

Pragmatic Influences on Argument Word Order in Karuk Narrative Texts¹

Department of Linguistics, University of California, Berkeley

Kevin Yu

May 2021

1 Introduction

Karuk is an indigenous Native American language spoken by the Karuk people, whose ancestral territory stretches along the Klamath River from *Panámnik* (modern day Orleans) to *Athithúfvuunupma* (modern day Happy Camp) (Garret et al., 2020). According to Golla (2011, p. 86), Karuk is a severely endangered language, as there are currently fewer than a dozen fluent first-language speakers left. However, there is a growing number of people who have achieved some degree of second-language fluency. This is in part thanks to language revitalization efforts, which include school-based programs for children and the AICLS Master-Apprentice program. Although Karuk is genetically a language isolate, it is classified under the Hokan phylum (Golla, 2011, p. 82–84).

Davis et al. (2020, p. 844) describe Karuk as a headmarking, polysynthetic language, and Mikkelsen (2014) argues that its basic word order is verb final. Additionally, Davis et al. (2020) also describe Karuk as a non-configurational language because it exhibits the three characteristics of non-configurational languages from Hale (1983): arguments may be freely ordered, omitted, and split. The first characteristic is of particular interest for this study, which seeks to tackle the following question: what factors influence the order of nominal arguments with respect to their verbs? In other words, is Karuk word order completely random, or are there certain factors which increase the likelihood of particular word orders arising?

Currently, there is very little research on the factors which influence Karuk word order. My study seeks to fill this gap by determining whether referential distance, topic persistence, predicate transitivity, animacy, or thematic continuity can be correlated with preverbal or postverbal argument position (see section 2 for detailed explanations of these factors). My study finds that referential distance, which measures the number of clauses between subsequent occurrences of referents, is a significant predictor of subject position. Specifically, subjects that have lower referential distance values (i.e. subject referents which are relatively

¹This paper is an adaptation of my undergraduate honors thesis, which I completed over the course of the 2020-2021 academic year. I would like to thank the Karuk Tribe, even though I did not get to work with them directly, for sharing such a rich collection of cultural and linguistic knowledge in *Ararahih'urípih* ('Karuk Language Net,' URL: <http://linguistics.berkeley.edu/~karuk/index.php>). Furthermore, I could not have written this paper without the wisdom, guidance, feedback, and encouragement of my advisor, Professor Line Mikkelsen, and my readers, Professor Eve Sweetser and Professor Isaac Bleaman.

Throughout this paper, I will use the following abbreviations: ANC: Anciliary, ANT: Anterior, DUR: Durative, FUT: Future tense, ITER: Iterative, KL: *Karok Language*, NEG: Negative, NOMZ: Nominalizer, PAST: Past tense, PERF: Perfect, PL: Plural, POSS: Possessive, SG: Singular, and WB: William Bright. Following Bright (1957, p. 58–64), glosses of verbal agreement prefixes indicate both the number and person of the subject and the direct object. For example, 3SG(>3) indicates a 3rd person singular subject acting upon a 3rd person object.

close to their previous occurrence) are more likely to appear postverbally, while referents with a higher referential distance are more likely to occur preverbally. On the other hand, I found that animate objects were more likely than inanimate ones to occur postverbally. Although referential distance did not have a significant effect on object position, it is worth noting that referential distance was very close to being a statistically significant predictor of object position and followed the same trend as subject referential distance (see Sections 3 and 4 for a more detailed discussion of object referential distance).

The non-configurationality of Karuk can be seen in the following examples, which also appear in Davis et al. (2020):

(1) a. **SOV word order:**

púyava kári pa='áraar pa='urípi u-p-ithyúru-ri-paa
 you.see then the-human the-net 3SG(>3)-ITER-pull.out
 'Then the Indian pulled the net out of the water'
 Julia Starritt "Salmon Fishing" (WB_KL-69: 16)

b. **SVO word order:**

xás pa-pihñiich u-píimni pa-mú-'aramah
 then the=old.man 3SG(>3)-fall.in.love the=3SG.POSS-child
 'And the old man fell in love with his child.'
 Julia Starritt "Coyote Marries His Own Daughter" (WB_KL-16: 3)

c. **VOS word order:**

ta'ítam kun-ífik-aheen pa-xuntápan pa-'asiktávaan-sa
 so 3PL(>3SG)-pick.up-ANT the=acorn the=woman-PL
 'Then the women gathered the acorns'
 Mamie Offield "Coyote Gives Salmon and Acorns to Mankind" (WB_KL-17: 34)

(2) **Freely dropped arguments:**

xás t-u-'áv
 then PERF-3SG(>3)-eat
 'Then he ate it'
 Julia Bennett "Screech Owl and Coyote" (ALK_14-35: 16)

(3) **Split arguments**

púyava táay tá kun-'ûupva pa-tayíith
 you.see much PERF 3PL(>3SG)-dig.roots the=Brodiaes
 'So they dug a lot of brodiaes'
 Nettie Ruben "The Story of Skunk" (WB_KL-46: 14)

Examples (1a) to (1c) demonstrate three potential word orders in Karuk. Additionally, VOS, OSV, and OVS are also attested in Karuk (Maier, to appear). Example (2) demonstrates that both the subject and direct object of a transitive verb can be dropped. Finally, (3) demonstrates that an NP consisting of a quantifier and a noun can be separated.

Maier (to appear) provides a Treebank of the Karuk language along with data on the distribution of different word orders. Table 1 displays the distribution of SV, VS, OV, and

VO word order. These include sentences that have one overt argument as well as those with two overt arguments.

Word Order	Number of Occurrences
SV	912
VS	291
OV	414
VO	274

Table 1: Distribution of SV, VS, OV, and VO

While Table 1 does not distinguish between clauses with one or two overt arguments, Table 2 displays the distribution of word orders for clauses with two overt arguments.

Word Order	Number of Occurrences
SOV	54
SVO	63
VSO	1
VOS	4
OVS	9
OSV	8

Table 2: Distribution of Word Orders in Clauses with Two Overt Arguments

In total, 912 clauses have SV word order while 291 have VS word order. As for objects, 414 clauses have OV word order while 274 have VO word order.

My study joins many other studies on languages with pragmatically controlled word order, including languages such as Klamath (a Penutian language of southern Oregon) (Meyer, 1992), Chamorro (a Malayo-Polynesian language of Guam) (Cooreman, 1992), Ute (a Uto-Aztecan language of Colorado and Utah), Paraguayan-Guarani (a Tupí-Guaraní language of Paraguay) (Tonhauser and Colijn, 2010), and Nez Perce (a Sahaptian language spoken in Idaho) (Rude, 1992). A common finding among many of these studies is that postverbal word order is correlated with more continuous referents, while preverbal word order is correlated with discontinuous referents (Meyer, 1992; Rude, 1992). Indeed, (Givón, 1983b, p. 33-34) predicts that preverbal position is correlated with discontinuous referents while postverbal position is correlated with continuous referents. This entails that referents that are more predictable and those which were mentioned more recently tend to occur postverbally, while other referents tend to occur preverbally. Thus, one may expect lower referential distance and higher topic persistence to correlate with postverbal word order, since these characteristics indicate salient referents. Cooreman (1992) also finds that breaks in thematic continuity (i.e. clauses which introduce new lines of action) correlate with preverbal word order.

My findings from Karuk help to confirm the prediction from Givón (1983b) that postverbal position encodes continuous referents while preverbal position encodes discontinuous referents. The tendency for low referential distance subjects to occur postverbally more often than high referential distance ones indicates that continuous subjects appear postverbally more often than discontinuous ones. However, the lack of a statistically significant

effect of referential distance on object position seems to go against Givón’s prediction. Topic persistence, on the other hand, was found not to have a significant effect for either subjects or objects. Similarly, Rude (1992, p. 205) also finds that topic persistence did not have any discernible correlation with word order in Nez Perce. As for thematic continuity, I found that this predictor did not have a significant effect on the position of subjects or objects.

In addition to referential distance, topic persistence, and thematic continuity, my study also considered predicate transitivity and animacy as potential predictors of word order. Tonhauser and Colijn (2010) examined the effect of these factors (among others) on word order in Paraguayan-Guaraní. This study found that inanimate noun phrases were more likely to appear postverbally than animate/human referents, and it also reported that only animate/human direct objects were realized preverbally. In contrast, my findings concerning animacy were rather different from the findings of Tonhauser and Colijn (2010). I found that animacy only had a significant effect on object position, so not all referents were affected by this factor. Furthermore, while Tonhauser and Colijn (2010) found that inanimate noun phrases tended to appear postverbally, I found that *animate* objects tended to appear postverbally. This observation may be explained in terms of iconicity and Differential Object Marking. Animate objects are marked (Aissen, 2003, p. 8), and postverbal position in Karuk is also marked, so what we see is that objects that are marked in terms of animacy have a tendency to appear in a marked position.

Regarding transitivity, Tonhauser and Colijn (2010) found that intransitive subjects had a slight preference for preverbal word order, while transitive subjects had a slight preference for postverbal word order. However, my study did not find that transitivity was a significant predictor of either subject position or object position.

Section 2 describes my methodology and includes a discussion of the coding decisions that I made when annotating referents for referential distance, topic persistence, thematic continuity, animacy, and predicate transitivity. It also describes these factors in more detail and describes the quantitative analysis that I performed using logistic regression. Section 3 lays out the results of this quantitative analysis and shows that referential distance was the only factor which had a significant effect on the placement of subjects. Next, Section 4 discusses the results of the quantitative analysis, and compares my findings to those of studies which examine other languages with flexible word order. Finally, Section 5 concludes by summarizing my findings and proposing directions for future research.

2 Methodology

The goal of this study is to test the relationship between word order and the following factors: referential distance (henceforth abbreviated RD), topic persistence (henceforth abbreviated TP), thematic continuity, animacy, predicate transitivity, and definiteness. Both RD and TP are defined in Givón (1983b). These measurements have been applied in numerous studies, including Cooreman (1992) and Meyer (1992). The principle of thematic continuity also comes from Givón (1983b), and its application in this study follows the methodology of Cooreman (1992).

I annotated six Karuk narratives for each of the five factors listed above. In total, these texts contained a total of 447 clauses, including those which did not overtly real-

ize any nominal arguments. These narratives, along with their speakers, are listed here: “Coyote’s Journey” by Chester Pepper (WB_KL-03), “Coyote and the Sun” by Chester Pepper (WB_LA78.1-016b), “Coyote Steals Fire” by Julia Starritt (WB_KL-10), “Coyote Goes to a War Dance” by Julia Starritt (WB_KL-06), “Coyote Goes to the Sky” by Julia Starritt (WB_KL-08), and “The Perils of Weasel” by Lottie Beck (WB_KL-18). I accessed these texts through an online corpus called *Ararahih’urípih* (‘Karuk language net,’ <http://linguistics.berkeley.edu/~karuk/index.php>). In *Ararahih’urípih*, these texts are labelled using an alphanumeric code that indicates their specific origin. For example, “WB” stands for “William Bright” and “KL” stands for “Karok Language,” indicating that these texts come from *The Karok Language* by William Bright (Bright, 1957). This corpus segments each text into individual sentences, and each sentence comes with a gloss as well as a translation. For example, here is a sentence from “Coyote’s Journey”:

- (4) víri-va káan ta’ítam pihnêefich káan u-thívruuhma
 so-so there so coyote there 3SG(>3)-float.to
 “So then Coyote floated to there”
 Chester Pepper “Coyote’s Journey” (WB_KL-03:101)

The sentences, along with their glosses and translations, constituted the dataset for my study. The following sections describe the procedures behind my annotations, as well as cases that were difficult to resolve. I begin by discussing my method for separating subjects and objects, as well as the decisions that I made when classifying referents into either category. I then turn to cases that required special attention during my annotation process. Next, I discuss each of the annotation measurements in the following order: RD, TP, thematic continuity, transitivity, and animacy. Finally, I describe the logistic regression model that I constructed, as well as the reasons for choosing this particular analysis.

2.1 Annotation process and coding decisions

All overt arguments were categorized as subjects, direct objects, or indirect objects. This was done to examine whether subjects and objects had different word order properties. This was the case in Paraguayan-Guaraní. Tonhauser and Colijn (2010) found that while direct objects and indirect objects almost always occurred postverbally, subjects exhibited a greater degree of flexibility. Within the texts that I analyzed, only a handful of overt indirect objects were found and annotated. All of these indirect objects were excluded due to their scarcity. Subject identification was rather straightforward. Direct objects, on the other hand, were more difficult to classify. This was in part due to the fact that I needed to decide whether to include arguments that were added by applicative suffixes. For each overt subject and each overt direct object, I applied all five of the measurements being examined in this study (RD, TP, thematic continuity, animacy, and predicate transitivity).

