

1 INTRODUCTION

Tshivenda (Guthrie S21,¹ ~1.3m speakers in South Africa (Limpopo Province) & Zimbabwe.), ~1.3m speakers in South Africa & Zimbabwe) shares with other Southern Bantu languages a distinctive morphological alternation in the form of the present tense prefix, traditionally termed the **conjoint / disjoint alternation**. As shown below, the simple present is expressed either by the prefix /a-/ (termed the disjoint form) or the /ø-/ (termed the conjoint).

- | | | | | |
|-----|----|--|----|-------------------------------------|
| (1) | a. | ndi (a) ɭa ɲemeɲeme
1sg A eat termite
“I eat termite.” | b. | ndi *(a) ɭa
1s A eat
“I eat.” |
|-----|----|--|----|-------------------------------------|

I will present the results of recent fieldwork showing that the availability of the disjoint prefix in Tshivenda, unlike in the other Southern Bantu languages, shows a three-way distribution: It's obligatory in some contexts, impossible in others, and optional elsewhere. In other languages with the alternation, e.g. isiZulu (Halpert 2016), the conjoint and disjoint forms are typically in complementary distribution – no optionality is possible.

I will also show that the same three-way distribution also shows up in the domain of prosody. Penultimate lengthening is a phenomenon common to many Bantu languages and applies to some large prosodic unit (typically called the intonational phrase; see Hyman 2009). Previous scholarship on Tshivenda has found that the penultimate syllable of the utterance is always lengthened and that some utterance-internal penults may also be lengthened (Polous 1990, Cassimjee 1992, Westphal 1962); the conditions under which this utterance-internal lengthening occurred have not been fully documented previously. My own fieldwork has shown that this internal lengthening shows a three-way split in frequency: In some syntactic contexts it is frequent, in others rare, and in still others does not occur. Strikingly, the contexts conditioning these three frequencies are the same as those conditioning the disjoint prefix: The syntactic contexts which trigger frequent internal lengthening are exactly those which require the disjoint prefix; those which disallow lengthening require the conjoint; and those with an intermediate frequency of lengthening allow either form of the verb.

I will propose in this paper that the coincidence of these two distributions may not be an accident, but rather that the presence or absence of the disjoint prefix in Tshivenda is a prosodic phenomenon in the sense that it is conditioned by the position of the verb within the prosodic phrasing, not by a purely syntactic process. In particular, I will propose the following generalization:²

- (2) **Conjoint / disjoint generalization (Tshivenda):**
- a. **Disjoint (/a-/):** appears when the verb is last in an ι P.
 - b. **Conjoint (/ø-/):** appears elsewhere.

In addition, I will argue that this phenomenon lends support to the model of prosody argued for in Selkirk & Lee (2015), in which prosodic structure formation is a distinct process from, and strictly precedes, structure-sensitive phonology. Perhaps controversially, this model allows for the possibility of prosodic structure which goes entirely unmarked by any particular phonological process. In section 6, I will argue that something like this must be happening in Tshivenda: While the distributions of utterance-internal penult lengthening and the disjoint prefix allow us to diagnose the presence of an intonational phrase boundary in certain syntactic structures, in any given utterance that boundary may be marked by neither lengthening (which is a variable process) nor the disjoint prefix (which only appears in the present tense).

¹The Guthrie scheme divides Bantu languages into areal groups (indexed by a letter), then into loosely-genetic groups indexed by the first digit (called decades), with a second digit uniquely identifying the particular language. Group S covers Southern Africa; Tshivenda is alone in decade 2, indicating a comparatively distant relationship from the other languages in its geographical region. For more information on the Guthrie system, see Nurse & Phillipson (2003).

²This proposal closely mirrors one made in Cheng & Downing (2009) for isiZulu. However, Halpert (2016) has convincingly argued that the isiZulu case cannot be prosodic in nature and must have a purely syntactic origin.

This analysis has bearing on the issue of direct vs indirect reference in the prosody (Selkirk & Lee 2015; Pak 2008). This question is ultimately a question of what syntactic information is available to the structure-sensitive phonology. In a direct reference system, the phonology has access to the syntactic structure itself, and may make use of syntactic notions like sisterhood, domination, c-command, and labelling in its calculations. In an indirect system, the phonology has no direct access to the syntax; instead, a process of prosodic structure formation creates an intermediate structure. In a direct reference theory, the analysis presented here has no way to unify the various syntactic conditions on the disjoint prefix; in an indirect reference theory, the prosodic structure provides the unifying condition. Combining an indirect reference theory with a theory of phonological variability naturally yields a theory in which the prosodic structure may be present but not overtly marked in the speech signal.

The structure of this paper is as follows. In section 2, I will discuss the disjoint alternation in Tshivenda, comparing and contrasting with other Southern Bantu languages. I will then present in section 3 the results of a survey on the acceptability of conjoint and disjoint verb forms in different syntactic contexts, showing that there is a three-way split in the availability of this prefix by syntactic context. In section 4, I'll go on to discuss the results of a study on sentence-internal penult lengthening across a variety of syntactic contexts, showing that the same three-way split in the distribution emerges. In section 5 I will present a model of prosodic structure and structure-sensitive phonology which can account for this three-way split and relate the disjoint alternation to prosodic phenomena. Finally, in section 6, I will discuss the consequences of this model for the broader theory of prosody.

2 THE CONJOINT / DISJOINT ALTERNATION

Southern Bantu languages frequently show an alternation in the form of the verb under certain tenses. For instance, in isiZulu³, the simple present takes a prefix /ya-/ in some contexts, but is /ø-/ elsewhere:

- | | | | | | |
|-----|----|--|-----|----|--|
| (3) | a. | uMlungisi u- pheka iqanda
M. 3s- cook egg
“Mlungisi is cooking an egg.” | (4) | a. | *uMlungisi u- pheka
M. 3s- cook |
| | b. | *uMlungisi u- ya- pheka iqanda
M. 3s- YA- cook egg | | b. | uMlungisi u- ya- pheka
M. 3s- YA- cook
“Mlungisi is cooking.” |

The short form of the verb /ø-/ is traditionally termed the ‘conjoint’ form; the long form /ya-/ is called the ‘disjoint’. Halpert (2016) gives the following generalization for the distribution of these forms:

- (5) **Conjoint-disjoint generalization (isiZulu):**
- a. **Conjoint (ø):** appears when vP contains material (after A movement)
 - b. **Disjoint (ya):** appears when vP does not contain material (after A movement)

Note two key properties of this generalization:

1. The conjoint and disjoint forms of the verb are in complementary distribution.
2. The distribution is predictable based on syntactic context.

This seems to be the norm across Southern Bantu: The disjoint alternation is a strictly syntactic fact. In fact, in isiZulu and other languages the alternation appears in several different tense / aspect / polarity combinations with different morphological realizations, but with the same structural generalization governing which form is realized. In Tshivenda, by contrast, the disjoint alternation only appears in the simple present tense (Polous 1990) – all other tense / aspect / polarity combinations do not alternate. Polous gives the following generalization about the distribution of the disjoint prefix:

³ All Zulu examples taken from Halpert (2016).

- (6) **Conjoint-disjoint generalization (Tshivenda, to be revised):**
- a. The disjoint is available everywhere.
 - b. The conjoint is ungrammatical when the verb is last in the sentence.

In contrast to isiZulu, this generalization does not place the conjoint & disjoint forms in complementary distribution – rather, it seems to suggest that the disjoint is the default form, with a specialized conjoint form required only in certain contexts. It also makes no reference to anything strictly syntactic in nature, but instead refers to the linear order of constituents. This generalization should arouse suspicion: morphosyntactic processes do not typically make reference to linear order, and closely-related languages to Tshivenda show much more complexity in the distribution of the disjoint prefix. In particular, in other languages certain classes of post-verbal material appear not to ‘count’ for the calculation of the verb form. The specific syntactic relationship of the verb and the post-verbal material is taken into account.

In order to arrive at a fuller understanding of the distribution of the conjoint & disjoint forms, I undertook a pilot study in the form of a grammaticality judgments survey. The experimental stimuli were designed to survey a wide range of different types of post-verbal material – that is, the post-verbal material varied both in syntactic category and in presumed syntactic position. The results of this experiment show that while the details of Polous’s generalization are in fact inadequate – the disjoint form is not in fact available everywhere, and the conjoint is ungrammatical in some cases where the verb is not last in the sentence – the underlying nature of the generalization is correct: The Tshivenda conjoint & disjoint forms are not in complementary distribution, but rather show a three-way split in acceptability.

3 SURVEY DESIGN AND RESULTS

I conducted a pilot study on the conjoint / disjoint alternation at the University of Venda in Thohoyandou, Limpopo Province, South Africa. The study consisted of a short written questionnaire asking for grammaticality ratings on a variety of sentences.

3.1 *Survey design*

The target stimuli comprised 8 conditions, varying the syntactic nature of the post-verbal material. Each condition comprised 4 sentences, where possible balanced between transitive & intransitive verbs, except the **dislocated** and ***in situ* object** conditions which had only 2. Each sentence was presented twice: once in the conjoint form, once in the disjoint. This resulted in a total of 56 test items. More detail on the specific conditions and stimuli is given below.

