77:13)

(128) xás u-kyâa-hi-ti pa-káan kun-iváxraahmath-ti pa-'áama
then 3SG>3-make-ESS-DUR SUB-there 3PL>3-dry.(something)-DUR the-salmon
káru vúra fâat vúra pa-kun-tâarahi-tih
also INTENS what INTENS SUB-3PL>3-have-DUR
'And they were made so that they dried fish there and whatever (else) they had.', (WB_KL-77:14)

(129) xás âapun vúra uum p-oo-tâayhi-ti
then on.the.ground INTENS 3SG.PRO SUB-3SG>3-be.many-DUR
pa-mukun-'ásip káru vúra fâat vúra pa-kun-tâarahi-ti
the-3PL.POSS-bowl also INTENS what INTENS SUB-3PL>3-have-DUR
pa-kun-imnísh-eesh
SUB-3PL>3-cook-PROSP
'And on the floor were their cooking baskets and whatever else they had when they were going to cook.', (WB_KL-77:15)

(130) xás pa-'ávansa vúra kích mukun-'ikrívkir u-tâayhi-ti
then the-man INTENS only 3PL.POSS-disk-seat 3SG>3-be.many-DUR
'And only the men's seats were there.', (WB_KL-77:16)

(131) xás kun-sáanvu-ti áhup ú-thvuuy-ti imtháatvar káru
then 3PL>3-carry.(things)-DUR wood 3SG>3-be.named-DUR shinny.stick also
tákasar
shinny.tossel

'And they carried sticks, they were called shinny sticks and a 'tossel' (i.e., a double ball).', (WB_KL-78:4)

(132) xás pa-tákasar uum ahup-tunvêech-as u-nhítunv-ahi-tih
then the-shinny.tossel 3SG.PRO wood-small.(pl.)-PL 3SG>3-tie.together-ESS-DUR
'And the tossel was little sticks, they were tied together.', (WB_KL-78:5)

(133) xákarari áxak pa-'ávans-as káru áachip áxak
on.both.sides two the-man-PL also middle two
'There were two men at each end and two in the middle.', (WB_KL-78:6)

(134) púya-va hâari uum pa-yu'-kúkam pa-'ávans-as píshiip
and.so-so sometime 3SG.PRO the-downriver-side the-man-PL first
t-u-'úum pa-tákasar u-phíriv-irak
PER-3SG>3-arrive SUB-shinny.tossel 3SG>3-lie.(two)-where
'Sometimes the men on the downriver end arrived first where the tossel lay.', (WB_KL-78:25)

(135) fátaak kúna t-óo sriv
somewhere in.addition PER-3SG>3 target-shooting
'They did target-shooting someplace.', (WB_KL-82:14)

(136) xás ikxúrar xás káh'ir t-u-várak
then evening then upriver.world-renewal PER-3SG>3-come.down
'And in the evening they did the war dance.', (WB_KL-82:21)

(137) yáan vúra u-súpaahi-tih
recently INTENS 3SG>3-become.day-DUR
'It was just becoming day.', (WB_KL-83:27)

(138) u-peechkanvichv-á-ra-hi-tih
3SG>3-gamble-DEVERB-having-DENOM-DUR
'There was gambling.', (WB_KL-83:42)

(139) káru káh'ir t-u-várak
also upriver.world-renewal PER-3SG>3-come.down
'And they did the war dance.', (WB_KL-83:44)

(140) káruk u-'ír-ahiv
upriver 3SG>3-celebrate.world-renewal-time
'There was a world renewal upriver.', (WB_KL-84:1)

(141) xás vaa yíiv yúruk xumvaroo-máruk áh-kaam t-óo kyav
then so far downriver placename-uphill fire-large PER-3SG>3 make
'And he made a big fire far downriver, uphill from xumvároov.', (WB_KL-84:11)

(142) xás taakrípaak kúuk t-u-'uum
then placename to PER-3SG>3-arrive
'And they went to taakrípaak.', (WB_KL-84:20)

(143) vúra fátaak xás yáv u-'íihya
INTENS somewhere then good 3SG>3-stand.(long.object)
'Some places (the trees) are good', (WB_KL-85:28)

(144) xás pa-mukun-yafusa-yêepsha vúra uum yâamach
then the-3PL.POSS-dress-good.(pl.) INTENS 3SG.PRO pretty
u-kyâah-ahi-tih
3SG>3-make-ESS-DUR
'And their good dresses were made pretty.', (WB_KL-86:5)

(145) píshiip panyúrar u-taxapkóo-hi-tih
first beargrass 3SG>3-braid.on-ESS-DUR
'First bear-lily leaves were braided on.', (WB_KL-86:6)

(146) pa-mukun-'ápxaan u-kyâar-ahi-ti sárip káru sárum
the-3PL.POSS-cap 3SG>3-make.with-ESS-DUR hazel.twigs also Jeffrey.pine.root
'Their hats were made with hazel twigs and pine-roots.', (WB_KL-86:10)

(147) xás pa-mukun-'ápxaan uum u-vík-ahi-ti pa-sárum
then the-3PL.POSS-cap 3SG.PRO 3SG>3-weave-ESS-DUR the-Jeffrey.pine.root
mûuk káru pa-panyúrar káru p-eekritápkir káru pa-tíiptiip
with also the-beargrass also the-maidenhair.fern also the-Woodwardia.fern
'And their hats were woven with the pine-roots and the bear-lily leaves and the five-finger fern and the chain fern.', (WB_KL-86:13)

(148) pa-'arara-'asiktávaan uum ishváak u-thúkinh-ahi-tih
the-human-woman 3SG.PRO chin 3SG>3-tattoo-ESS-DUR

'The Indian women were tattooed on the chin.', (WB_KL-87:1)

(149) kuyráak u-sasip-úniih-va yítha achipyâach xás xákararih
three 3SG>3-be.in.line.(pl.)-down-PL.ACT one very.center then on.both.sides

'There were three stripes running down, one right in the middle and (two) on each side.', (WB_KL-87:2)

(150) hâari tírih-shas káru hâari vúra tûupichas kuynákmahich
sometime wide-PL also sometime INTENS small.ones three.at.a.time
p-oo-sasip-úniih-va
SUB-3SG>3-be.in.line.(pl.)-down-PL.ACT
'Sometimes they were wide and sometimes they were narrow, and sometimes they were each (composed of) three little ones running down.', (WB_KL-87:3)

(151) puxích kúnish u-páthriih-tih víri pu-xú-tih-ap kíri nu-pêer
very.much sort.of 3SG>3-rain-DUR so NEG-think-DUR-NEG I.wish 1>2SG-say.to
pa-axíitich-as ôok kóova nu-'am
the-child-PLhere so 1>2SG-eat
'It's sort of raining hard, so they don't want to ask the (neighbor's) children to eat here with them.', (WB_KL-89:6)

(152) túuyship u-vêehrim-va xás u-tíshraam-hi-tih
mountain 3SG>3-stand.(things)-PL.ACT then 3SG>3-valley-DENOM-DUR
mu-súrukam
3SG.POSS-under
'Mountains are standing, and a valley is below them.', (WB_KL-92:52)

(153) páy uum pichas-'ípaha u-vêehrim-va xás simsímtas
this 3SG.PRO peaches-tree 3SG>3-stand-PL.ACT then wire.fence
u-taaspáth-ahi-tih
3SG>3-fence.around-ESS-DUR
'These peach trees are standing, and there is a wire fence around.', (WB_KL-92:73)