Following Maier (2020), if a verb only takes two nominal arguments, then one must be the subject while the other is the direct object. Even if the verb could take a clausal complement in addition to the two nominal arguments, the two nominal arguments were still annotated as subject and direct object. For example, despite the fact that *ipêer* ‘to say to’ can take up to three arguments, one of these arguments is a clausal complement, so the two nominal

arguments are labeled as ‘subject’ and ‘direct object.’ Example (5) illustrates this principle (note that *ipeer* is phonologically realized as *ipeen* in this context):

- (5) xás kun-ípeen-ti pihnêefich payêem
 then 3PL(>3SG)-say.to-DUR coyote now
 pu-kin-ípkookanp-eesh-ara
 NEG-1>2SG-go.with.again-FUT-NEG
 ‘Then they told Coyote, “Now you won’t go with us again.”’
 Chester Pepper “Coyote’s Journey” (WB_KL-03:163)

In this example, *pihnêefich* ‘Coyote’ would be coded as a direct object, rather than an indirect object.

Applicative suffixes were also important to consider because they could introduce additional, often directional, arguments to verbs. The question that arises, then, is whether these applied arguments should count as direct objects or as some other type of argument. Applicative suffixes in Karuk are highly complex, and Macaulay (2004) discusses them in greater detail. For now, I will list a couple of examples that I encountered while annotating my section of the corpus. In each example, the applicative suffix and its corresponding applied argument are bolded for emphasis.

- (6) kári xás p-oo-vôonupuk **ikmaháchraam** u-vôonupuk
 then then NOMZ-3SG(>3)-leave.house sweathouse 3SG(>3)-leave.house
 ‘And when he went out, (the person sweating) came out of the sweathouse.’
 Chester Pepper “Coyote’s Journey” (WB_KL-03:23)
- (7) ta’ítam máh’iit yáan vúra u-súpaahi-tih u-vôor**uraa**
 so morning recently Intensive 3SG(>3)-become.day-DUR 3SG(>3)-crawl.up
pa=’ípaha
 the=tree
 ‘So in the morning it was scarcely dawn, he climbed up the tree.’
 Lottie Beck “The Perils of Weasel” (WB_KL-18:12)

In (6), the suffix *-rupuk* (realized as *-nupuk*) adds the argument *ikmaháchraam* ‘sweathouse.’ In (7), the suffix *-uraa* adds the argument *pa’ípaha* ‘the tree.’

Following Maier (2020), I separated applied objects from direct objects and indirect objects. As a result, applied objects were excluded from the categories of subjects, direct objects, and indirect objects. Because my study is only concerned with subjects and direct objects, the word order properties of applied arguments lie beyond the scope of my study, so they are excluded from my analysis. Returning to examples (6) and (7), *ikmaháchraam* ‘sweathouse’ and *pa’ípaha* ‘the tree’ are not treated as direct objects and are consequently excluded from my analysis.

Subsections 2.1.1 to 2.1.3 discuss notable types of cases that required special attention.

2.1.1 Ambiguous cases

In rare instances, it was unclear which verb a particular referent was an argument of. If an argument occurs preverbally with respect to one verb and postverbally with respect to

another, then this creates a dilemma for my annotation. In order to determine which factors influence argument word order, I must be able to determine whether a particular nominal argument is postverbal or preverbal with respect to its verb. Example (8) demonstrates a case of ambiguity that I encountered.

- (8) xás ta'ítam ú-kriihv-aheen **pa=sípnuuk**
 then so 3SG(>3)-fish.with.set.net-ANT **the-storage.basket**
 u-pakurîihvu-tih
 3SG(>3)-sing.songs-DUR
 ‘Then the storage basket fished, it was singing.’
 Julia Starritt “Coyote Goes to a War Dance” (WB_KL-06:16)

Here, the ambiguous referent in question is *pasípnuuk* ‘the storage basket.’ Both *úkriihvaheen* ‘to fish with a set net’ and *upakurîihvutih* ‘singing songs’ have the same verbal agreement, which indicates a third person singular subject for both verbs. Thus, *pasípnuuk* ‘the storage basket’ is a compatible subject for both of these verbs. To resolve this ambiguity, I employed the principle of ‘following the translation’ from Maier (2020, p. 9). This means that in order to resolve these cases of ambiguity, I referred to the translation in order to choose the interpretation that most closely matched the meaning expressed by the translation.

Returning to (8), the translation contains two clauses, one which explicitly mentions ‘the storage basket’ and one which only uses the pronoun ‘it.’ I assume that the explicit reference to ‘the storage basket’ corresponds to an overt realization of *pasípnuuk*, while the pronoun corresponds to an implicit reference. Thus, I consider *pasípnuuk* to be an argument of *úkriihvaheen* ‘to fish with a set net,’ rather than *upakurîihvutih* ‘singing songs.’ The verb *upakurîihvutih* has an implied argument, corresponding with the use of ‘it’ in the translation. Based on these decisions, *pasípnuuk* was coded as a postverbal argument of the verb *úkriihvaheen*.

2.1.2 Constituent Questions

According to Davis et al. (2020, p. 848), wh-words within constituent questions in Karuk have a tendency to be fronted to the left periphery. Thus, in order to prevent constituent questions from skewing the results of my study, I excluded them from my annotations. The sentence in (9) provides an example of an excluded wh-question.

- (9) xás fâat chími vúra kun-páxraam-eesh
 then what soon Intensive 3PL(>3SG)-bet.again-FUT
 ‘Then what were they to bet?’ Julia Starritt “Coyote Steals Fire” (WB_KL-10: 5)

In (9), the wh-question *fâat* ‘what’ appears preverbally near the beginning of the question. Based on the analysis from Davis et al. (2020, p. 848), *fâat* ‘what’ appears preverbally because in general, wh-words are moved to the left periphery in Karuk. Thus, I excluded this sentence from my analysis. On the other hand, while constituent questions were excluded from this study, polar questions were not. This is because I did not find enough evidence in my data to suggest that the word order properties of polar questions differ substantially from those of declarative sentences in Karuk.

2.1.3 An Appositional Construction That Straddles the Predicate

According to Maier (2020, p. 41), apposition refers to a phenomenon in which two noun phrases refer to the same entity. In rare cases, one referent occurs before a verb, while the other referent occurs after the same verb. These cases defy word order characterizations, since I cannot determine whether the referent in question is preverbal or postverbal. I only came across one such case, so I decided to exclude it from my analysis, as doing so would not have a large impact on my study. Example (10) is the sole example of this phenomenon that I encountered:

- (10) *xás vúra uum yâamach mu-'asiktaván-'aramah*
then Intensive 3SG pretty 3SG.POSS-woman-child
'And her female child was pretty.'
Lottie Beck "The Perils of Weasel" (WB_KL-18:2).

In example (10), the pronoun *uum* 'he/she/it' and the noun *mu'askitaván'aramah* 'her female child' both refer to the female child. However, *uum* is preverbal while *mu'askitaván'aramah* is postverbal. Thus, it is unclear as to whether the referent 'the female child' is preverbal or postverbal in this sentence. As a result, I excluded (10) from my analysis.

2.2 Measuring RD (Referential Distance)

Givón (1983b, p. 13) defines RD as the number of clauses between the current occurrence of a referent and its most recent occurrence. RD is a look-back measure because it counts clauses starting with the current occurrence and then progressing backwards through the previous clauses. By measuring how far the previous occurrence is, RD acts as a way to measure how easy it is for a speaker/listener to recall a particular referent. Referents that have a higher RD have been inactive in the discourse for a longer period of time; therefore, they are harder to recall. On the other hand, lower RD signals a referent that has been mentioned recently and is easier to recall.

While RD was only applied to overt arguments of verbs, the previous occurrence of a referent does not need to be overt (Givón, 1983b, p. 14). Examples (12a) to (12c) illustrate this rule. Additionally, in some cases, the same entity may be referred to using two different NPs. For these cases, I consider both NPs to be the same referent for the purposes of calculating RD. Consider the following example, which contains two, non-adjacent sentences from the same narrative:

- (11) a. *xás u-árihish pa-kéevniikich*
then 3SG(>3)-sing the-old.woman.(dimin.)
'And the old woman sang.'
Lottie Beck "The Perils of Weasel" (WB_KL-18: 62)
- b. *xáyva âanxus pa-keechxâach mâaka u-paathrípaa*
by.luck weasel the-widow.(old) little.uphill 3SG(>3)-throw.inland
'By luck Weasel threw the widow into (the corner) uphill.'
Lottie Beck "The Perils of Weasel" (WB_KL-18: 66)

Both *pakéevniikich* ‘the old woman’ and *pakeechxâach* ‘the old widow’ refer to the same character, so for the purpose of calculating RD, they are occurrences of the same referent.

The minimum value that can be assigned for RD is 1, representing a maximally continuous referent which occurs in the immediately preceding clause. Givón (1983b) arbitrarily sets the maximum value of RD at 20, and many other studies, such as Cooreman (1992), Meyer (1992), and Rude (1992), follow this practice. Following these studies, I have also set the maximum value of RD to 20, and any referent that has an RD of 20 or higher is assigned this value. Additionally, this maximum value is assigned to first time occurrences of referents and to most indefinite referents (as many indefinites were also first time occurrences). Givón (1983b, p. 11) claims that first time occurrences of referents and indefinite referents, like referents with high RD values, are more difficult to process because they require the speaker/hearer to open a new file for them, i.e. they require the speaker/hearer to store a new referent in their memory. Because of this, these referents are assigned an RD value of 20.

To illustrate how RD is applied, the following example takes three consecutive clauses from “Coyote’s Journey” by Chester Pepper. In this example, RD is applied to the occurrence of *pihnêefich* ‘Coyote’ in sentence (12c).

- (12) a. ta’ítam u-p-thívruuhvarak
 so 3SG(>3)-ITER-float.down.from.upriver
 ‘So he floated back down from upstream’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 91)
- b. víri víra uum táay pa-’áhup-tunvêech-as
 so Intensive 3.SG much the-wood-small.(PL.)-PL
 u-’áthanvarak-tih
 3s(>3)-float.down.from.upstream-DUR
 ‘There were a lot of little sticks floating down from upriver.’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 92)
- c. xás pihnêefich u-xús chími ahup-yâamach kan-párihish
 then coyote 3SG(>3)-think soon wood-pretty 1SG(>3)-be.transformed
 ‘And Coyote thought, “Let me become a pretty stick!”’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 93)

Applying referential distance to the occurrence of *pihnêefich* ‘Coyote’ in (12c), we obtain a value of 2. The clause in (12b) does not contain any reference to Coyote, either overt or implied, so it counts as a gap. However, (12a) does contain an implied reference to Coyote. This can be seen in the translation (as indicated by the word ‘he’), as well as the 3rd person agreement on the verb *upthívruuhvarak* ‘to float down from upstream.’ The previous occurrence of a referent does not have to be overt, so we find that the previous reference to *pihnêefich* is two clauses behind the current reference.

The methodology that I employed when handling quotations and subordinate clauses is based on the methodologies of Meyer (1992) and Rude (1992). For the purpose of calculating RD, I calculated RD based on the number of sentences rather than the number of clauses. This means that relative clauses, adverbial clauses, and complement clauses were not counted as separate clauses. Instead, they were grouped together with their main clause and counted

as one clause. RD was still applied to referents occurring within these subordinate clauses. For example, (13) presents an example of an adverbial clause.

- (13) *xás pa=pihnêefich u-’úum yánava axíich kích*
 then NOMZ=coyote 3SG(>3)-arrive visible child only
kun-’áaraahi-tih
 3PL(>3SG)-live.(PL)-DUR
 ‘And when Coyote got there, he saw there were nothing but children’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10: 23)

The entirety of (13) is considered to be one clause, even though it contains an adverbial clause: *papihnêefich u’úum* (‘when Coyote got there’). The beginning of this adverbial clause is marked with *pa-*, which is a clitic that functions as a complementizer. RD is still calculated for the reference to *pihnêefich* (‘Coyote’) in the adverbial clause. Additionally, RD is also applied to *axíich* (‘child’). However, when calculating RD for subsequent referents, the sentence in (13) is counted as just one clause.

For quotations, referents inside of quotations were not measured for RD. This means that in example (12c), RD is not calculated for the referent *ahupyâamach* ‘pretty stick.’ This is because RD is intended to reflect the difficulty that the speaker/hearer faces when identifying a referent in discourse (Givón, 1983b, p. 10–11). Quotations represent the thoughts of characters within the narration and reflect their own knowledge internal to the world of the narrative. Thus, it is not entirely clear whether the RD of referents within quotations reflect the speaker’s/hearer’s ability to identify a referent, or if they reflect a character’s ability to identify a particular referent. For example, it is possible for a referent to be repeatedly mentioned in a narrative and thus easily identifiable by both the speaker and hearer. However, for a character in the narrative, they could be learning about this referent for the first time, and so their ability to identify the referent does not necessarily reflect the speaker’s/hearer’s ability to identify the referent. In order to avoid potential complications arising from referents within quotations, I excluded these cases from my analysis. As for how I counted quotations while calculating RD, I considered both the quotative margin and the direct quotation to be one single clause (so (12c) is a single clause).