In addition to the 56 test items, the test included 22 ungrammatical control sentences and 22 grammatical fillers, making a total of 100 questions. In all the sentences (stimuli, fillers, & controls), the grammatical subjects were non-human animals drawn from noun classes 7 and 9; these subjects thus controlled an agreeing noun class prefix on the verb. Fillers were all grammatical sentences constructed by transposing test stimuli into the past tense or the present progressive aspect, where the conjoint / disjoint alternation does not occur; examples are given in (7). Controls were in a mix of tense/aspect conditions (including the present tense with overt disjoint prefix), but were rendered ungrammatical by mismatching the agreement prefix with the subject; examples are given in (8).

- (7) Fillers (grammatical):
- a. Tshipuka tsho zwima mulovha.
7.animal 7.SUBJ.PST drink yesterday
“The animal drank yesterday.”
 - b. Nḁou i khou fema muya.
9.elephant 9.SUBJ PROG breath air
“The elephant is breathing air.”

- (8) Controls (ungrammatical):
- a. *Tshinoni i a fhufha tshi tshi ya muri -ni.
7.bird 9.SUBJ DISJ fly 7.SUBJ SEC go tree LOC
“The bird flies into the tree.”
 - b. *Nguluvhe tshi bambela.
9.pig 7.SUBJ bathe
“The pig bathes.”

The 100 sentences were presented in a fixed, randomized order, each accompanied by a 5-point Likert scale. Subjects were instructed to circle a rating from 1 to 5, where 1 was labeled ‘mistaken or incomplete’, while 5 was ‘natural and complete’; the experimenter explained to each subject that they should rate the sentences based on whether they would expect the sentence to have been said by a good Tshivenda speaker. 12 undergraduates were recruited from the introductory linguistics course at the University of Venda. It took subjects around 10 minutes to complete the survey.

3.2 Test conditions

The 8 test conditions varied in the nature of the post-verbal material; conditions were chosen to cover a wide range of possible syntactic positions of the post-verbal element. For a complete list of stimuli, see Appendix A.

In the **final** condition, the verb was sentence final, i.e. there was no post-verbal material. This condition included both intransitive verbs and transitive ones; in the transitive case, the verb had an object marker prefix, which renders an overt object optional:

- (9) **Final condition:**
- a. Nguluvhe i (a) bambela.
9.pig 9.SUBJ (DISJ) bathe
“The pig bathes.”
 - b. Tshinoni tshi (a) i zwima.
7.bird 7.SUBJ (DISJ) 9.OBJ drink
“The bird drinks (it).”

Two conditions were constructed using overt post-verbal objects. These were divided between the *in situ* **object** block and the **dislocated** block. In the former, ordinary transitive verbs were presented with an overt object; in the latter, the object was co-referenced by an object marker on the verb.⁴ The inclusion of the **dislocated** block follows from the fact that related languages (e.g. isiZulu) show that dislocated objects ‘don’t count’ for the calculation of the disjoint prefix – that is, they are high enough in the structure that verbs followed only by a dislocated object pattern the same as verbs that are sentence-final. This gives us reason to expect that they may do the same in Tshivenda. Examples of the **object** and **dislocated** blocks are given in (10) and (11), respectively.

- (10) **Object condition:**
- Nḁou i (a) fema muya.
9.elephant 9.SUBJ (DISJ) breath 3.air
“The elephant breathes air.”

- (11) **Dislocated condition:**

⁴In many Bantu languages, including Tshivenda, objects coreferenced in this manner are not generally in their base position inside the vP (Buell 2005). For instance, it is generally possible to separate a coreferenced object from the verb with an adverb (i-b); this is not possible when the object marker is not present (i-a).

- (i) a. Tshinoni tshi a dzhia (*zwino) thanga
7.bird s.7 DSJ now 9.seed
“The bird takes (*now) a seed.”
- b. Tshinoni tshi a i dzhi zwino thanga
7.bird s.7 A 9.OBJ now 9.seed
“The bird takes it now, the seed.”

Nḁou i (a) u fema muya.
 9.elephant 9.SUBJ (DISJ) 3.OBJ breath 3.air
 “The elephant breathes it, air.”

Three conditions were constructed with varying classes of adverb following the verb, on the basis that different adverb classes are known to attach at different heights in the syntactic structure and thus may have varying effects on the acceptability of the disjoint prefix. The three classes of adverb chosen were **manner** adverbs (such as ‘well’ or ‘badly’), which typically attach very low in the structure; **locative** adverbs (‘at home’, ‘in the forest’) which typically attach at an intermediate point in the structure; and **temporal** adverbs (‘today’, ‘now’) which typically attach high in the structure. Both transitive and intransitive verbs were used; in the transitive case, the object was omitted and an object marker was placed on the verb, parallel to the **final** block. Examples are given below.

(12) **Manner** adverbs:

- a. Nḁou i (a) bambela zwavhudi.
 9.elephant 9.SUBJ (DISJ) bathe well
 “The elephant bathes well.”
- b. Tshiḁoni tshi (a) zwi ḁa nga-u-ongolowa.
 7.bird 7.SUBJ (DISJ) 8.OBJ eat slowly
 “The bird eats (them) well.”

(13) **Locative** adverbs:

- a. Tshiḁoni tshi (a) fhufha muri -ni.
 7.bird 7.SUBJ (DISJ) fly tree LOC
 “The bird flies in the trees.”
- b. Tshipuka tshi (a) i zwima daka -ni.
 7.animal 7.SUBJ (DISJ) 9.OBJ hunt forest LOC
 “The animal hunts (it) in the forest.”

(14) **Temporal** adverbs:

- a. Nguluvhe i (a) bambela zwino.
 9.pig 9.SUBJ (DISJ) bathe now
 “The pig bathes now.”
- b. Tshiḁoni tshi (a) i dzhia zwino.
 7.bird 7.SUBJ (DISJ) 9.OBJ take now
 “The bird takes (it) now.”

The **secondary predicate** condition comprised verbs followed by a clausal adjunct marked with the dependent prefix /tshi-/ (van Warmelo 1989). This construction is often the most natural way of conveying the destination of a verb of motion in Tshivenda. It is not presently clear where in the syntactic structure the secondary predicate attaches in Tshivenda, but analogy to related languages suggests that it attaches quite low. However, because the post-verbal material is an entire clause, rather than a DP or adverb, we might expect conjoint / disjoint alternation to behave differently from the other blocks regardless of the attachment-point of the adjunct clause.

(15) **Secondary predicate** condition:

- a. nḁou i gidima i tshi ya daka -ni
 9.elephant 9.SUBJ (DISJ) run 9.SUBJ DEP go forest LOC
 “The elephant runs into the forest.”
- b. Tshiḁoni tshi (a) i dzhia tshi tshi ya daka -ni.
 7.bird 7.SUBJ (DISJ) 9.OBJ take 7.SUBJ DEP go forest LOC
 “The bird takes (it) into the forest.”

Finally, one further block included verbs followed by the focus-sensitive item *shedzi*, which means roughly ‘only’. The intention was for *shedzi* to scope over the VP and to provide a test of whether focus changes the availability of the conjoint / disjoint prefixes. In fact, the results will show that the meaning of these *shedzi* sentences was difficult for the subjects to arrive at, and the sentences in this

block were generally rated poorly regardless of the form of the verb.

(16) **fhedzi** block:

- a. ??Tshinoni tshi (a) fhufha fhedzi
 7.bird 7.SUBJ (DISJ) fly only
 Intended: “The bird only flies (i.e. doesn’t also swim).”
- b. ??Nḡou i (a) zwi la fhedzi
 9.elephant 9.SUBJ (DISJ) 8.OBJ eat only
 Intended: “The elephant only eats them (i.e. doesn’t also throw them).”

There are at least two potential hypotheses about results. The first is that the disjoint patterns according to Polous’s generalization – available everywhere, and required when the verb is last in the sentence. More specifically, this hypothesis predicts that disjoint sentences will receive high grammaticality ratings in every condition, while the conjoint will give high ratings everywhere except the **final** condition. The second available hypothesis is that the disjoint patterns the same as in isiZulu, namely that the conjoint form is used when low attaching material follows the verb and the disjoint otherwise. More specifically, this predicts that the conjoint will receive high ratings in the **object, manner, & secondary** conditions and low ratings elsewhere; and that the disjoint will receive high ratings in the **final, dislocated, locative, & temporal** blocks and low ratings elsewhere. These predictions are summarized in the table below:

(17) Predicted availability of conjoint and disjoint by condition:⁵

Condition	Polous	isiZulu
final	Disjoint	Disjoint
dislocated	Either	Disjoint
object	Either	Conjoint
locative	Either	Disjoint
manner	Either	Conjoint
temporal	Either	Disjoint
secondary	Either	Conjoint
fhedzi	Either	?