Furthermore, there may be cases where the referent in question occurs outside of a quotation, while its previous occurrence occurs inside of a quotation. In these cases, the previous occurrence inside the quotation still counts as an occurrence. However, clauses inside of quotations which do not contain the referent under consideration do not count as gaps. These principles are illustrated in the following example:

- (14) a. *payêem pa=ni-máh-aak pa=saamvároo pa=niní-vaas*
 now NOMZ=1s(>3)-see-when the=creek the=1SG.POSS-blanket
kúuk ni-paathm-éesh
 to.there 1SG(>3)-throw.toward-FUT
 ‘‘Now when I see a creek, I’ll throw my blanket at it’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 73)
- b. *xasík vaa ni-pachipchíp-eesh*
 then.(future) that 1SG(>3)-suck.on-FUT

‘Then I can suck on that.’ ’

Chester Pepper “Coyote’s Journey” (WB_KL-03: 74)

- c. xás p=oo-máh pa=saamvároo ta’ítam kúuk
then NOMZ=3SG(>3)-see the=creek so to.there
u-paathm-áheen
3SG(>3)-throw.toward-ANT

‘And when he saw the creek, then he threw (the blanket) at it.’

Chester Pepper “Coyote’s Journey” (WB_KL-03: 75)

Here, we are concerned with the referent *pasaamvároo* ‘the creek’ in (14c). Note that the quotation beginning in (14a) does not contain a quotative margin. Occasionally, quotations will occur without a verb of saying, and because Karuk is a pro-drop language, the subject of the omitted verb of saying may also be dropped. In these cases, the subject of the verb of saying is still considered to be present. The referential distance of *pasaamvároo* ‘the creek’ is 1. The clause in (14b) does not count as a gap, since it is located inside of a quotation. However, (14a) contains a reference to *pasaamvároo*. Thus, the referential distance between the reference to *pasaamvároo* in (14a) and the reference in (14c) is just 1.

2.3 Measuring TP (Topic Persistence)

Givón (1983b, p. 14–15) defines TP as the number of clauses in which a referent continues to have an uninterrupted presence as an argument of a verb. Thus, while RD is a backward looking measurement, TP is a forward looking measurement, as it counts the number of clauses after the current occurrence of a referent under consideration. Givón (1983b, p. 14–15) notes that while RD reflects the difficulty of the speaker’s/hearer’s task of identifying referents, TP reflects the topic’s importance in the discourse. Givón assumes that more important referents occur more frequently and therefore have a higher probability of persisting longer.

In some studies, including a study on Klamath (Meyer, 1992) and a study on Ute (Givón, 1983a), subjects and objects were treated differently when calculating TP. In these studies, when TP is measured for subjects, only subsequent occurrences as subjects are counted. On the other hand, direct objects can be counted regardless of whether they reoccur as subjects or objects. In my study, I also follow this practice. According to Givón (1983a, p. 157), subjects are the main topic of clauses. Thus, by counting only subsequent occurrences as the subject of a clause, TP measures the number of clauses in which a subject continues to be the *main* topic. If a referent switches from subject position to object position, then it loses its status as the main topic of a clause, and this drop in importance constitutes a break in TP.

As was the case when calculating RD, relative clauses, adverbial clauses, and complement clauses were not considered as separate clauses independent of their main clauses. Additionally, quotations and their quotative margins were treated as single clauses, and subsequent occurrences in quotations were also counted. However, if a referent does not occur within a quotation, then this does not count as a gap.

The sentences in examples (15a) to (15e) demonstrate the application of TP. All of these clauses are consecutive clauses from the narrative “Coyote Steals Fire” by Julia Starritt.

- (15) a. xás pihnêefich u-’áasish
 then coyote 3SG(>3)-lie.down
 ‘Then Coyote lay down’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10: 38)
- b. naa tá ní-kviit-ha
 1SG PERF 1SG(>3)-sleep
 ‘(He said), “I’m going to sleep.”’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10:38)
- c. káruma íp uum t-óo yuunk-at ahtúun
 in.fact PAST 3SG PERF-3SG(>3) poke.with.long.object-PAST oak.bark
 pa=mu-fithih-’ípan
 the=3SG.POSS-foot-end
 ‘He had put oak bark in his toes.’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10:39)
- d. xás káan ú-yruuhriv
 then there 3SG(>3)-(SG).lie
 ‘And he lay there’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10:40)
- e. áak u-piivkírih-tih
 in.the.fire 3SG(>3)-stick.feet.in.fire-DUR
 ‘He stuck his feet in the fire’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10:40)
- f. xás vúra pa-t-u-’iink-áyaachha xás téé
 then Intensive NOMZ-PERF-3SG(>3)-be.on.fire-well then PERF
 imnaká-kaam
 coal-large
 ‘And when (the bark) had burned well, then there was a big coal.’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10:41)

Here, we are concerned with calculating the TP value of the referent *pihnêefich* ‘Coyote’ in (15a). Note that (15b) does not contain an overt reference to Coyote. Instead, it contains a quotation that is spoken by Coyote and nothing else. TP does not require every subsequent reference to be overt, so because Coyote is still understood to be the one doing the speaking in (15b), this clause is still counted when calculating TP. Likewise, note that (15d) and (15e) both contain implicit references to Coyote, as indicated by the verbal agreement. These clauses are also counted. However, (15f) does not include a reference to Coyote, so we stop counting TP at this point. Thus, the TP value of the referent *pihnêefich* in (15a) is 4, as clauses (15b) to (15e) contain occurrences of this referent, all of which are in subject position. If there were overt occurrences of *pihnêefich* in the clauses after (15a), then these occurrences would have TP values under 4 because they come after the occurrence in (15a) and thus have a smaller number of occurrences after them.

The example in (15) demonstrated a case in which TP was broken by a clause that did not contain the referent under consideration. This next example demonstrates the added

constraint on subject TP by providing an example of TP being broken by downgrading. In this example, we are calculating TP for the first occurrence of *pamu'afupchúrax* ‘the anus’ in (16a).

- (16) a. hinupáy pa-mu-'afupchúrax p-oo-'iinkú-tih
 surprise the-3SG.POSS-anus NOMZ-3SG(>3)-be.on.fire-DUR
 ‘There it was his anus burning’
 Chester Pepper “Coyote’s Journey” (WB_KL-03:51)
- b. hinupáy íp pa-'axváha mûuk u-p-sívshaap-at
 surprise PAST the-pitch with.(by.means.of) 3SG(>3)-ITER-seal.up-PAST
 hinupáy vaa p-oo-'iinkú-tih
 surprise that NOMZ-3SG(>3)-be.on.fire-DUR
 ‘There it was the pitch he had sealed it with that was burning’
 Chester Pepper “Coyote’s Journey” (WB_KL-03:52)

In sentence (16a), *pamu'afupchúrax* ‘his anus’ occurs in subject position. Next, in (16b), the object of the verb *upsívshaapat* ‘seal up’ is *pamu'afupchúrax* ‘the anus.’ Thus, without any constraints on subject TP, this would count as an occurrence of *pamu'afupchúrax* ‘the anus’ and TP would be incremented. However, I am only counting subsequent occurrences of subject referents if those occurrences are in subject position. In (16b), *pamu'afupchúrax* ‘the anus’ is the object of *upsívshaapat* ‘seal up,’ so this is a case in which a subject is demoted to object position in the next clause. Therefore, (16b) constitutes a break in TP, and the TP value of *pamu'afupchúrax* ‘the anus’ is 0.

2.4 Thematic Continuity

Cooreman (1992, p. 244) defines a thematic paragraph as “a narrative unit in which one or more participants are involved in one or a series of activities which form a unified whole and which move towards a general goal.” For example, in “Coyote’s Journey” (WB_KL-03), the second paragraph describes a series of events in which Coyote leaves an unspecified location, travels upriver, and announces to a group of people that he is going to Klamath Lakes. Broadly speaking, this series of events takes place near the starting point of Coyote’s journey, and its goal is to initiate his journey. Thus, these activities are all a part of one theme: Coyote’s departure.

Because thematic paragraphs are characterized by a coherent sequence of events, large shifts in sequences of events can signal the beginning of new thematic paragraphs. Thematic continuity, then, is an assessment of whether a referent serves to continue a particular line of action in a thematic paragraph (in which case, it is considered continuous), or if it initiates a new line of action and hence a new thematic paragraph. My methodology for assessing thematic continuity follows the methodology of Cooreman (1992), which utilized thematic continuity to determine its influence on word order in Chamorro narrative texts.

For my study, I used the paragraph boundaries in the online corpus as the boundaries for each thematic paragraph. Without access to audio recordings of the narratives, I had no clear way of using phonetic or prosodic clues to determine paragraph boundaries. I decided to choose the boundaries in *Ararahih'urípih* (which originate from William Bright’s transcriptions) because they coincided with the beginnings of new lines of action and were readily

available. For each referent, I decided whether they were located at either a “break” in thematic continuity, or a “continuation” of thematic continuity. Breaks in thematic continuity represent referents that occur at the beginning of thematic paragraphs. On the other hand, continuations of thematic continuity represent referents that occur after the first clause of a paragraph (i.e. in paragraph-medial position or paragraph final position).

In many cases, paragraphs within the online corpus only contained a single clause. This was usually the case in dialogues, where attention shifted from one speaker to another. For cases in which a paragraph contains only a single clause, referents within the clause are classified as breaks in thematic continuity. To illustrate this principle, (17) provides an example of a dialogue between a group of people and Coyote.

- (17) a. *xás kun-ípeen-ti pihnêefich payêem*
 then 3PL(>3)-say.to-DUR coyote now
pu-kin-ípkookanp-eesh-ara
 NEG-1>2SG-go.with.again-FUT-NEG
 Then they told Coyote, “Now you won’t go with us again.”
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 163)
- b. *xás pihnêefich u-píip payêem ník vúra ishávaas*
 then coyote 3SG(>3)-say now a.little Intensive child.of.deceased.sibling
vaa pu-nee-p-kuph-êesh-ara
 so NEG-1SG(>3)-ITER-do-FUT
 ‘And Coyote said, “I won’t do it again this time, nephew.”’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 164)

Sentence (17a) mentions a group of people speaking to Coyote. Then, sentence (17b) shifts attention to Coyote by focusing on what he says in response. Thus, the topic between these two sentences is different, so a break in thematic continuity exists between (17a) and (17b). As a result, the occurrence of *pihnêefich* ‘Coyote’ in (17b) is classified as being located at a break in thematic continuity.

2.5 Assessing predicate transitivity

Verbs in Karuk can be classified as intransitive, transitive, or ditransitive based on how many nominal arguments they take. In example (18) below, the verb *xúriha* (‘to be hungry’) provides an example of an intransitive verb with only one nominal argument:

- (18) *xás vúra t-óo xúriha pihnêefich*
 then Intensive PERF-3SG(>3) be.hungry coyote
 ‘And Coyote was hungry.’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 130)

In sentence (18), *xúriha* has only one argument: *pihnêefich* ‘Coyote.’ In addition to intransitive verbal predicates, nominals in Karuk may also act as non-verbal predicates. These non-verbal predicates were classified as intransitives, and none of the factors that were applied to nominal arguments were applied to these non-verbal predicates. This is because my study is concerned with the order of nominal arguments with respect to their verbs, and

because nominal predicates are not arguments, they are not subjected to the same factors as other nominal referents.

Non-verbal predicates were occasionally found in copular constructions. In these constructions, the 3rd person singular pronoun *uum* can function as a copula (Garret et al., 2020, p. 14). In these cases, *uum* is not considered to be the predicate. Rather than describing the subject as a predicate would, *uum* acts as a copula because it mediates between a subject and its non-verbal predicate (Maier, 2020, p. 43).

In other cases, *uum* is omitted as a copula, resulting in non-verbal predicates occurring with just its subject. In the following example, *athkúrit* ‘fat’ provides an example of such a non-verbal predicate:

- (19) pa=mukun-patúmkir káru vúra athkúrit
 the=3PL.POSS-head.rest also Intensive fat
 ‘Their pillows were fat too.’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 128)

In example (19), *athkúrit* ‘fat’ is not classified as an indefinite referent, and so the five factors that I applied to argument NPs are not applied to non-verbal predicates like *athkúrit*. However, because this non-verbal predicate still takes a single argument, it is classified as an intransitive predicate, and all factors are applied to its subject.

Examples (20a) and (20b) demonstrate the use of a transitive verb and a ditransitive verb, respectively:

- (20) a. xás pa=mu-’asíp-haar u-thaxávxav
 then the=3sPOSS-bowl-and.all 3SG(>3)-chew.up
 ‘And he chewed up (the person’s) baskets to boot.’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 20)
- b. xás pa=t-óo kfuyyshur xás kári pa=yítha u-’êe
 then NOMZ=PERF-3SG(>3) be.tired then then the=one 3s(>3)-give
 pá=’aah
 the=fire
 ‘And when he got tired, then he gave the fire to the (next) one.’
 Julia Starritt “Coyote Steals Fire” (WB_KL-10: 45)

In example (20a), the transitive verb is *uthaxávxav* ‘he chewed up.’ The direct object is *pamu’asíphaar* ‘the person’s basket’ and the subject is Coyote. Here, only the direct object is overt. In example (20b), the subject is Coyote (not overt), the direct object is *pá’aaah* ‘the fire,’ and the indirect object is *payítha* ‘the (next) one.’

In addition to intransitive, transitive, and ditransitive verbs, I also created a category for verbs that took one nominal argument and one clausal complement, as well as a category for verbs that took two nominal arguments and one clausal complement. The first category was referred to as the “transitive with clausal complement” category, since it takes two arguments, one of which is a clausal complement. I labelled the latter category as the “ditransitive with clausal complement” category. The reason why I created these separate categories is because clausal arguments and nominal arguments have different word order properties. According to Davis et al. (2020, p. 847), complement clauses must follow the verb. Nominal arguments, on

the other hand, are not restricted to either preverbal or postverbal position due to the non-configurationality of Karuk (Davis et al., 2020, p. 845). Because clausal complements and nominal arguments exhibit different word order properties, I decided to create two separate categories for verbs that took clausal complements.

Example (21) provides an example of a verb with one nominal argument and one clausal complement.

- (21) kári xás u-xús-aanik chími káruk ishpuúk kan-ikyâan
 then then 3SG(>3)-think-ANC soon upriver dentalium.shells 1SG(>3)-go.gather
 kahyúras
 Klamath.Lakes
 ‘And he thought, “Let me go upriver to get money at Klamath Lakes!”’
 Chester Pepper “Coyote’s Journey” (WB_KL-03: 4)

In (21) the only nominal argument is ‘he,’ and it is not overtly expressed (in this case ‘he’ refers to Coyote). The thought itself is the clausal complement of the verb *uxúsaanik* ‘thought.’

On the other hand, (22) provides an example of a verb with two nominal arguments and one clausal complement.

- (22) xás kun-ípeen-ti pihnêefich payêem
 then 3PL(>3SG)-say.to-DUR coyote now
 pu-kin-ípkookanp-eesh-ara
 NEG-1>2SG-go.with.again-FUT-NEG
 ‘Then they told Coyote, “Now you won’t go with us again.”’
 Chester Pepper “Coyote’s Journey” (WB_KL-03:163)

One nominal argument is ‘they,’ which in this sentence is not overtly expressed. The second nominal argument is *pihnêefich* ‘Coyote.’ Finally, the content of the speech is the clausal complement of the verb *kunípeenti* ‘they told.’

2.6 Assessing animacy

To assess animacy, I used a modified animacy hierarchy based off of the one used in Tonhauser and Colijn (2010). While Tonhauser and Colijn (2010) distinguished human, animate, and inanimate referents, my study distinguishes only animate and inanimate referents. I chose this two-way distinction in order to simplify the statistical model, which I describe in more detail in Section 2.7. The category of animate referents includes animals, humans, and animals with human-like behavior (e.g. Coyote). The category of inanimate referents include all other referents.

Examples (23a) and (23b) provide examples of an animate and inanimate referent, respectively. In each example, the relevant example is bolded for emphasis.

- (23) a. Animate:
 xás axmáy **chishíi** kun-íkvuuhvu-naa
 then suddenly **dog** 3PL(>3SG)-howl-PL

‘And suddenly dogs howled’
 Julia Starritt “Coyote Steals Fire” (WB-KL_10: 61)

b. Inanimate:

xás **pa=mu-sípnuuk** uum ipshûunkinich
 then **the=3SG.POSS-storage.basket** 3SG low

And his storage basket was short

Julia Starritt “Coyote Goes to the Sky” (WB-KL_08: 11)

In (23a), *chishíi* ‘dog’ is an animal capable of moving and howling on its own, so it is classified as animate. On the other hand, *pamusipnuuk* ‘his storage basket’ is inanimate because it cannot move on its own.

2.7 Logistic Regression Analysis

I used multivariate logistic regression to determine whether or not the five variables described above had a significant impact on word order. I chose this particular analysis because word order, the dependent variable of this study, was encoded as a binary categorical variable (i.e. any given argument was either preverbal or postverbal). In addition, there were five independent variables, some of which were categorical while others were numerical. Additionally, logistic regression was a useful analysis for my data because my data was not normally distributed. For example, RD had a bimodal distribution rather than a normal distribution, as most referents either had a value of 1 or a value of 20.

In order to control for grammatical function, I split the data into two separate groups: one for subject referents and another for object referents. This was done in case subjects and direct objects had different word order properties. I then implemented two logistic regression models, one for each dataset. The specification of each model is listed in 24.

- (24) a. **Subject word order** is modeled as a response to RD + TP + thematic continuity + animacy + predicate transitivity
 b. **Object word order** is modeled as a response to RD + TP + thematic continuity + animacy + predicate transitivity

3 Quantitative Results

When collecting a sample for a statistical analysis, it is important to consider the size of this sample relative to the size of the population that it came from. Thus, I will begin this section by discussing the size of each of my datasets and the size of the full corpus in *Ararahih’urípih*. The six narratives that I examined contained a total of 447 clauses, including ones which did not realize either of the verb’s arguments. Of these 447 clauses, 200 of them contained either an overt subject, an overt direct object, or both. Of these clauses, 146 of them contained an overt subject. On the other hand, 64 of these clauses contained an overt direct object. Clauses that contained both an overt direct object and an overt subject were relatively rare, and only 10 clauses fell into this category. As I noted in Section 2.1, overt referents were classified as either subjects or direct objects in case subjects and objects had different word order properties. Indirect objects were excluded due to their scarcity.

According to Maier (to appear), the corpus that I consulted contained a total of 5310 clauses, each with a unique predicate. Of these clauses, 1481 of them contained an overt subject, while 887 of them contained an overt direct object. Of the clauses containing an overt subject, 912 of them have SV word order while 291 of them have VS word order. As for the clauses containing an overt object, 414 of them have OV word order while 274 of them have VO word order. The distribution in (2) (repeated here for convenience) displays the number of occurrences of all word orders for clauses that had both an overt subject and an overt direct object.

Word Order	Number of Occurrences
SOV	54
SVO	63
VSO	1
VOS	4
OVS	9
OSV	8

Table 3: Distribution of Word Orders in Clauses with Two Overt Arguments

All five of the predictors that I described in Section 2 were applied to both overt subjects and overt direct objects. Sections 3.1 to 3.5 discuss the quantitative results of applying these predictors to subjects, and Section 3.6 tests whether any of them were a significant predictor of subject position. Similarly, Sections 3.7 to 3.11 discuss the quantitative results of each predictor after being applied to overt direct objects, and Section 3.12 tests correlations.

3.1 Subject Referential Distance

Table 4 displays both the counts and percentages of preverbal and postverbal subjects for each value of referential distance (RD) that I observed. These values range from 1, indicating maximally continuous subjects, to 20, indicating maximally discontinuous subjects.

Value	Preverbal Subjects	Percent Preverbal	Postverbal Subjects	Percent Postverbal
1	35	79.54%	9	20.45%
2	12	80%	3	20%
3	8	72.73%	3	27.27%
4	7	77.78%	2	22.22%
5	3	75%	1	25%
6	3	100%	0	0%
7	3	60%	2	40%
8	1	100%	0	0%
9	0	0%	1	100%
10	1	100%	0	0%
12	1	100%	0	0%
20	50	98.04%	1	1.96%

Table 4: Distribution of Subject Referential Distance

In Table 4, we see that subjects tend to occur preverbally, regardless of their RD. We can also observe that the number and the percentage of postverbal subjects appears to steadily decrease as RD increases. However, RD is not equally distributed at each value, resulting in many intermediate values having very small counts. To get a better comparison of subjects with low RD values and subjects with high RD values, we can divide all of the subjects into those with an RD value less than or equal to 10, and those with an RD value greater than 10. The category of subjects with an RD value greater than 10 also includes subjects being mentioned for the first time. These first time mentions, along with subjects that have been absent from the discourse for a relatively long amount of time, represent discontinuous subjects that are harder to identify.

Figure 1 represents this partition visually by displaying the proportions of postverbal and preverbal subjects for subjects with RD values greater than 10 and subjects with RD values less than or equal to 10.

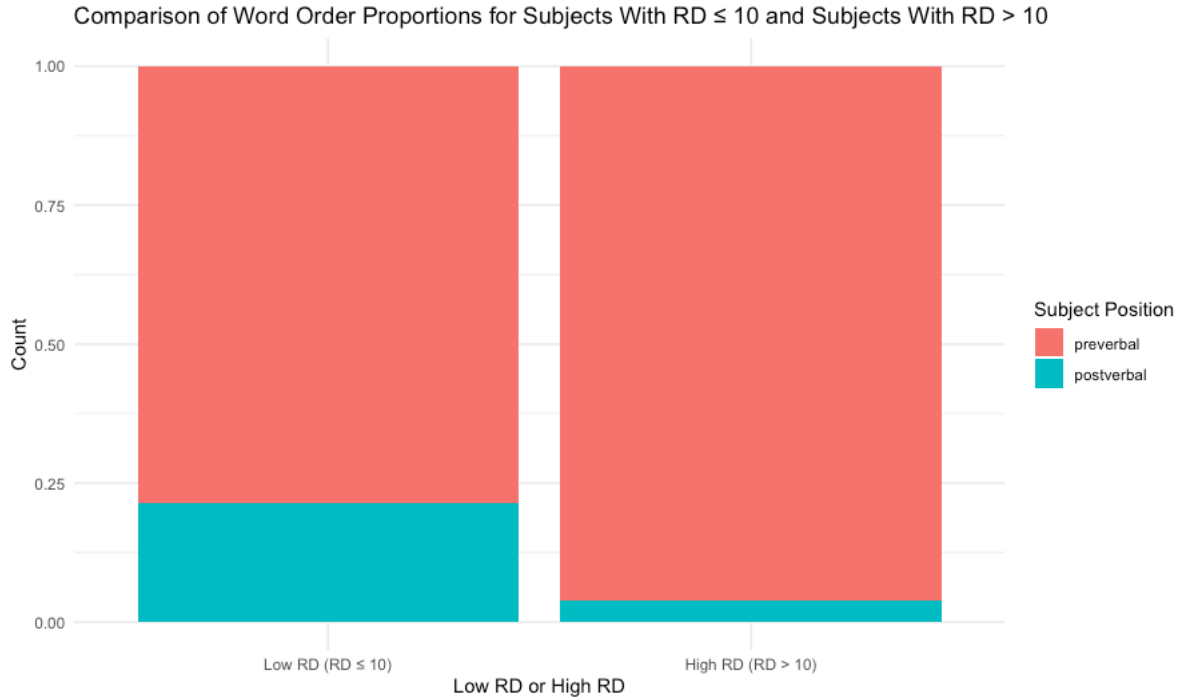


Figure 1: Word Order Proportions of Low RD and High RD Subjects

In this graph, we do indeed see that the proportion of postverbal subjects with high RD values (i.e. $RD > 10$) is much lower than the proportion of those with low RD values. This observation suggests that for subjects, lower values of RD are correlated with postverbal position.

3.2 Subject Topic Persistence

Table 5 shows the counts and percentages of preverbal and postverbal subjects for each value of TP that I observed for subjects.

Value	Preverbal Subjects	Percent Preverbal	Postverbal Subjects	Percent Postverbal
0	80	86.96%	12	13.04%
1	20	83.33%	4	16.67%
2	13	76.47%	4	23.53%
3	4	100%	0	0%
4	2	100%	0	0%
5	1	100%	0	0%
6	2	66.67%	1	33.33%
7	0	0%	1	100%
10	1	100%	0	0%
12	1	100%	0	0%

Table 5: Number and Percentage of Preverbal Subjects and Postverbal Subjects for Each Value of Topic Persistence

Looking at the percentages of postverbal subjects in column 5, we can see that the percentages generally increase as TP increases. However, it is also worth noting that subjects with TP values greater than or equal to three were relatively rare. Because of the scarcity of subjects with higher TP values, it is difficult to infer any trends in subject position based on the data in Table 5.