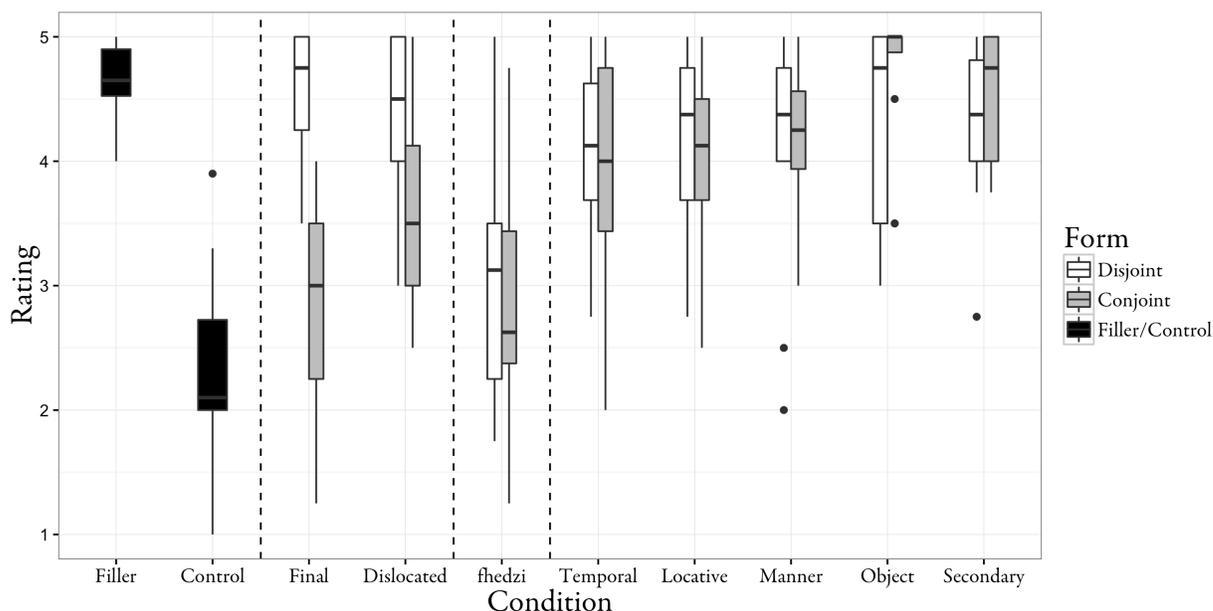
3.3 Results and analysis

Figure (18) shows the speaker mean ratings per condition, including controls and fillers.⁶ The dashed lines separate out conditions into groups with similar behavior.

⁵I’m not currently aware of how focus-sensitive operators parallel to *fhedzi* behave in isiZulu.

⁶This box-and-whisker plot should be read as follows: The dark horizontal mark indicates the median overall rating. The box extends out on either side to the edges of the 1st and 3rd quartiles, while the ‘whiskers’ extend out to 1.5 times the interquartile range; if no box or whisker is drawn, this indicates that the quartiles are at the median itself, i.e. that most responses are at the median itself. Speakers whose average response in that condition fell outside of the extent of the whiskers are regarded as outliers and plotted as individual points.

(18) Speaker mean ratings of conjoint / disjoint forms, by condition



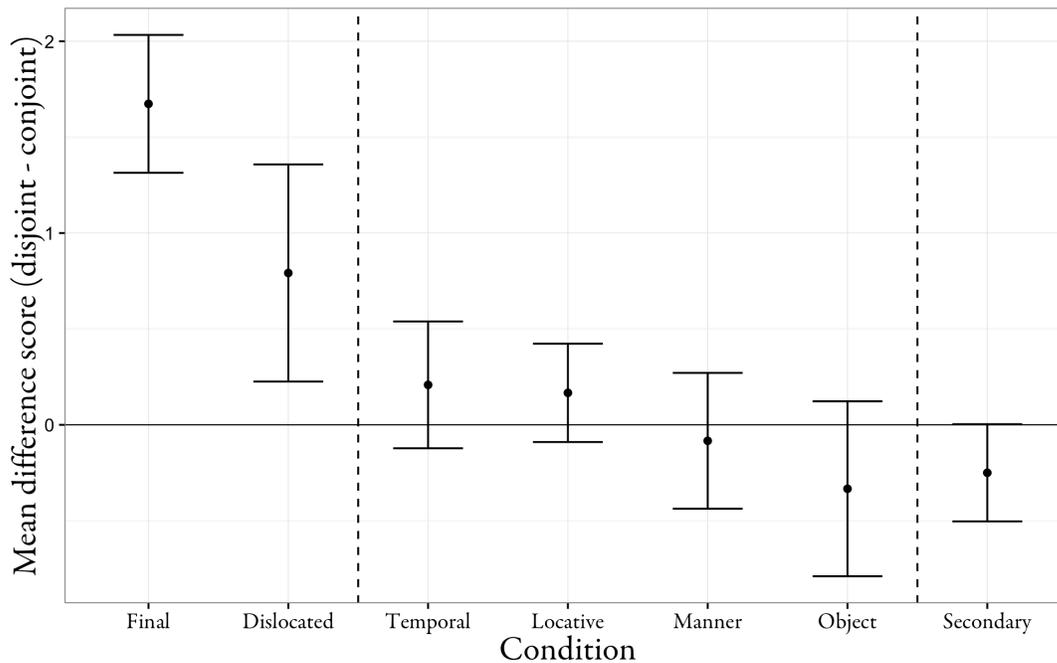
The leftmost group comprises filler and control sentences, both of which behave as expected, getting high and low ratings respectively. The next group comprises the **final** and **dislocated** conditions; Here, we can see a difference between the median rating given to disjoint sentences (in both cases, > 4) and the median rating given to the conjoint (in both cases, < 4). The **fhedzi** group is grouped on its own: Speakers rated both conjoint and disjoint forms as equally bad. I hypothesize that this is because the necessary focus reading is pragmatically challenging. I will be disregarding this condition in all that followed.

Finally, in the three adverb conditions (**temporal, locative, & manner**) and the *in situ* **object** condition, ratings are very nearly at ceiling, with the majority of sentences receiving the rating 4 or 5. (The **secondary** block seems, in these raw scores, to be behaving similarly, but see below for evidence that it patterns differently within speakers.) This indicates that speakers happily accept either the conjoint and disjoint form of the verb in these contexts, lending credence to the idea that the forms are in free variation here.

While the raw scores are insightful insofar as they indicate speakers rate both forms at ceiling in some conditions, they don't allow us to assess the significance of any preference speakers may have. As such, I calculated a by-speaker mean difference score between ratings given to the disjoint and to the conjoint sentences within each condition. In the resulting score, a positive value indicates that the speaker preferred the disjoint form of the verb, and a negative score that they preferred the conjoint. If the score is not significantly different from zero, then there is no preference. The results are presented in (19). The error bars indicate 95% confidence intervals around the across-speaker mean for the condition, while the point indicates the mean itself.⁷

⁷(19) leaves out the fillers and controls, which didn't vary by form of the verb and so a mean difference score cannot be calculated. It also leaves out the **fhedzi** condition, for reasons discussed above.

(19) Conjoint / disjoint preferences, by condition



From (19), it can be seen that the **final** and **dislocated** conditions show a significant⁸ preference for the disjoint; the adverb (**temporal, locative & manner**)⁹ conditions and **object** condition show no significant difference from zero; and only the **secondary** condition shows a significant preference for the conjoint.

Together with the fact that the **adverb** and **object** blocks received at-ceiling ratings, the lack of preference in those conditions indicates that speakers willingly accept either form there. Thus, we have the following observed availability of the conjoint and disjoint forms:

(20) Observed and predicted availability of conjoint and disjoint:

Condition	Polous	isiZulu	Observed
final	Disjoint	Disjoint	Disjoint
dislocated	Either	Disjoint	Disjoint
object	Either	Conjoint	Either
locative	Either	Disjoint	Either
manner	Either	Conjoint	Either
temporal	Either	Disjoint	Either
secondary	Either	Conjoint	Conjoint

The observed distribution thus matches neither Polous's generalization nor the isiZulu facts. Polous's generalization is false on two counts: First, the disjoint form is not in fact available everywhere – in particular, when a secondary predicate follows the verb, the disjoint is ungrammatical. Second, the conjoint is ungrammatical in some situations where the verb is not last in the sentence. However, neither does the observed distribution match what is found in isiZulu and related languages. In particular, it is true that it least in some contexts the conjoint and disjoint forms are equally acceptable.

I take this latter point as evidence that the disjoint alternation in Tshivenḁa is a different phenomenon from in other Bantu languages. In other Bantu languages, the two forms are in complementary distribution – where one is judged acceptable, the other is judged ungrammatical. This is the

⁸Significance was assessed at the 0.05 level using the Holm-Bonferroni correction for multiple comparisons.

⁹As the three categories of adverbs all behave alike, I will hereafter discuss them as a single **adverb** condition.

profile we expect of syntactic phenomena – an obligatory process occurring when its structural conditions are met. The TshivenḐa alternation does not have this signature. In the sections that follow, I will propose that the alternation is prosodically conditioned in TshivenḐa and that the optionality of the disjoint prefix corresponds precisely to optionality in the prosodic phrasing. The evidence for this claim comes from the distribution of a purely prosodic phenomenon, namely penultimate lengthening. The data presented here indicate that the conjoint / disjoint alternation shows a three-way split in acceptability, summarized in (21), across syntactic contexts; below, I will present data on penultimate lengthening that shows a three-way split in frequency across the same groupings of syntactic contexts.

(21) **Conjoint / disjoint availability by context:**

FINAL	Disjoint
DISLOCATED OBJECT	Disjoint
ADVERB	Either
IN SITU OBJECT	Either
SECONDARY PREDICATE	Conjoint

4 PENULTIMATE LENGTHENING

In this section, I will report the results of a second experiment which show that the same syntactic contexts which condition the availability of the conjoint and disjoint forms also differ systematically in their prosodic properties, specifically in the distribution of penultimate lengthening. Penultimate lengthening is a process commonly found across the Bantu family (Hyman 2009) in which the penultimate syllable of some large (roughly clause-sized) prosodic unit is phonemically lengthened. Penultimate lengthening is typically taken to be diagnostic of the intonational phrase in Bantu languages.