3.3 Thematic Continuity and Subject Word Order

Recall that thematic continuity was measured based on a referent’s position within a paragraph. Each paragraph represented its own line of action and revolved around a central theme or idea, so referents that occurred at the beginning of paragraphs represented breaks in thematic continuity. On the other hand, referents in paragraph-medial or paragraph-final position represented continuations of thematic continuity.

Table 6 shows the quantities and percentages of preverbal and postverbal subjects that occurred at breaks in thematic continuity and continuations of thematic continuity.

Thematic Continuity	Preverbal Subjects	Percent Preverbal	Postverbal Subjects	Percent Postverbal
Break	41	85.42%	7	14.58%
Continuous	83	84.69%	15	15.31%

Table 6: Quantities and Percentages of Preverbal Subjects and Postverbal Subjects in Thematically Non-Continuous and Thematically Continuous Positions

Based on the third column (“Percent Preverbal”) of Table 6, we can see that the difference between the percentage of preverbal subjects in the “break” category (85.42%) and the percentage of preverbal subjects in the “continuous” category (84.69%) is less than 2%. Likewise, column five shows a very small difference (<1%) between the percentage of postverbal subjects at thematic breaks (14.58%) and the percentage of postverbal subjects that were

thematically continuous (15.31%). These results suggest that thematic continuity does not significantly affect the position of subjects.

3.4 Effect of Predicate Transitivity on Subject Position

Originally, predicates were split into the following categories: intransitive, transitive, transitive with clausal complement (reflecting verbs with one nominal argument and one clausal complement), ditransitive, and ditransitive with clausal complement (two nominal arguments and one clausal complement). The last two categories ultimately had a very small number of subject referents. As a result, I combined the “ditransitive” and “transitive” categories, as well as the “transitive with clausal complement” and “ditransitive with clausal complement” categories.

For each category of transitivity, Table 7 provides the counts and percentages of preverbal subjects and postverbal subjects.

Transitivity	Preverbal Subjects	Percent Preverbal	Postverbal Subjects	Percent Postverbal
Intransitive	75	84.27%	14	15.73%
Transitive	25	86.21%	4	13.79%
Transitive with clausal complement	24	85.71%	4	14.29%

Table 7: Word Order Distribution Among Subjects for Different Levels of Predicate Transitivity

In Table 7, we see that for each level of predicate transitivity, the percentages of preverbal subjects are all rather similar. Looking at column 3 (“Percent Preverbal”) the difference between the intransitive category and the transitive category is about 1.94%. Likewise, the percentages of postverbal subjects in each category are also rather similar. In column 5 (“Percent Postverbal”), the difference is less than 2%. Because these differences are so small, it is reasonable to suspect that predicate transitivity is not a significant predictor of subject position.

3.5 Subject Animacy and Subject Position

Table 8 displays the counts and percentages of preverbal and postverbal subjects for each level of animacy. The rows are arranged in order of increasing level of animacy from top to bottom.

Transitivity	Preverbal Subjects	Percent Preverbal	Postverbal Subjects	Percent Postverbal
Inanimate	25	83.33%	5	16.67%
Animate	99	85.34%	17	14.66%

Table 8: Counts and Percentages of Preverbal and Postverbal Subjects for Each Animacy Level

In Table 8, we find that the percentages in Column 3 (“Percent Preverbal”) and Column 5 (“Percent Postverbal”) are very close together. There is only a 2.01% difference between inanimate and animate subjects in terms of the percentage of subjects that occurred preverbally/postverbally in each animacy category. The lack of a large difference between animacy categories suggests that animacy is not a significant predictor of subject position.

3.6 Logistic Regression Model for Subjects

To summarize the observations from Sections 3.1 to 3.5, lower RD values seemed to be correlated with postverbal realizations of subjects. On the other hand, higher values of TP were very scarce, so no clear inferences about the connection between TP and subject position can be made. In addition, thematic continuity, predicate transitivity, and animacy all seemed to have no effect on subject position. To test whether there was any correlation between these predictors and subject position, I used a logistic regression model. Logistic regression assumes that the response variable is binomially distributed (as opposed to linear regression, which assumes that the response variable is generated by a Gaussian, or normal, process). This is the case in my data, since subject position is either preverbal or postverbal.

To implement this model, I modelled the log odds of the probability of a particular subject appearing postverbally based on the values of the five predictors discussed in Section 2. Table 9 displays the output of running this model using R’s `glm()` function.

Term	Coefficient Estimate	Standard Error	Z Statistic	P Value	N
Intercept	0.09282	0.91550	0.101	0.91925	146
Subject RD (Baseline: RD = 0)	-0.11703	0.04211	-2.779	0.00545	146
Subject Topic Persistence (Baseline: TP = 0)	0.04259	0.12228	0.348	0.72760	146
Subject Animacy = Inanimate (Baseline)	N/A	N/A	N/A	N/A	30
Subject Animacy = Animate	-1.08925	0.75727	-1.438	0.15032	116
Transitivity = Intransitive (Baseline)	N/A	N/A	N/A	N/A	89
Transitivity = Transitive	-0.27383	0.67380	-0.406	0.68445	29
Transitivity = Transitive with clausal complement	-0.19477	0.70300	-0.277	0.78174	28
Thematic Break = Break (Baseline)	N/A	N/A	N/A	N/A	48
Thematic continuity = continuous	-0.18113	0.56451	-0.321	0.74831	98

Table 9: Results of Logistic Regression Model for Subjects

Based on this model, we see that RD had a significant impact on subject word order. Its coefficient is negative, indicating that subject RD is negatively correlated with the probability of a subject appearing postverbally. This means that subjects with high RD values are significantly less likely to appear postverbally. No other predictor had a significant impact on subject word order. To visualize the effect of subject RD on subject position, Figure 2 illustrates the model’s predictions of obtaining a postverbal subject for each value of RD.

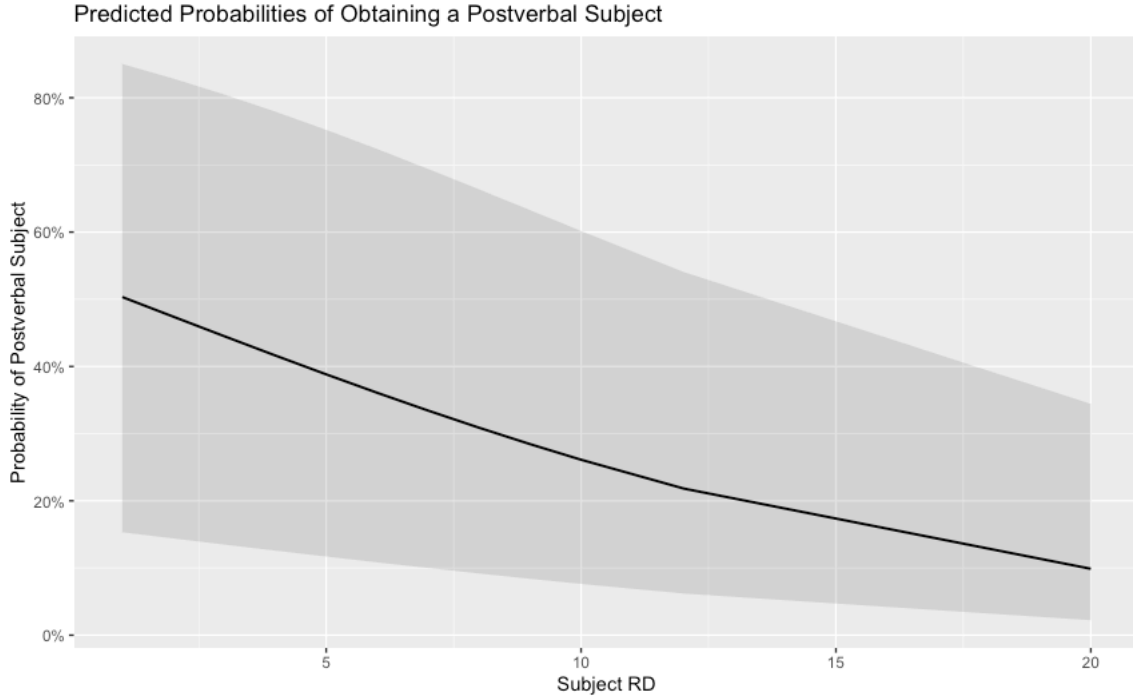


Figure 2: Model Predictions of the Probabilities of Obtaining a Postverbal Subject Based on RD

According to Figure 2, as subject RD increases, the probability of observing a postverbal subject does indeed decrease.

Because RD was the only significant predictor of subject position, we can test whether removing the non-significant factors results in a different fit. Table 10 shows the results of removing the non-significant factors and running this new model in R.

Term	Coefficient Estimate	Standard Error	Z Statistic	P Value
Intercept	-1.12095	0.29870	-3.753	0.000175
Subject RD	-0.08983	0.03689	-2.435	0.014889

Table 10: Pruned Logistic Regression Model With RD

This model still confirms that RD has a significant impact on subject word order. The coefficient is also negative, which confirms that as the RD of a subject increases, the probability of that subject appearing postverbally decreases. To illustrate this trend, Figure 3 plots the new model’s predicted probabilities of observing a postverbal subject.

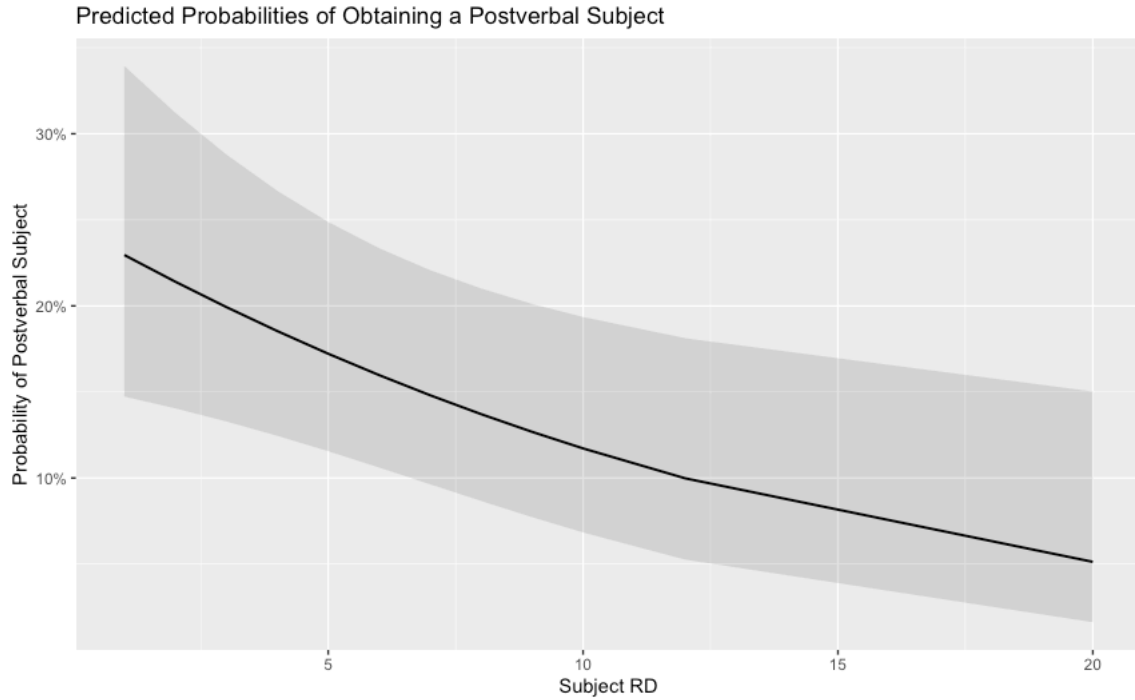


Figure 3: Model Predictions of the Probabilities of Obtaining a Postverbal Subject after Removing Non-Significant Factors

Figure 3 again confirms that there is a negative correlation between subject RD and the probability of observing a postverbal subject.

3.7 Direct Object Referential Distance

Moving on now to a discussion of direct objects, Table 11 displays the counts and percentages of both preverbal and postverbal objects for each RD value that I observed.

Referential Distance	Preverbal Objects ($N = 39$)	Percent Preverbal	Postverbal Objects ($N = 25$)	Percent Postverbal
1	7	43.75%	9	56.25%
2	3	60%	2	40%
3	3	75%	1	25%
4	1	50%	1	50%
5	2	66.67%	1	33.33%
6	0	0%	1	100%
7	0	0%	1	100%
10	0	0%	1	100%
14	1	100%	0	0%
20	21	72.41%	8	27.59%

Table 11: Counts and Percentages of Preverbal Objects and Postverbal Objects for Each Observed Value of RD

Looking at the percentages in Table 11 (columns 3 and 5), we can see that the percentage of preverbal objects generally increases as RD increases. On the other hand, the percentage of postverbal objects generally seems to decrease as RD increases. This is the same trend that we saw in the subjects dataset. Another notable similarity is the fact that the distribution of direct object RD also follows a bimodal distribution. The intermediate values are rather scarce, while a relatively high number of objects have an RD value of 1 or an RD value of 20. We can split the range of RD in half and compare the proportions of preverbal objects and the proportions of postverbal objects in both the low RD half and the high RD half.

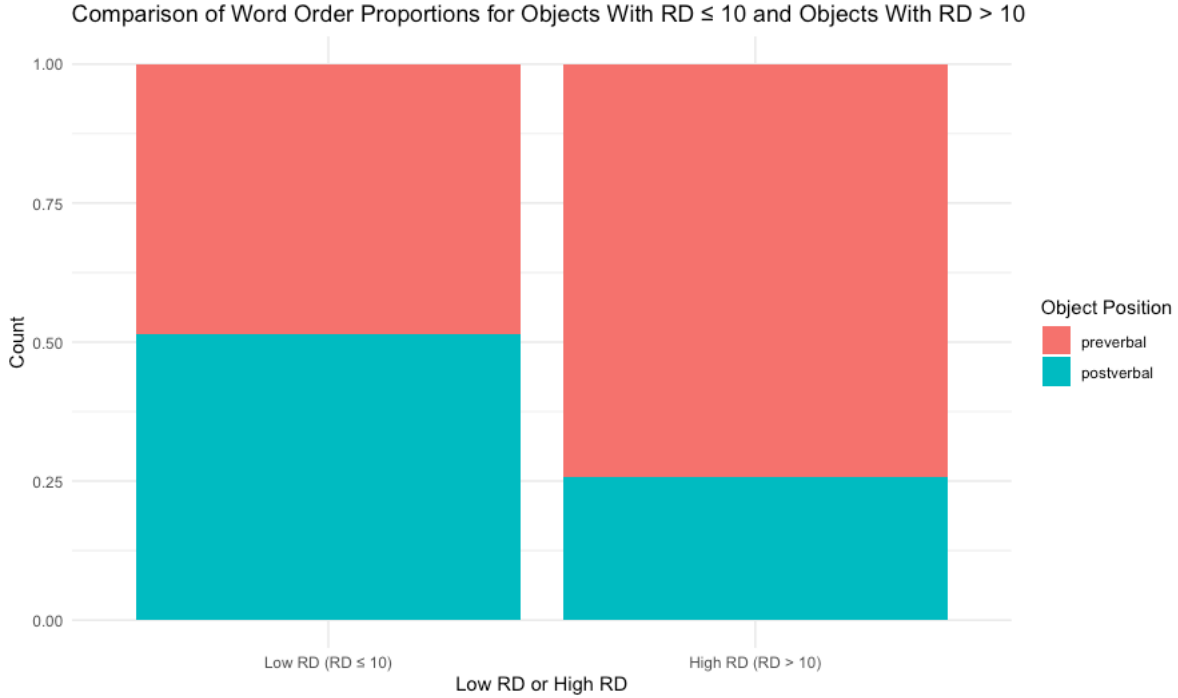


Figure 4: Comparison of Low RD Objects ($RD \leq 10$) and High RD Objects ($RD > 10$)

Figure 4 shows that low RD objects have a greater tendency to occur postverbally. Thus, this graph suggests that for direct objects, lower values of RD are correlated with postverbal position.

3.8 Direct Object Topic Persistence

Table 12 displays the counts and percentages of preverbal objects and postverbal objects for each TP value that I observed.

Topic Persistence	Preverbal Objects ($N = 39$)	Percent Preverbal	Postverbal Objects ($N = 25$)	Percent Postverbal
0	24	63.16%	14	36.84%
1	7	53.85%	6	46.15%
2	4	80%	1	20%
3	2	66.67%	1	33.33%
4	1	33.33%	2	66.67%
6	1	50%	1	50%

Table 12: Distribution of Topic Persistence Among Preverbal/Postverbal Direct Objects

In columns 3 and 5 (“Percent Preverbal” and “Percent Postverbal”), the percentages fluctuate without showing any clear trend. Furthermore, the number of objects with TP

values greater than 1 is rather small. Both of these observations suggest that object TP is not a significant predictor of object position.

3.9 Direct Object Thematic Continuity

Table 13 displays both the counts and percentages of preverbal/postverbal direct objects in each category of thematic continuity. “Break” indicates objects that occurred at the beginning of thematic paragraphs, and “continuous” indicates objects that occurred paragraph-medially or paragraph-finally.

Object Thematic Continuity	Preverbal Objects	Percent Preverbal	Postverbal Objects	Percent Postverbal
Break	6	66.67%	3	33.33%
Continuous	33	60%	22	40%

Table 13: Counts and Percentages of Preverbal/Postverbal Direct Objects Within Categories of Thematic Continuity

In Column 3, the percentage of preverbal objects in the “Break” category is rather close to the percentage of preverbal objects in the “Continuous” category (likewise for the percentages of postverbal objects in Column 5). The difference between the two categories ultimately comes down to about 6.67%. While it seems that thematically continuous objects occur postverbally more often than thematically discontinuous ones, the small difference suggests that thematic continuity may not be a significant predictor of object word order.

3.10 Predicate Transitivity and Direct Object Position

Of the 64 overt direct objects that I encountered, 57 of them occurred as one of two nominal arguments in a transitive construction. Three of them occurred with ditransitive verbs that take up to three nominal arguments. Because these ditransitive verbs were so scarce, I collapsed the category of transitive and ditransitive verbs together into one category, which I labelled as “(Di)transitive.”

On the other hand, only two direct objects occurred with verbs that take one nominal argument and one clausal complement, and two occurred with ditransitive verbs that take two nominal arguments and one clausal complement. Due to the extremely small number of referents in each of these categories, I collapsed both of the categories into one category labelled as “(Di)transitive With Clausal Complement.”

Note that because direct objects cannot occur with intransitive verbs, Table 14 does not have a row for intransitive predicates.

Predicate Transitivity	Preverbal Objects ($N = 39$)	Percent Preverbal	Postverbal Objects ($N = 25$)	Percent Postverbal
(Di)transitive	35	58.33%	25	41.67%
(Di)transitive With Clausal Complement	4	100%	0	0%

Table 14: Counts and Percentages of Preverbal/Postverbal Direct Objects in Each Transitivity Category

The data in Table 14 indicate that nearly all objects occurred with verbs that took only two nominal arguments or only three nominal arguments. There were also objects that occurred with verbs that took one or two nominal arguments and one clausal complement. However, this category was very small and only contained four objects. Thus, the number of objects outside of the “(Di)transitive” category is too small for me to make any inferences about correlations between predicate transitivity and object position.

3.11 Direct Object Animacy

For each level of animacy, Table 15 summarizes the counts and percentages of preverbal and postverbal direct objects.

Object Animacy	Preverbal Objects	Percent Preverbal	Postverbal Objects	Percent Postverbal
Inanimate	31	68.89%	14	31.11%
Animate	8	42.11%	11	57.89%

Table 15: Distribution and Percentages of Preverbal Objects and Postverbal Objects for Each Animacy Level

Table 15 reveals that there were 45 inanimate direct objects compared to 19 animate ones. The relatively high number of inanimate objects could be due to a tendency for objects to be acted upon. In addition, we can also observe that in Column 5 (“Percent Postverbal”), the percentage of postverbal animate objects is 26.78% higher than the percentage of postverbal inanimate objects ($57.89\% - 31.11\% = 26.78\%$). This hints at a possible correlation between animate objects and postverbal word order, with animate objects occurring postverbally significantly more often than inanimate ones.

3.12 Logistic Regression Model for Direct Objects

To summarize my observations of direct objects from Sections 3.7 to 3.11, my examination of object RD suggested that lower values of RD were correlated with postverbal position for objects. Animacy also suggested a possible correlation, with animate objects appearing postverbally more often than inanimate ones. On the other hand, TP and thematic continuity did not show any clear trends. As for predicate transitivity, the number of objects

occurring outside of the “(Di)transitive” category was extremely small, so no clear correlations between predicate transitivity and object position can be inferred from this data.

I will now turn to a discussion of the logistic regression model that I implemented for direct objects. The logistic regression model for direct objects uses the same predictors as the logistic regression model for subjects (i.e. RD, TP, thematic continuity, predicate transitivity, and animacy). The results of running this model with R’s `glm()` function are displayed in Table 16.

Term	Coefficient Estimate	Standard Error	Z Statistic	P Value	N
Intercept	-0.83687	1.08639	-0.770	0.4411	64
Direct Object RD (Baseline: RD = 0)	-0.06523	0.03422	-1.906	0.0566	64
Direct Object Persistence (Baseline: TP = 0)	0.03318	0.22025	0.151	0.8803	64
Object Animacy = Inanimate (Baseline)	N/A	N/A	N/A	N/A	45
Object Animacy = Animate	1.90608	0.75533	2.523	0.0116	19
Transitivity = Transitive (Baseline)	N/A	N/A	N/A	N/A	60
Transitivity = Transitive With Clausal Complement	-19.22558	1944.30869	-0.010	0.9921	4
Thematic Continuity = Break (Baseline)	N/A	N/A	N/A	N/A	9
Thematic Continuity = Continuous	0.82457	0.97775	0.843	0.3990	55

Table 16: Results of Logistic Regression Model for Direct Objects

Based on the p values of this model, only one predictor was a significant predictor of direct object word order: animacy. When direct objects were animate, they were significantly more likely to occur postverbally. Figure 5 helps to illustrate this trend.

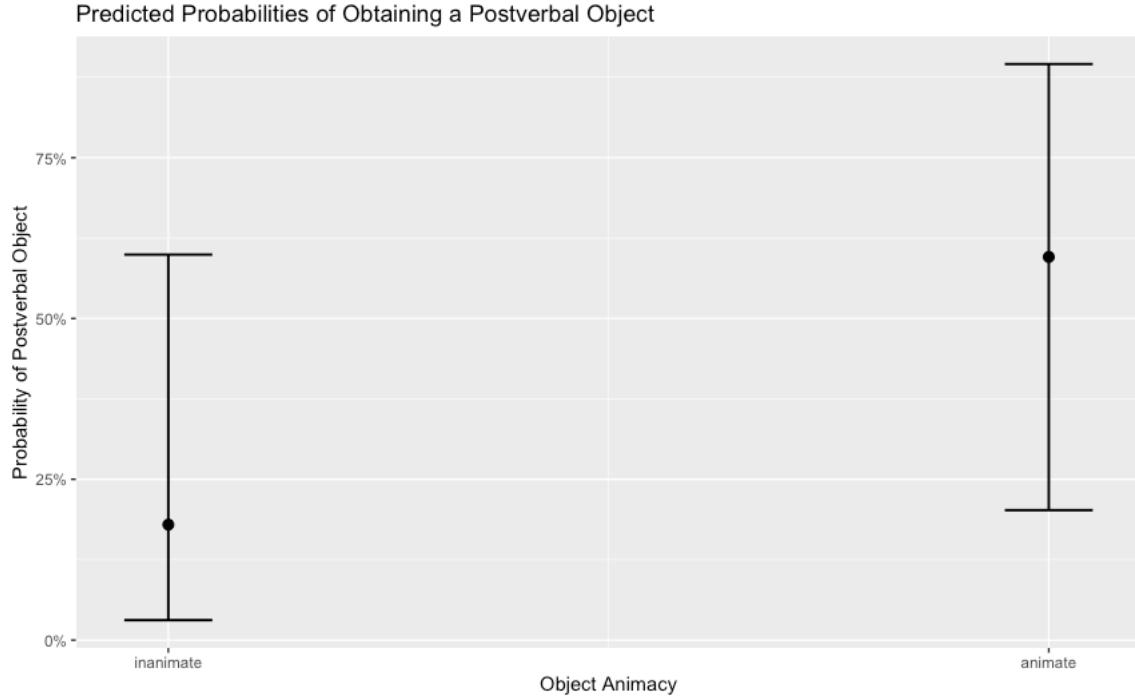


Figure 5: Predicted Probabilities of Obtaining a Postverbal Object Based on Animacy

As we can see from Figure 5, the logistic regression model for direct objects predicts that animate objects are significantly more likely than inanimate ones to appear postverbally.

Although this predictor was the only one to reach statistical significance, direct object RD was also very close to reaching statistical significance. The results indicated that as direct object RD increased, direct objects became less likely to occur postverbally. However, because the p value for direct object RD is not less than 0.05, I cannot reject the null hypothesis that RD and direct object word order have no relationship.

Similar to the logistic regression model for subjects, the logistic regression model for objects also had many non-significant predictors. However, after performing stepdown model comparison by removing predictors other than object RD, I still found that object RD was not a significant predictor of object position. Based on these results, stepdown model comparison did not seem to yield any additional insights.