As with many Bantu languages, TshivenḐa has no contrastive vowel length, but does lengthen the penult of (at least) the utterance in declarative clauses (Polous 1990). This lengthening is to be regarded as a phonological effect (rather than phonetic), in that TshivenḐa only allows contour tones on lengthened penults, indicating that the process has added a tone-bearing unit (e.g. a mora) to the target syllable.

- (22) a. ndo mbindimedza **ludambwa:na**
 1sg.PST destroy 1 I.dam
 “I destroyed the dam.”
- b. ndo mbindimedza ludambwana **namu:si**
 1sg.PST destroy 1 I.dam today
 “I destroyed the dam today.”

The penult of the entire (declarative) utterance is always lengthened. However, there may be utterance-internal lengthening, as well. For example, in (23) *ludambwa:na* shows penult lengthening despite not being utterance-final.

- (23) ndo mbindimedza **ludambwa:na namu:si**
 1sg.PST destroy 1 I.dam today
 “I destroyed the dam today.”

Comparing (23) and (22-b), we see that internal lengthening in this syntactic context is apparently variable. However, there is room for uncertainty about the source of this variability: If penult lengthening is associated with the intonational phrase level of prosodic structure, then the contrast between (23) and (22-b) may indicate a contrast in intonational phrasing. Alternatively, we might propose that (23) still has an intonational phrase boundary after the object, and what is variable is not the structure but the lengthening itself. If the variability lies in the prosodic structure formation, then we might expect to find some syntactic contexts in which the prosodic structure is not variable

and internal lengthening happens 100% of the time. By contrast, if variability lies in the structure-sensitive phonological lengthening only, then even in syntactic contexts where the prosodic structure was fixed, we might expect lengthening to be variable.

Previous literature on Tshivenḁa does not resolve this issue in that it provides no good description of the distribution of utterance-internal penultimate lengthening. Cassimjee (1992) simply notes that some verbs may be ‘phrase-final’ without being utterance final; Polous is similarly vague. As such, I undertook a production study on the distribution of utterance-internal penultimate lengthening as part of a larger study on Tshivenḁa prosody. I will show below that the distribution of utterance-internal lengthening in my study shows a complicated three-way distribution that indicates variability in both the structure-sensitive phonology and the prosodic structure formation components. Furthermore, the syntactic conditions on this three-way distribution coincide directly with the conditions on the conjoint / disjoint split.

4.1 Study design

As part of a larger study on Tshivenḁa prosody, I conducted a production experiment on the distribution of penultimate lengthening. The study consisted of a production task with written stimuli: Speakers were presented with a series of sentences on a computer and asked to read the sentence out loud and to press a key to advance. All speakers were recorded using head-mounted microphone. Speakers read the entire list of sentences 3 times; each time the stimuli were presented in a different randomized order (with the order being fixed across speakers). 12 speakers were recorded; I’m here reporting on a subset of the data including 5 speakers and 1 repetition each.

The study comprised a number of different syntactic conditions, each varying what material followed the verb; in the randomized order, the conditions were interspersed so as to serve as fillers for one another. As this was part of a larger study, there were several conditions not relevant to sentence-internal penultimate lengthening. I will here report on four syntactic conditions, parallel to the conditions used in the conjoint / disjoint study: *in situ* **objects**, **dislocated** objects, **adverbs** (balanced across temporal, manner, and locative adverbials), and **secondary** predicates. All stimuli used a first-person subject. Within each condition, the stimuli manipulated other prosodic factors such as the length of the verb (1 or 2 syllables versus 4 or 5 syllables), the lexical tone of the verb (high or low), and the length & lexical tone of the post-verbal material where possible such that each condition was balanced on all these dimensions. Because of the complexity of these manipulations, no attempt was made to control the segmental material of the lexical items, or to reuse the same lexical items across different conditions.

Examples of stimuli in all four conditions are presented below; a full list of stimuli can be found in Appendix B.

(24) Object condition:

- a. Ndó lá nemeḁeme.
1 sg.PST eat 9.termite
“I ate a termite.”
- b. Ndó mbíndimedza ludambwana.
1 sg.PST destroy 1 I.dam
“I destroyed the dam.”

(25) Dislocated object condition:

- a. Ndó í lá nemeḁeme.
1 sg.PST 9.OBJ eat 9.termite
“I ate it, a termite.”
- b. Ndó lú mbíndimedza ludambwana.
1 sg.PST 1 I.OBJ destroy 1 I.dam
“I destroyed it, the dam.”

(26) Adverb condition:

- a. Ndó wa vhunamelo -ni.
1 sg.PST fall bus.stop LOC
“I fell at the bus stop.”

- b. Ndó námaila ṅamúsi.
1sg.PST stagger today
“I staggered today.”
- c. Ndó námaila nga-u-ongolowa.
1sg.PST stagger slowly
“I staggered slowly.”
- (27) **Secondary predicate condition:**
- a. Ndi réila ndi tshí ya hayá -ni.
1sg drive 1sg DEP go home LOC
“I drive home.” (Lit.: I drive such that I go home.)
- b. Ndi gidima rí tshí ḍa hayá -ni
1sg run 2sg DEP leave home LOC
“I chase us from home.” (Lit.: I run such that we leave home.)

4.2 Results

After marking all syllable boundaries in Praat, I initially hand-coded all syllables as long or short. I then ran two statistical tests to confirm the validity of my transcriptions. First, I ran a mixed-effects linear model to determine whether the two classes of syllables identified by my transcriptions really showed different distributions of length. Only the penult of the verb was considered, and these were compared across all four study conditions in addition to control conditions in which I believed the targets to be all short or all long. In order to control for both speech rate and the varying lengths of the words in question, I took as my dependent variable the observed / expected ratio defined in (28). This represents the observed length of the syllable divided by a naive ‘expected’ duration in which the observed duration of the word is divided evenly by the number of syllables.

$$(28) \text{ Syllable observed / expected ratio} = \text{Duration}_{\text{target}} * n_{\text{syllables}} / \text{Duration}_{\text{word}}$$

The mixed-effects model looked for a fixed effect of the transcription on the syllable duration ratio, with a random effect of the specific word used and of the speaker. The first random effect is intended to control for phonetic properties specific to that word, such as inherent vowel length or adjacent segments. The model run was:

$$(29) \text{ Duration ratio} \sim \text{Transcription} + (1 | \text{Word}) + (1 | \text{Speaker})$$

This model showed a significant main effect of transcription at $p < .001$, confirming that the transcription does indeed reflect a real difference in duration.

As a second confirmation, I trained a logistic classifier on the transcribed data and tested its predictions. The model was constructed as follows:

$$(30) \text{ Transcription} \sim \text{Duration ratio} + (1 | \text{Word}) + (1 | \text{Speaker})$$

I used a fold / rotation method for testing in which the data was randomly divided into four groups, with the classifier trained on three and then tested on the fourth. The average accuracy over all four runs was 88.6%, again confirming that the transcriptions reflect measurable differences in relative syllable duration.

4.3 Internal lengthening distribution

Once the accuracy of the transcriptions had been confirmed, I tabulated the percentage of tokens displaying utterance-internal penultimate lengthening on the verb within each syntactic condition:

(31) Percentage of tokens with internal penult lengthening:

(SENTENCE-FINAL) DISLOCATED OBJECT	(100%) 60%
ADVERB	25%
IN SITU OBJECT	15%
SECONDARY PREDICATE	5%

Strikingly, the distributions also show a three-way split: Utterance-internal lengthening is common when only a dislocated object follows the verb; when an *in situ* object or an adverb follows the verb, lengthening is less common; and when only a secondary predicate follows the verb, lengthening is vanishingly rare.¹⁰ Notably, the syntactic conditions on this distribution are the same as for the conjoint / disjoint alternation: That is, verbs followed by dislocated objects pattern the same as sentence-final verbs; *in situ* objects and adverbs pattern together, and secondary predicates pattern a third way. This overlap suggests a common origin for both phenomena; in the next section, I will outline a model of Tshivenda prosody that explains the commonalities.

5 EXPLAINING THE THREE-WAY SPLIT

I have shown that both the conjoint / disjoint alternation and utterance-internal penultimate lengthening show a three-way split in their distributions, and that the syntactic conditions underlying both phenomena pattern alike between the two. In this section, I will begin by arguing that the Tshivenda utterance-internal lengthening facts are challenging for a theory of prosody in which structure-sensitive phonological marking is calculated directly off the syntactic structure. In order to fully capture the distribution, we need an indirect reference theory of prosodic structure. I will propose two such models: One in which variability is present both in the syntax-prosody mapping and in the prosody-phonology mapping, and the other in which variability is present only in the syntax-prosody mapping but in which a particular prosodic structure is always mapped to a unique phonological string.

I will further argue against a model in which all prosodic variability is attributed to variation in prosodic structure due to such phonetic factors as speech rate (e.g. Pak 2008). While speech rate may play some role, such a model cannot fully account for the Tshivenda lengthening data. Finally, I'll tentatively propose that the model with two sources of variation allows for a parsimonious analysis of the conjoint / disjoint alternation in Tshivenda; specifically, I'll propose that the disjoint prefix occurs exactly when the verb is last in an intonational phrase and that the conjoint appears everywhere else.