4 Discussion

In Section 3, the logistic regression model for subjects found that RD was a significant predictor of subject position. Specifically, higher subject RD values were correlated with preverbal position while lower values of subject RD were correlated with postverbal position. The other four predictors (TP, thematic continuity, animacy, and predicate transitivity) did not have a significant effect on subject word order. On the other hand, although object RD seemed to follow the same trend as subject RD (i.e. higher object RD was correlated with preverbal word order), the logistic regression model for objects did not find that this was a

significant predictor. This model did, however, find that animate objects were significantly more likely to occur postverbally than inanimate ones.

These findings on subject RD and object animacy could be attributed to either language specific explanations or language general explanations. For example, to explain the correlation between subject RD and subject position, one may propose a Karuk specific explanation which states that Karuk uses marked word order to encode subjects with low RD. This could explain why low RD subjects tend to occur postverbally more often than preverbal ones, since Mikkelsen (2014, p. 1) argues that the unmarked word order in Karuk is verb final. In this section, I will argue that the tendency for lower RD subjects to appear postverbally is not related to the markedness of postverbal position. Instead, I will argue that postverbal position is generally used to encode referents with low RD in multiple languages with pragmatically controlled word order. On the other hand, I will argue that animate objects occur postverbally more frequently because they are marked by animacy and therefore appear in a marked word order.

Section 4.1 discusses the implications of my findings concerning RD and subject position. Additionally, it compares my findings with those from other languages, in particular Chamorro, Klamath, and Ute. Section 4.2 discusses the correlation between animacy and object position. Afterwards, Section 4.3 discusses the predictors which did not have a statistically significant effect. Finally, Section 4.4 summarizes the discussion of my findings.

4.1 Correlations Between RD and Word Order

Givón (1983b, p. 19–20) proposes the word order hierarchies in (25) in order to connect the continuity of a subject/object and word order. These scales are arranged in order of decreasing continuity of the nominal argument. Because referents with lower RD values are more continuous, these scales imply that lower RD subjects/objects are more likely to occur postverbally.

- (25) a. VS (most continuous) > SV (less continuous)
b. VO (most continuous) > OV (less continuous)

Givón (1983b, p. 19) claims that evidence for this hierarchy comes from languages such as Latin-American Spanish (a Romance language; this study’s data consists of transcriptions from Mexico City, Caracas, and Santiago de Chile) (Bentivoglio, 1983), Biblical Hebrew (a Semitic language from Israel) (Fox, 1983), and Ute (Givón, 1983a).

The hierarchy in (25) is a more specific instance of the following hierarchy, which lists different ways of ordering comments and topics in order of decreasing continuity of the topic:

- (26) Comment > Comment-Topic > Topic-Comment > Topic (Givón, 1983b, p. 20)

In (25), the verb is the comment and the subject/object is the topic.

Givón claims that this scale can be explained by the following psychological principle:

- (27) “Attend first to the most urgent task” (Givón, 1983b, p. 20)

When topics are highly continuous (for example when they have very low RD values), they are dropped because the comment is much more important and the topic can be easily inferred from context. For cases where they are still relatively continuous, the comment is

still more important than the topic, so it will be mentioned first. For topics that are even less continuous, the topic will be mentioned first because identifying it is “the most urgent task” at hand. Finally, on the most discontinuous end of the scale in (26), topics that are highly discontinuous will be repeated without their comment.

My findings concerning RD in Karuk, along with studies on Ute (Givón, 1983a), Klamath (Meyer, 1992), and Chamorro (Cooreman, 1992), provide additional evidence for the scale in (26). For example, Givón (1983a) reports that postverbal position in Ute favors continuous referents. Similarly, Meyer (1992) also notes that postverbal position typically encodes more continuous referents in Klamath, including those with lower RD values. Regarding subject referents, Meyer (1992) finds that postverbal subjects on average have an RD of 3.92, while preverbal subjects on average have an RD of 7.82.

Furthermore, while Cooreman (1992) found that subject RD in general did not have a significant effect on word order, she did find that subject RD had a significant impact on word order among subjects within clauses that maintained thematic continuity (i.e. paragraph-medial or paragraph-final clauses). This finding suggests that while thematic continuity was the most significant predictor of word order, whenever thematic continuity was maintained, SV word order marked disruptions in continuity while VS word order marked subjects that maintained continuity. Thus, these findings, while slightly different from those of Givón (1983a) and Meyer (1992), do still support the claim that postverbal position commonly marks continuous referents.

The continuity of referents and its effect on word order may be related to the concept of thematic information from Tomlin and Rhodes (1992). Tomlin and Rhodes (1992, p. 117) define thematic information as “information which is central to the development of a particular text or sub-text,” and they note that this type of information plays a role in influencing word order. In contrast to thematic information, rhematic information includes information which is not central to a text’s development. Tomlin and Rhodes (1992) demonstrate that in Ojibwa (an Algonquian language spoken near Ontario province in Canada), thematic information follows rhematic information. This finding could help to explain why referents with low RD appeared postverbally more often than those with high RD. If we assume that referents with lower RD tend to occur more frequently, then lower RD referents are more likely to constitute thematic information. This is because referents that are important to a text are more likely to be frequently mentioned. Thus, if thematic information tends to occur postverbally, then lower RD referents would also tend to occur postverbally. However, RD is not a perfect measurement of a referent’s importance to a text’s development, as it is entirely possible for a highly unimportant referent to occur two clauses in a row and then stop occurring. Section 5 discusses an alternative measurement for directly measuring a referent’s frequency and hence its centrality in a text’s development.

So far, we have seen that within languages with pragmatically controlled word order, low RD referents tend to occur postverbally while high RD referents tend to occur preverbally. Many of these languages differ in terms of their unmarked word order, suggesting that the pattern noted above may be unrelated to markedness. According to Mikkelsen (2014), Karuk is a verb final language. Evidence in support of this claim include (but is not limited to) the following observations: adverbial complements must be preverbal, verb final word order is offered in unmarked elicitation contexts, and a majority of clauses in Karuk are verb final. Based on these observations, postverbal position in Karuk is a marked position for both

subjects and objects. This means that subjects with low RD have a greater tendency to occur in a marked position (i.e. postverbally). Similar to Karuk, Ute is a primarily SOV language with pragmatically controlled word order (Givón, 1983a, p. 145). This study makes a similar observation on word order, noting that low RD referents tend to prefer postverbal position. In both languages, then, we see that low RD referents have a tendency to appear postverbally in a marked position.

On the other hand, Cooreman (1992, p. 244) notes that Chamorro has a basic word order of VSO. This would mean that preverbal position is marked. Cooreman (1992, p. 252) finds that subject RD does significantly influence word order for subjects in paragraph-medial and paragraph-final position. Among these subjects, those with low RD were more likely to appear postverbally while those with high RD were more likely to appear preverbally. In terms of subject RD, the difference between Karuk and Chamorro is that in Chamorro, subjects with low RD occur in an unmarked position, while in Karuk, subjects with low RD occur in a marked position. This seems to suggest that the position of a referent with high RD/low RD is not related to how marked its position is relative to the language's unmarked word order. Rather, it seems that regardless of what the unmarked word order is, preverbal position seems to favor high RD referents while postverbal position tends to encode low RD referents.

The correlation between preverbal position and new information may be related to focalization. Davis et al. (2020, p. 849-850) argue that focused constituents must occur preverbally when marked by a focus particle. In addition, they also indicate that any focused constituent, regardless of whether it is marked by a focus particle or not, must occur preverbally. This indicates that preverbal position serves a purpose of emphasizing focused information. If we assume that new information (including information that was absent for a long time) is often focused, then this would help explain why high RD subjects had a greater tendency to occur preverbally. If new information is correlated with focus, and focus must be preverbal, then there is a correlation between new information and preverbal position.

While RD was not a significant predictor of object word order, object RD still followed the same trend as the one found in the studies discussed above. That is, objects with low RD values tended to appear postverbally more often than those with higher RD values. However, the p value was just short of significance ($p = 0.0566$), so I cannot reject the null hypothesis that RD is not significantly correlated with word order.

There are multiple possible explanations for why object RD ultimately did not have a significant effect. One such possibility is that the sample of objects was too small. Within my dataset, 64 of the clauses that I collected had overt objects; on the other hand, I found 146 clauses with overt subjects. It may be possible that the lack of a significant result was due to chance, and that collecting more data could find that object RD was also a significant predictor of object word order. It is also possible that objects and subjects in Karuk behave differently in terms of the pragmatic factors that influence their word order. Section 4.2 discusses the observation that animacy had a significant effect on object word order but not on subject word order. It may be the case that word order marks different kinds of subjects and objects. For subjects, continuity (measured via RD) may be more significant in influencing word order, while animacy, as opposed to RD, is more influential on object position.

In summary, the findings from my study, along with those of Givón (1983a), Meyer

(1992), and Cooreman (1992), all seem to suggest that postverbal position favors referents with low RD while preverbal position favors referents with high RD. However, my study only finds that subject RD is a significant predictor of subject position. These observations also suggest that Karuk subjects follow a rheme-theme distribution of information that is similar to the information distribution of Ojibwa (Tomlin and Rhodes, 1992).

4.2 Animacy and Direct Object Word Order

While RD had a significant effect on subject word order, animacy only had a significant effect on the position of objects. This finding may be connected to the markedness of animate objects and the findings of Aissen (2003). Animate objects in my dataset were in a minority compared to inanimate objects, with 19 animate objects (and 45 inanimate ones) out of a total of 64 objects. The relative scarcity of animate objects already suggests that they constitute a marked category.

Aissen (2003) discusses the role that animacy plays in a phenomenon known as Differential Object Marking (DOM). In cases of DOM, some objects that are more marked than others receive marked structure (e.g. through case marking), and Aissen (2003, p. 12) captures this relationship using iconicity constraints. Objects in higher animacy levels are more marked, and this markedness results in marked structure. Aissen (2003, p. 8) lists the following animacy scale, which defines the relative ranking of three levels of animacy:

(28) Animacy scale: Human > Animate > Inanimate (Croft, 1988)

Thus, based on this scale, animate objects are more marked than inanimate ones. Based on the iconicity relationship from Aissen (2003), this would mean that animate objects are more likely to appear with marked structure. Aissen (2003) discusses several examples of DOM involving case-marking, revealing how (by the iconicity constraints) marked objects receive marked structure via case-marking. For instance, Aissen (2003, p. 19) states that Sinhalese only allows case-marking on animate referents. In Hindi, case-marking is allowed on human objects and animate objects denoting ‘higher animals,’ e.g. elephants and lions (Mohanani, 1993, p. 28). These cases, along with the others discussed by Aissen (2003), illustrate how the iconicity principle results in marked objects receiving marked structure via case.

Based on the iconicity principle, we may also expect cases of DOM that utilize other kinds of structure, such as word order, to indicate marked objects. Indeed, Aissen (2003, p. 35–36) discusses the possible relationship between DOM and object shift. A number of these cases have constraints that resemble the definiteness constraints on DOM discussed in Aissen (2003). For example, in Persian, while definite objects are permitted to shift, indefinite objects can only undergo object shift if they are interpreted as specific objects (Browning and Karimi, 1994).

If word order truly can be a way to express marked substance, then a possible explanation for the effect of animacy on the position of Karuk direct objects emerges. Mikkelsen (2014, p. 1) argues that the unmarked word order of Karuk is verb-final. This entails that postverbal objects are more marked than preverbal ones. Thus, by the iconicity principle, because animate objects are marked, they are more likely to appear postverbally.

The tendency for animate objects to appear postverbally in Karuk differs from the cases of object shift that are discussed in Aissen (2003). For example, animate objects in Karuk

tend to appear to the right of the verb rather than within the left periphery. Furthermore, many cases of object shift involve the definiteness of an object, rather than its animacy (Aissen, 2003, p. 36). Despite these differences, there are some notable similarities. In both cases, marked objects are signalled by marked word order. Additionally, one may also speculate that postverbal objects in Karuk appear outside of the VP. Thus, prototypical cases of object shift and postverbal appearances of animate objects in Karuk could both involve movement of an object to a location outside of its base position in the VP.

4.3 Discussion of Null Results

While RD had a significant effect on the position of subjects, TP did not demonstrate a significant effect on either subject position or object position. This contrast may be due to the different purposes of RD and TP. According to Givón (1983b, p. 13-14), RD is supposed to correlate with the difficulty of identifying a referent and filing it in memory. On the other hand, TP is a measurement of a topic's importance and assumes that more important discourse topics "have a higher probability of persisting longer in the register after a relevant measuring point" (Givón, 1983b, p. 15). The significance of RD as a predictor of subject position seems to indicate that the preceding discourse has a stronger impact on the position of subjects than the subsequent discourse. Based on the purposes of RD and TP, this entails that the accessibility of a nominal argument, rather than its importance in the upcoming discourse, is more influential on its position with respect to its verb.

Findings concerning TP varied within other studies on languages with pragmatically controlled word order. Similar to my study, Rude (1992, p. 205) finds that TP did not have a significant influence on word order in Nez Perce. On the other hand, Givón (1983a, p. 186-187) finds that for subjects, postverbal subjects on average had lower TP values than preverbal subjects. Additionally, Givón (1983a, p. 187-188) also finds that preverbal objects on average had lower TP values than postverbal objects. Without any statistical tests, however, we do not have a way of determining whether this is a genuine effect.

Following Cooreman (1992), my study also incorporated a measure for thematic continuity, which assessed whether a particular referent occurred at the beginning of a thematic paragraph or paragraph medially/finally. While my study ultimately found that thematic continuity was not a significant predictor of subject or object word order, Cooreman (1992, p. 253) does find that thematic continuity is a statistically significant predictor of subject word order in Chamorro. It may be the case that within Chamorro, preverbal position can serve the purpose of signalling new lines of action or drastic shifts in themes and settings. These large shifts would then indicate boundaries of thematic paragraphs. On the other hand, preverbal position does not seem to serve this purpose in Karuk.

Predicate transitivity is another predictor which ultimately did not have a significant effect on subject position and object position. With regards to objects, nearly all direct objects occurred with transitive predicates, so there was not enough data outside of the transitive category to make any inferences about potential correlations between object position and predicate transitivity. Subjects, on the other hand, had more data in multiple categories but still found that predicate transitivity was not a significant predictor of subject word order. This is because the proportions of preverbal/postverbal subjects in all transitivity categories was nearly the same.

4.4 Summary

In summary, RD was the only significant predictor of subject position, and lower RD values were correlated with postverbal subjects. At first, this finding seemed to hint at a connection between marked word order and low RD, as postverbal position is marked in Karuk. However, findings from Ute (Givón, 1983a), Klamath (Meyer, 1992), and Chamorro (Cooreman, 1992) also found that postverbal position favored low RD referents, even though these languages differed in terms of their unmarked word order (in the case of Chamorro, this correlation was only statistically significant when thematic continuity was maintained). Thus, in languages with pragmatically controlled word order, postverbal position seems to encode referents with low RD, and the tendency for low RD subjects to appear postverbally in Karuk does not seem to be due to a language-specific explanation in terms of markedness.

On the other hand, animacy was found to be the only statistically significant predictor of object position, while RD was not a significant predictor. In this case, the tendency for animate objects to appear postverbally does seem to be due to the markedness of animate objects. According to the iconicity constraints of Aissen (2003), marked objects are more likely to receive marked structure. In the case of Karuk, animate objects are marked and have a tendency to occur in a marked position, so postverbal position (as opposed to case) seems to be a reflection of the markedness of animate objects.

5 Conclusion

As a non-configurational language, Karuk has the potential to realize every word order permutation of subjects, verbs, and objects. This study sought to answer the following question: Is Karuk word order truly random, and if it is not, what factors influence the preverbal or postverbal realization of nominal arguments in Karuk? The factors that I examined are as follows: RD (referential distance), TP (topic persistence), thematic continuity, predicate transitivity, and animacy.

By constructing a multivariate logistic regression model for subjects and objects, I was able to obtain the following results: subject referential distance was negatively correlated with postverbal position ($p = 0.00545$). That is, as subject RD increased, the chances of a subject appearing postverbally decreased. With regards to subjects, no other predictor had a significant effect. On the other hand, object RD also seemed to be negatively correlated with postverbal position; however, this result was just short of significance ($p = 0.0584$). Even after I removed the non-significant factors from the model, object RD still did not have a significant effect on object position. Thus, I cannot reject the null hypothesis that object RD and object position are not correlated. While object RD did not have a significant effect, object animacy did. The logistic regression model for objects found that animate objects had a significantly higher chance of appearing postverbally than inanimate objects ($p = 0.0115$).

My findings concerning subject RD resemble the findings on RD from Ute (Givón, 1983a), Klamath (Meyer, 1992), and Chamorro (Cooreman, 1992). In several languages with pragmatically controlled word order, referents with high RD tend to occur more often in preverbal position while referents with low RD tend to occur more often in postverbal position (Givón, 1983b). While the findings from the subjects dataset in my study help to support this claim,

it is interesting to note that the objects dataset fell just short of showing statistically significant support for this claim. This could point to a difference between subjects and objects in terms of how sensitive they are to certain pragmatic factors. Alternatively, however, this may be the result of random chance. The object dataset was significantly smaller than the subjects dataset (64 objects compared to 146 subjects), so finding a significant effect would be more difficult. Additional data on objects could help verify whether or not object RD is a significant predictor of object position, and it is entirely possible that larger datasets of objects will find more significant results.

While collecting more data can confirm whether or not a negative result truly is negative, it can also be a good way to verify positive results. While this study did find a significant effect for subject RD and object animacy, additional quantitative and statistical studies should be carried out to verify whether there is truly a connection between RD and word order or animacy and word order. Winter (2020, p. 177-178) notes that it is possible for studies to obtain a statistically significant result even when the null hypothesis is true. Even though there is always a chance that a statistically significant result was obtained due to chance, additional studies can bolster a theory and strengthen the connection between a set of independent variables and dependent variables. Thus, future research can expand my dataset and use even more clauses to examine the effect of the pragmatic factors that I investigated.

One notable way to expand my study's dataset would be to incorporate additional genres. My current study only examined texts in the narrative genre. Thus, a possible extension of my study would be to examine the effect of pragmatic factors on word order in other genres. *Ararahih'urípih* does contain several other genres, such as dialogues and procedural texts. Narratives as a genre are unique because they are memorized and retold. Furthermore, many of the characters within narratives are culturally very significant (e.g. *pihneefich* 'Coyote'). Thus, even when they occur for the first time in a narrative, they are likely easier to retrieve than non-character referents and are definite. The salience of mythological characters in the narratives that I examined possibly influenced their probability of occurring preverbally or postverbally, so examining other texts that do not involve these culturally salient referents could yield different results. In order to compare the effects of pragmatic factors in various genres, it is important to collect enough data from each genre. In other words, a large number of clauses should be drawn from each genre in order for samples to be compared.

In addition to examining more data, additional predictors/measurements may be considered, or the factors within my dataset can be modified. In addition to RD, TP, and thematic continuity, Givón (1983b) also provides another predictor that could potentially influence word order. This predictor is potential interference, also referred to as potential referent interference. Givón (1983b, p. 14) defines this predictor as a way to assess the effect that other referents in the immediately preceding register can have on topic identification. Here, the immediately preceding register is arbitrarily defined as one to five clauses to the left. In this range, a referent is only counted if it is just as compatible as the referent under consideration in terms of its compatibility with the clause's predicate.

I also propose that the frequency of a referent's occurrence can yield insight into the ways in which pragmatic factors influence word order. I will refer to this measurement as **referential frequency**, or RF. There are two ways in which this measurement can be defined: globally for a text, or *thus far* in a text. RF can be computed globally for a particular

referent by dividing the number of occurrences of that referent by the total number of clauses within the text. The resulting number can be thought of as the average rate at which the referent under consideration occurs in a particular text. Highly important referents, such as main characters, are likely to be referenced extremely frequently, and so they will have much higher RF values. To connect this with word order, one can examine whether a referent's global RF is correlated with the percentage of postverbal/preverbal overt occurrences within a text.

Alternatively, RF *thus far* can be computed for specific occurrences of a referent within a text. It is calculated by taking the number of occurrences of a referent up until and including the current one and dividing this by the number of clauses that have elapsed so far, including the current one. Both global RF and RF *thus far* are proportions, so they take on values between 0 and 1 (or between 0 and 100 if converted to a percentage).

RF incorporates the concept of thematic information from Tomlin and Rhodes (1992). Tomlin and Rhodes (1992) define thematic information as information which contributes to the development of a text or sub-text. While RD can attempt to measure this by identifying referents that occur closely together, it is not a direct measurement of a referent's importance in a text or the extent to which it contributes to the text's development. It is possible for two occurrences of a referent to occur in two consecutive clauses but nowhere else in a text. In this case, it is unlikely for the referents to contribute much to a text's development.

For languages like Ojibwa that express rhematic information before thematic information, I would propose the following hypotheses concerning global RF and RF *thus far*, respectively: overt referents with higher values of global RF will have a higher percentage of appearances occurring postverbally, and overt referents with higher values of RF *thus far* are more likely to occur postverbally within their clause. By providing a quantitative measure of a referent's frequency of occurrence and hence its importance, RF can provide another way to explore how pragmatic factors can play a role in influencing the position of nominal arguments with respect to their verbs.

References

- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, 21:435 – 483.
- Bentivoglio, P. (1983). Topic continuity and discontinuity in discourse: A study of spoken Latin-American Spanish. In Givón, T., editor, *Topic Continuity in Discourse*, pages 255 – 312. John Benjamins, Amsterdam.
- Bright, W. (1957). *The Karok Language*. University of California Press, Berkeley.
- Browning, M. and Karimi, E. (1994). Scrambling to object position in Persian. In Corver, N. and Riemsdijk, van, H., editors, *Studies on Scrambling. Movement and Non-Movement Approaches to Free Word-Order Phenomena*, pages 61 – 100. Mouton de Gruyter, Berlin.
- Cooreman, A. (1992). The pragmatics of word order variation in Chamorro narrative text. In Payne, D. L., editor, *Pragmatics of Word Order Flexibility*, pages 243–263. John Benjamins, Amsterdam.
- Croft, W. (1988). Agreement vs. case marking and direct objects. In Barlow, M. and Ferguson, C., editors, *Agreement in Natural Language: Approaches, Theories, Descriptions*, pages 159 – 179. CSLI, Stanford.
- Davis, C. S., Smith, V., Super, N. n. J., Super, P. S., Thom, C. S., and Mikkelsen, L. (2020). Forms and functions of backward resumption: the case of Karuk. In *Language*. Linguistics Society of America.
- Fox, A. (1983). Topic continuity in Biblical Hebrew narrative. In Givón, T., editor, *Topic Continuity in Discourse*, pages 215–254. John Benjamins, Amsterdam.
- Garret, A., Gehr, S., Maier, E. H., Mikkelsen, L., Richardson, C., and Sandy, C. S. (2020). Karuk. In Jany, C., Mithun, M., and Rice, K., editors, *The Languages and Linguistics of Indigenous North America: A Comprehensive Guide*, pages 13–15. De Gruyter Mouton, Berlin.
- Givón, T. (1983a). Topic continuity and word-order pragmatics in Ute. In Givón, T., editor, *Topic Continuity in Discourse*, pages 143–214. John Benjamins, Amsterdam.
- Givón, T. (1983b). Topic continuity in discourse: An introduction. In Givón, T., editor, *Topic Continuity in Discourse*, pages 1–41. John Benjamins, Amsterdam.
- Golla, V. (2011). *California Indian Languages*. University of California Press, Berkeley.
- Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory*, 1:5 – 47.
- Macaulay, M. (2004). On the Karuk directional suffixes. In Ettliger, M., Fleisher, N., and Park-Doob, M., editors, *Proceedings of the Thirtieth Annual Meeting of the Berkeley Linguistics Society: Special Session on the Morphology of Native American Languages*, volume 30, pages 85–101. Berkeley Linguistics Society, Berkeley.

- Maier, E. (2020). Karuk treebank annotation guidelines. Previous revision on April 3, 2020.
- Maier, E. (To appear). *A treebank of the Karuk Language*. PhD thesis, University of California, Berkeley.
- Meyer, K. S. (1992). Word order in Klamath. In Payne, D. L., editor, *Pragmatics of Word Order Flexibility*, pages 167–192. John Benjamins, Amsterdam.
- Mikkelsen, L. (2014). What goes postverbal in a verb-final language? The interplay of syntactic category, information structure, and word order in Karuk. Accessed at <http://linguistics.berkeley.edu/~mikkelsen/talks.html>.
- Mohanan, T. (1993). Case alternation on objects in Hindi. *South Asian Language Review*, 3:1 – 30.
- Rude, N. (1992). Word order and topicality in Nez Perce. In Payne, D. L., editor, *Pragmatics of Word Order Flexibility*, pages 193–208. John Benjamins, Amsterdam.
- Tomlin, R. and Rhodes, R. (1992). Information distribution in Ojibwa. In Payne, D. L., editor, *Pragmatics of Word Order Flexibility*, pages 117 – 135. John Benjamins, Amsterdam.
- Tonhauser, J. and Colijn, E. (2010). Word order in Paraguayan Guaraní. *IJAL*, 76:255–288.
- Winter, B. (2020). *Statistics for linguists: an introduction using R*. Routledge, New York.