5.1 *The inadequacy of direct reference*

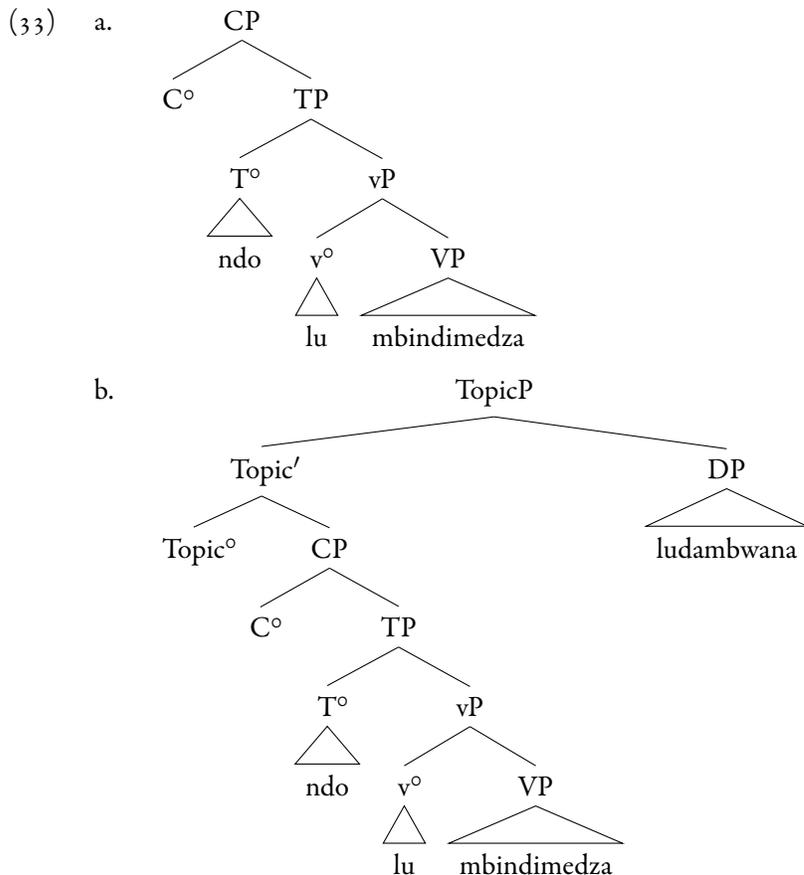
I will propose that a direct reference theory of prosody makes it difficult to state a parsimonious analysis of the Tshivenda facts. In a direct reference theory, prosodic marking is calculated directly from the underlying syntactic structure. As I have shown above, Tshivenda internal lengthening shows a complicated three-way frequency distribution across syntactic structures. This three-way distribution does not allow for a simple syntactic generalization about the frequency of lengthening. In particular, there are two cases in which the frequency of lengthening differs between two contexts that are structurally identical.

The first problematic difference concerns utterance-internal versus utterance-final lengthening. Recall that the penultimate syllable of the entire utterance is always lengthened. In particular, there is a contrast in the frequency of lengthening between object-marked transitive verbs followed by a null object versus those followed by a right-dislocated object:

¹⁰ All but one of the secondary predicate cases showing internal lengthening come from the same speaker, who shows many signs of list intonation in general.

- (32) a. Ndo lu mbindime:dza
 1sg.PST 1I.OBJ destroy
 "I destroyed it." **Lengthening obligatory: 100%**
- b. Ndo lu mbindime(:)dza ludambwa:na
 1sg.PST 1I.OBJ destroy 1I.dam
 "I destroyed it, the dam." **Lengthening frequent: 60%**

I will assume, following Buell (2005) and others, that the dislocated object in (32-b) is merged quite high in the syntactic structure, adjoining to TopicP above CP. This gives the two sentences the structures in (33):¹¹

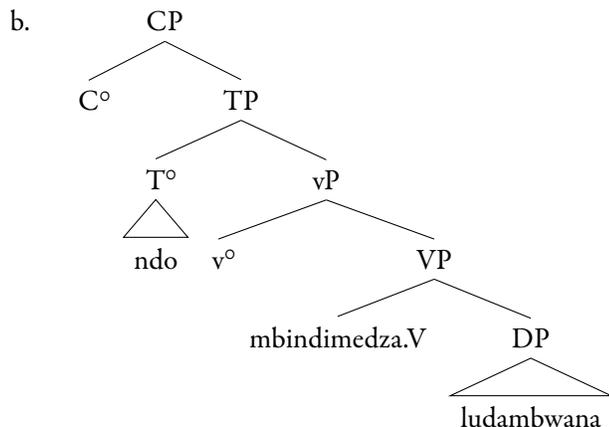


Note that the structure in (33-a) is a strict subset of the structure in (33-b). That is, the CP in (33-a) appears embedded within (33-b), with no change to its internal composition. However, in (33-a) it receives penultimate lengthening with 100% frequency, while in (33-b) it receives it with only 60% frequency. This discrepancy is problematic for a theory of prosody which would tie lengthening directly to the syntactic structure.

The second problem comes from the intermediate frequency with which transitive verbs followed by *in situ* objects undergo lengthening. That is even when the verb is not last in any phrase, not even in the VP, they still undergo lengthening, albeit with reduced frequency. This is shown in (34).

¹¹Nothing in this analysis hinges on the specific projects occupied by the subject and object prefixes on the verb.

- (34) a. Ndo mbindime(:)dza ludambwa:na
 1sg.PST destroy 1 I.dam
 “I destroyed the dam.” Lengthening infrequent: 25%



To account for this lengthening, we'd need to specify that there is some possibility of lengthening the verb regardless of its structural position – even when it is not last in any phrase, it receives lengthening some of the time. However, the data shows that this is not true: When the verb is followed by a secondary predicate, there is no possibility at all of it being lengthened. That is, the rule which triggers infrequent lengthening in (34) must be sensitive to the exact nature of the post-verbal material. Capturing this data in a direct reference system, then, quickly reduces to simply restating the facts: For each kind of post-verbal material, a different frequency of lengthening is specified.

5.2 Prosodic variability

By contrast, an indirect reference theory allows us to capture the Tshivenḁa facts quite simply. In an indirect reference theory, prosody involves two separate components: First, the prosodic structure building component consults the syntactic structure and builds a prosodic structure; then, the phonology calculates the surface form from the prosodic structure itself. We can capture the Tshivenḁa facts by assuming that there is some unit of prosodic structure associated with penultimate lengthening, and that there is variability in both the prosodic structure building component and the structure sensitive phonology component.

Since penultimate lengthening has traditionally been taken to be diagnostic of the intonational phrase (ι P) in Bantu languages (Hyman 2009), I will proceed to use this term to denote the unit of prosodic structure which is the environment for lengthening; nothing in my analysis hinges on this specific label. Thus, for a sentence in which lengthening is present on both the verb and the (dislocated) object, I propose that there are right ι P boundaries as shown in (35):

- (35) ndo lu mbindime(:)dza)_l ludambwa:na)_l
 1sg.PST 1 I.OBJ destroy 1 I.dam
 “I destroyed the dam.”

This distribution of boundaries corresponds to at least two different structures. The first is a flat, non-recursive structure, illustrated in (36-a); the second is a recursive structure as in Selkirk (2011), illustrated in (36-b).

- (36) a. (_l ndo lu mbindime(:)dza)_l (_l ludambwa:na)_l
 1sg.PST 1 I.OBJ destroy 1 I.dam
 “I destroyed the dam.”
 b. (_l (_l ndo lu mbindime(:)dza)_l ludambwa:na)_l
 1sg.PST 1 I.OBJ destroy 1 I.dam
 “I destroyed the dam.”

However, consider the fact that the lengthening on the verb is variable in this context, while the

lengthening on the final word of the sentence is categorical. We thus need some way to distinguish the two ι Ps. The non-recursive structure gives us no way to do this: both ι Ps are structurally identical. The recursive structure, however, does allow us to differentiate these: Ito & Mester (2009) have proposed that the structure-sensitive phonology can make reference to maximal phrases (which are not dominated by another phrase of the same type) and non-maximal ones (which are). That is, the prosodic structure visible to the phonology is really as in (37):

- (37) $(_{\iota\text{-Max}} (_{\iota} \text{ndo} \quad \text{lu} \quad \text{mbindime}(:)\text{dza})_{\iota} \text{ludambwa:na})_{\iota\text{-Max}}$
 1sg.PST 1 I.OBJ destroy 1 I.dam
 “I destroyed the dam.”

This gives us a simple generalization for Tshivenda penultimate lengthening:

- (38) **Penultimate lengthening rules:**
 a. Always lengthen the penultimate syllable of a maximal ι P.
 b. Variably lengthen the penultimate syllable of a non-maximal ι P.

So far we have considered the dislocated object case, in which lengthening is frequent but variable. Next, consider the secondary predicate case, in which lengthening is impossible. The secondary predicate is itself a clause, so under standard assumptions about prosodic mapping will be mapped to an intonational phrase itself. Thus, we expect a prosodic structure like the following:

- (39) $(_{\iota\text{-Max}} \text{ndi gidima} (_{\iota} \text{ndi tshi ya haya:ni})_{\iota})_{\iota\text{-Max}}$
 1sg run 1sg DEP go home.LOC
 “I run home.”

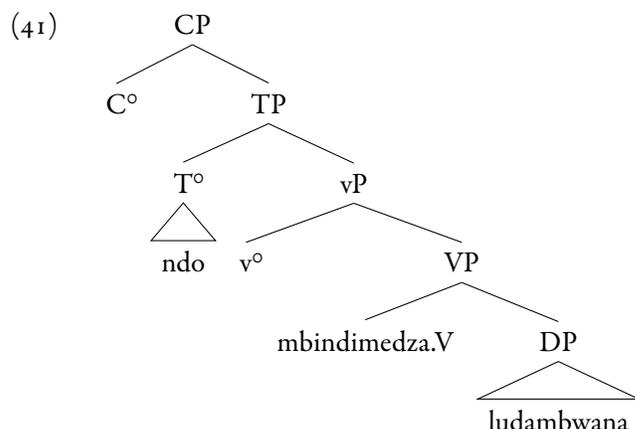
Here, the goal *hayani* is final in a maximal ι P and so is predicted to always be lengthened. The main verb *gidima*, however, isn't final in any ι P and so is predicted never to be lengthened. This matches the observed data.

Finally, consider the other two cases – adverbs and *in situ* objects. If the post-verbal material were separated from the verb by a right ι P boundary as in the dislocated object case (40-a), we would expect lengthening to be frequent; if, by contrast, the post-verbal material were not separated from the verb (40-b), we would expect lengthening to be impossible. What we observe is rather something in between: Lengthening is infrequent, but possible.

- (40) a. $(_{\iota\text{-Max}} (_{\iota} \text{ndo} \quad \text{namai}(:)\text{la})_{\iota} \text{namu:si})_{\iota\text{-Max}}$
 b. $(_{\iota\text{-Max}} \text{ndo} \quad \text{namaila} \quad \text{namu:si})_{\iota\text{-Max}}$
 1sg.PST stagger today
 “I staggered today.”

We can understand this by proposing that there is an additional source of variability: Sentences like (40) may be mapped to either of the two structures. We would thus expect the overall frequency of lengthening to be the joint probability of selecting structure (40-a) and of marking that structure with lengthening. This will be less than with dislocated objects (which always select the structure in (40-a)), but greater than with secondary predicates (which always select a structure parallel to (40-b)).

It's worth noting at this point that it is surprising to see lengthening on the verbs at all in these structures. If an ι P is the prosodic analogue of something like a clause, then we don't expect the verb to ever be final in an ι P intonational phrase in these syntactic contexts – there is no CP boundary separating the verb from an *in situ* object or adverb. This is most striking with the *in situ* object case, in which there is in fact no phrase boundary at all separating the verb and the object:



This is, in fact, a case of mismatch between the syntax and the prosody: Two items which are most certainly contained within the same clause are apparently being realized in different intonational phrases. It is not immediately clear what triggers this mismatch; I will discuss some possibilities in section 5.4. However, the presence of an apparent syntax-prosody mismatch provides additional support for an indirect reference analysis of this phenomenon.

5.3 *Where is the variability?*

One possible objection to the analysis proposed above is that it specifically requires intonational phrase boundaries to be present but unmarked. That is, non-maximal intonational phrases may be present in the structure without any overt marking in the speech signal. We might reasonably object to postulating additional structure for which we have no direct evidence within a particular utterance. As an alternative, we might propose that all intonational phrase boundaries are marked, but that Tshivenda allows optionality in whether to create sentence-internal boundaries in the first place.

The Tshivenda penultimate lengthening facts reported here do not fully rule out this analysis. Under this model, we can capture the 60% lengthening of the **dislocated object** case by proposing that the CP is mapped to an ι P with 60% probability, while the entire sentence is mapped to an ι P with 100% probability; this captures the contrast between verbs followed by dislocated objects (which are final in a CP but not the sentence) and final verbs. This same mapping principle accounts for the **secondary predicate** case – no right CP boundary separates the verb and the secondary predicate.

There is still the same difficulty in accounting for the **object** and **adverb** cases. In a model where the only source of variability is in the structure, we need to ensure that an intonational phrase boundary is inserted after the verb in these conditions at some low frequency. However we choose to model this, it will necessarily be a separate mechanism than the one by which we derive the variability in the **dislocated** case: It is sensitive to very different syntactic conditions and outputs ι P boundaries with a very different frequency. That is, in both models under discussion, there are necessarily two sources of variability: In one model, these are both in the prosodic structure; in the other, one source is in the structure, and the other in the marking.

5.4 *Speech rate*

In this section, I will consider one possible source for this variability. It has been frequently noted that fast speech rates are associated with less prosodic marking (Jun 1993, 2003; Pak 2008; Selkirk & Lee 2015; many others). Pak (2008) in particular proposes a mechanism whereby in fast speech prosodic domains may be ‘merged’; conversely, in slow speech prosodic domains may be ‘split’. Under such a model, we might explain the three-way split in utterance-internal lengthening as follows:

(42) ‘Merge/split’ model of Tshivenda lengthening:

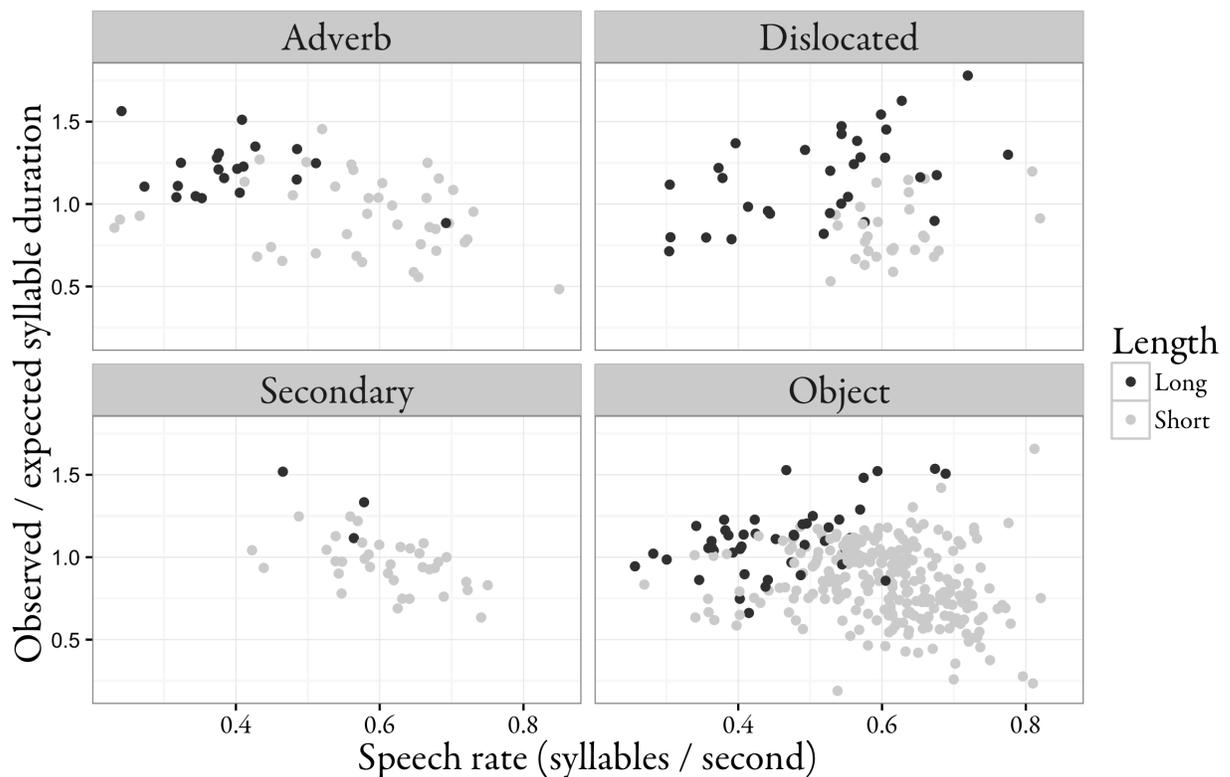
- a. Right-dislocated objects are separated from the verb by an intonational boundary that triggers lengthening; under fast speech, this boundary is removed, creating the appearance of variable lengthening in this syntactic condition.

- b. Objects and adverbs are not separated from the verb by an intonational boundary; under slow speech, an extra boundary is inserted, creating the appearance of variable lengthening in these syntactic conditions.
- c. The difference in frequency is attributable to the likelihood that a speaker will choose a fast speech rate versus a slow one.

Such an analysis is attractive in that it preserves the one-to-one correspondence between prosodic units and their marking. It also makes a clear prediction that the likelihood of lengthening should be inversely correlated with speech rate: In both the merging and the splitting cases, faster speech results in the absence of a prosodic boundary after the verb and thus no penultimate lengthening.

I tested this hypothesis on the production dataset. As a measure of speech rate, I used average syllables per second over the entire utterance. This was compared to the observed / expected syllable length ratio defined in section 4. While we expect the observed / expected ratio to fall into two distinct groups (namely long and short syllables), the correlation between this value and the speech rate will serve to measure the relative frequency of long versus short syllables at fast versus slow speech rates.

(43) Syllable duration ratio by speech rate



From (43), it can be seen that in the **adverb** and **object** conditions, there does seem to be a lack of long syllables at faster speech rates. However, this does not seem to be true for the dislocated block. I tested these observations using Pearson's r as a measure of correlation. The results, tabulated by the syntactic condition, are reported below. A negative value indicates that faster speech rates yield more short syllables. Significant results are marked with an asterisk.

(44) Correlation between speech rate and penult observed / expected length

Condition	r	p
Dislocated	0.09	0.48
Adverb	-0.44	< 0.001 *
Object	-0.27	< 0.001 *
Secondary	-0.54	< 0.001 *

The **dislocated** condition shows no overall correlation. Recall that the dislocated condition showed frequent lengthening; in a merge / split model of prosody, we would expect all of the non-lengthened cases to represent sentences where the fast speech rate caused two prosodic domains to merge. The lack of correlation indicates that this cannot be the source of variability in this condition. Thus, the merge / split model proposed in Pak (2008) cannot adequately account for the Tshivenda lengthening facts in at least this condition.

In the **secondary** condition, a negative correlation is observed; however, the tests in section 4 have already shown that there are no long syllables in this block.¹² That is, this negative correlation is reflecting an overall tendency for penults in this block to be somewhat shorter than the naive expected value (i.e. an equal division of the syllable) at fast speech rates. In fact, this correlation seems to hold of all short penults. Across all four blocks, penults transcribed as short show a correlation of $r = -0.19$ ($p < 0.001$), while long penults show the reverse correlation of $r = 0.35$ ($p < 0.001$). That is, as speech rate increases, long syllables lose duration much less than short syllables, resulting in them being longer proportionate to the word.

This same effect is likely responsible for the negative correlations in the **adverb** and **object** conditions. Recall that these conditions showed infrequent lengthening. With few long syllables overall, the trend for short syllables to grow shorter in fast speech dominates.

In the **adverb** and **object** conditions, matters are not so clear. There does seem to be a negative correlation, and inspecting the graph in ((43)), there may be some separation along the x-axis of the long and short syllables. However, recall that these conditions are also the ones that showed infrequent lengthening overall; with few long syllables, the trend for short syllables to grow shorter in fast speech may dominate. As such, the data presented here is ambiguous on the question of whether a split / merge model might account for these conditions.

If long syllables in the **adverb** and **object** conditions are in fact more likely at slower speech rates, then the results discussed in this section would be compatible with either of the models proposed above: There must be some mechanism by which the mismatch between syntax and prosody is derived in these contexts, and it must be a different source of variability than in the **dislocated** case. The results in this section show that variability of the **dislocated** case does not derive from speech rate, but that the variability in the **adverb** and **object** cases might do so.

5.5 *Other sources for variation*

If the variation in the **adverb** and **object** cases does not in fact relate to speech rate, there is at least one other factor that might play a role. It has long been noted that so-called 'information structure' factors such as focus plays a role in determining prosodic structure (see e.g. Katz & Selkirk 2011). It may be that the cases where the verb is separated from the object or adverb are exactly those cases where the object or adverb is under focus. My data did not control for information structure implicitly, and it's possible that speakers assumed different pragmatic contexts for individual utterances in this dataset.

¹²In fact, about 5% of the cases were lengthened, but as has already been discussed these cases are all attributable to lab speech from a single speaker. Even if those points are removed, the correlation reported in (44) still holds at the same significance level.

Tying this variability to information structure would solve at least two problems. For one, it would effectively eliminate one source of variability in the sense that information structure is assumed to be marked syntactically (e.g. by way of some [focus] feature); thus, in the **adverb** and **object** cases, there would in fact be a one-to-one mapping between the syntactic structure and the prosodic structure. For another, this model might explain why the **secondary** block is apparently never ‘split’ under fast speech – if the particular syntactic marking of information structure can occur on small phrases like DPs or AdvPs but not whole clauses, then we would never expect the secondary predicate to become separated from the verb in this way.

5.6 Explaining the conjoint / disjoint alternation

So far, I have proposed two possible models of the variability in Tshivenḁa prosody. In one, variability is present in two components of the grammar – both in the mapping from syntax to prosodic structure and in the mapping from prosodic structure to phonological string. In the other model, all the variability is in the mapping from syntax to prosody, with the prosodic structure uniquely determining a surface string. While I do not currently have a way to decide between these analyses, I will tentatively propose here that the first model (with variability in both components) allows us to give a parsimonious analysis of the conjoint / disjoint variability.

Recall that this analysis proposes the following prosodic structures for the conditions in question:

- (45) a. **Final:** $(_{t-\text{Max}} \text{ Subject Verb })_t$
 b. **Dislocated object:** $(_{t-\text{Max}} (_t \text{ Subject Verb })_t \text{ Object })_{t-\text{Max}}$
 c. **Adverb, *in situ* object:**
 (i) $(_{t-\text{Max}} (_t \text{ Subject Verb })_t \text{ Adverb/Object })_{t-\text{Max}}$
 $(_{t-\text{Max}} \text{ Subject Verb Adverb/Object })_{t-\text{Max}}$
 d. **Secondary predicate:** $(_{t-\text{Max}} \text{ Subject Verb } (_t \text{ Secondary predicate })_t)_{t-\text{Max}}$

In the **final** and **dislocated** condition, the verb is last in an intonational phrase – either a maximal one as in (45-a) or a non-maximal one as in (45-b). These are in fact exactly the two conditions in which the disjoint form of the verb is obligatory. By contrast, in the **secondary predicate** condition in (45-d), the verb is not last an intonational phrase; this is exactly the one condition where the conjoint is obligatory. In the **adverb** and **object** conditions, however, there is variability in the choice of prosodic structure, and the verb is sometimes last in an intonational phrase and sometimes not. These are, in fact, the two conditions in which either the conjoint or the disjoint are possible.

Thus, we can derive the following generalization:

- (46) **Conjoint / disjoint generalization (Tshivenḁa):**
 a. **Disjoint (/a-/):** appears when the verb is last in an ι P.
 b. **Conjoint (/ø-/):** appears elsewhere.

This generalization reduces the three-way split in acceptability to a simple two-way split, with the conjoint and disjoint forms in complementary distribution by prosodic structure. It also captures the generalization between the conjoint / disjoint form and the availability of sentence-internal lengthening:

CONDITION	LAST IN ι P?	LENGTHENED?	FORM?
Final, Dislocated	Always	Frequently	Disjoint
Adverb, Object	Sometimes	Sometimes	Variable
Secondary pred	Never	Rarely	Conjoint

6 CONCLUSIONS

I have shown here that the conjoint / disjoint alternation in Tshivenḁa behaves quite differently from the parallel alternation in other Southern Bantu languages. In particular, while other Southern Bantu languages typically show the disjoint and conjoint forms in complementary distribution across various syntactic structures, in Tshivenḁa there is a class of syntactic contexts in which the disjoint prefix

is optional. Furthermore, I've shown that the three-way split we see in the conjoint / disjoint alternation precisely mirrors a similar three-way split in the distribution of penultimate lengthening. I've proposed that we can understand this striking parallelism in an indirect reference theory of prosody where both the presence of the disjoint prefix and the frequency of lengthening are tied to whether the verb is final in an intonational phrase.

This analysis allows for a much more parsimonious description of the Tshivenda conjoint / disjoint facts: Instead of a three-way split conditioned by the syntax, we can state the generalization as complementary distribution conditioned by the prosody. This analysis seems particularly appropriate for Tshivenda, in comparison to the other Southern Bantu languages, in that the disjoint prefix is much more limited in distribution in Tshivenda than elsewhere: The alternation occurs only in the simple present (/ habitual) tense, and is only ever between /a-/ and /ø-/, rather than between two contentful morphemes. We might imagine, then, that the Tshivenda /a-/ prefix is really just the present tense morpheme, and that this morpheme undergoes a deletion process except when the verb is in a prosodically 'strong' position like the end of an intonational phrase. This would help us to understand why no /a-/ prefix appears when any other overt tense morphology is present. More work will be required to determine if this specific analysis is the correct one, but it seems clear that the disjoint prefix is best understood as a prosodic phenomenon.

It's worth noting that this analysis makes a strong language-internal prediction which I do not yet have the data to test: It predicts that conjoint-form verbs should never be lengthened, regardless of syntactic context, as they are by definition never tP final; conversely, it predicts that disjoint-form verbs should be lengthened ~60% of the time, as they are by definition always tP final. This prediction remains to be tested.

This analysis also has important consequences for our theory of prosody. First and most importantly, it is only compatible with a theory in which prosodic structure exists independently of its overt phonological marking. This view of prosody, in which structure-sensitive phonology happens at a later cycle than prosodic structure formation, has been latent in some recent literature, including Selkirk & Lee 2015, Elfner 2014, and others. Selkirk & Lee discuss the idea that the process of prosodic structure building strictly precedes structure sensitive phonology – that is to say, by the time prosodic marking (such as penultimate lengthening or the insertion of boundary tones) takes place, the prosodic structure is already fixed. This leaves open a question of where variability may be introduced. However, we know that many other phonological phenomena are variable (Coetzee & Pater 2011), and that some (e.g. t/d-deletion in English) are sensitive to at least morphological structure. It seems very likely, then, that structure-sensitive phonology is variable. If the structure-sensitive phonology operates on a fixed prosodic structure, as Selkirk & Lee suggest, then non-application of a variable process will result in a prosodic boundary which simply goes unmarked in that particular utterance. And indeed, something like this was observed by Elfner (2014) for pauses in English and Irish. She reports experimental results showing that pause likelihood and strength is correlated with the number of prosodic phrase boundaries stacked at a particular location, but that the pausing is variable – in fast speech, no pause may be placed at relatively weak boundaries. This closely parallels the situation described here for Tshivenda: The presence of an utterance internal tP boundary tells us the likelihood that lengthening will occur across a range of utterances with that structure, but in some utterances a boundary will be present with no associated lengthening.

A second consequence of this analysis is to lend support to indirect theories of prosodic structure. As was shown above, a direct reference analysis for the Tshivenda lengthening facts runs into the problem that no unified generalization covers the range of syntactic conditions; we're reduced to simply listing the various conditions and their respective outcomes. The conjoint / disjoint alternation is similarly unresolved in such a theory: While the analysis proposed in Halpert (2016) is sufficient to explain the cases where only one of the two forms is grammatical, it leaves out the cases where both forms are apparently acceptable. We're reduced to stipulating that the syntactic process triggering the disjoint form in Halpert's analysis is simply specified to be optional when the post-verbal material consists only of an adverb or an *in situ* object. That these two cases don't even form a natural syntactic class is telling: Such an analysis holds no explanatory power.

There is a remaining mystery however: What exactly is the source of the variation found in Tshivenda? It is unsatisfying to simply stipulate that penultimate lengthening occurs optionally, or that the infrequent lengthening cases simply pick one of two available prosodic structures. I have already

shown that the best candidate for conditioning this optionality, namely speech rate, fails: Faster speech rates do not, overall, correlate with less prosodic marking. Other possibilities remain, however. In the prosodic structure building component, we have long known that ‘information structural’ concerns such as focus can influence prosodic phrasing (see for instance Katz & Selkirk 2011). Perhaps when adverbs and *in situ* objects are separated from the verb by a prosodic boundary, this is the result of something like focus-marking on either the verb or the post-verbal element. By contrast, right-dislocated objects already have a special information-structural status (roughly topicality) and so are always phrased separately. On the other side, there are a host of phonological factors which might influence the application penultimate lengthening, such as the segmental and tonal compositions of the verb itself. Further study is needed to understand the sources of both kinds of variability.

WORKS CITED

- Buell, Leston. 2005. *Issues in Zulu verbal morphosyntax*. Doctoral dissertation. UCLA, Los Angeles, CA.
- Cassimjee, Farida. 1992. *An autosegmental analysis of Venda tonology*. Taylor & Francis.
- Cheng, Lisa Lai-Shen, and Laura J. Downing. 2009. Where’s the topic in Zulu? *The Linguistic Review* 26, no. 2-3: 207-238.
- Coetzee, Andries, and Joe Pater. 2011. “The place of variation in phonological theory.” In *The Handbook of Phonological Theory*, Goldsmith, Riggle and Yu (eds.)
- Elfner, Emily. 2014. Syntax-prosody mismatches in Irish and English verb-initial structures. Paper presented at *Exploring the Interfaces 3: Prosody and Constituent Structure*. McGill University, Montréal.
- Halpert, Claire. 2016. *Argument Licensing and Agreement*. New York, NY: Oxford University Press.
- Hyman, Larry M. 2009. “Penultimate lengthening in Bantu: Analysis and spread.” Presented at the Third Conference on Bantu Linguistics. Tervuren, Belgium.
- Ito, Junko & Armin Mester. 2012. Recursive prosodic phrasing in Japanese. *Prosody matters: Essays in honor of Elisabeth Selkirk*, pp.280-303.
- Jun, Sun-Ah. 1993. *The Phonetics and Phonology of Korean Prosody*. Doctoral dissertation, The Ohio State University, Columbus, Ohio.
- Jun, Sun-Ah. 2003. “The effect of phrase length and speech rate on prosodic phrasing.” Proceedings of the 15th International Congress of Phonetic Sciences.
- Katz, Jonah, and Lisa Selkirk. 2011. “Contrastive Focus vs. Discourse-New: Evidence from Phonetic Prominence in English”. *Language* 87.4:771-816.
- Nurse, Derek, & Gérard Philippson. 2003. “Towards a historical classification of the Bantu languages.” *The Bantu Languages*: 164-181.
- Pak, Marjorie. 2008 *The postsyntactic derivation and its phonological reflexes*. Ph.D. dissertation. University of Pennsylvania.
- Poulos, George. 1990. *A Linguistic Analysis of Venda*. Pretoria, ZA: Via Afrika.
- Selkirk, Lisa. 2011. The Syntax-Phonology Interface. In J. Goldsmith, J. Riggle, and A. Yu, eds., *The Handbook of Phonological Theory, 2nd edition*. Oxford: Blackwell, 2011.
- Selkirk, Lisa & Seunghun J. Lee. 2015. Constituency in sentence phonology: an introduction. *Phonology*, v. 32 no. 1. Cambridge University Press.
- van Warmelo, N.J. 1989. *Venda Dictionary: Tshivenda – English*. Pretoria, ZA: J.L. van Schaik.
- Westphal, E.O.J. 1962. “Venda: Tonal Structure and Intonation.” In *African Studies*, volume 21 no 2.

APPENDIX A: CONJOINT / DISJOINT ACCEPTABILITY SURVEY

Final:

1. Tshin̄oni tshi fhufha	CONJOINT	The bird flies.
2. Tshin̄oni tshi a fhufha	DISJOINT	"
3. Nguluvhe i bambela.	CONJOINT	The pig bathes.
4. Nguluvhe i a bambela.	DISJOINT	"
5. Tshipuka tshi i zwima.	CONJOINT	The animal hunts it.
6. Tshipuka tshi a i zwima.	DISJOINT	"
7. N̄dou i mmbona	CONJOINT	The elephant sees me.
8. N̄dou i a mmbona	DISJOINT	"

In situ object:

9. N̄dou i fema muya.	CONJOINT	The elephant breathes air.
10. N̄dou i a fema muya.	DISJOINT	"
11. Tshin̄oni tshi dzhia thanga.	CONJOINT	The bird takes a seed.
12. Tshin̄oni tshi a dzhia thanga.	DISJOINT	"

Dislocated object:

13. N̄dou i u fema muya.	CONJOINT	The elephant breathes it, air.
14. N̄dou i a u fema muya.	DISJOINT	"
15. Tshin̄oni tshi i dzhia thanga.	CONJOINT	The bird takes it, the seed.
16. Tshin̄oni tshi a i dzhia thanga.	DISJOINT	"

Manner adverb:

17. Tshipuka tshi sinyuwa nga ut̄avhanya.	CONJOINT	The animal gets angry quickly.
18. Tshipuka tshi a sinyuwa nga ut̄avhanya.	DISJOINT	"
19. N̄dou i bambela zwavhudi.	CONJOINT	The elephant bathes well.
20. N̄dou i a bambela zwavhudi.	DISJOINT	"
21. Tshin̄oni tshi zwī la nga u ongolowa.	CONJOINT	The bird eats it slowly.
22. Tshin̄oni tshi a zwī la nga u ongolowa.	DISJOINT	"
23. N̄dou i tshi dzhia nga maanda.	CONJOINT	The elephant takes it with effort.
24. N̄dou i a tshi dzhia nga maanda.	DISJOINT	"

Locative adverb:

25. N̄dou i bambela mulamboni.	CONJOINT	The elephant washes in the river.
26. N̄dou i a bambela mulamboni.	DISJOINT	"
27. Tshin̄oni tshi fhufha murini.	CONJOINT	The bird flies in the tree.
28. Tshin̄oni tshi a fhufha murini.	DISJOINT	"
29. Nguluvhe i zwī la bulasini.	CONJOINT	The pig eats it at the farm.
30. Nguluvhe i a zwī la bulasini.	DISJOINT	"
31. Tshipuka tshi i zwima dakani.	CONJOINT	The animal hunts it in the forest.
32. Tshipuka tshi a i zwima dakani.	DISJOINT	"

Temporal adverb:

33. Tshipuka tshi sinyuwa zwino.	CONJOINT	The animal grows angry now.
34. Tshipuka tshi a sinyuwa zwino.	DISJOINT	"
35. Nguluvhe i bambela zwino.	CONJOINT	The pig bathes now.
36. Nguluvhe i a bambela zwino.	DISJOINT	"
37. Tshin̄oni tshi i dzhia zwino.	CONJOINT	The bird takes it now.
38. Tshin̄oni tshi a i dzhia zwino.	DISJOINT	"
39. N̄dou i mmbona namusi.	CONJOINT	The elephant sees me today.
40. N̄dou i a mmbona namusi.	DISJOINT	"

Secondary predicate:

41. N̄dou i gidima i tshi ya dakani.	CONJOINT	The elephant runs into the forest.
--------------------------------------	----------	------------------------------------

42. Nḁou i a gidima i tshi ya dakani.	DISJOINT	"
43. Tshinoni tshi fhufha tshi tshi ya murini.	CONJOINT	The bird flies into the tree.
44. Tshinoni tshi a fhufha tshi tshi ya murini.	DISJOINT	"
45. Nḁou i tshi dzhia i tshi ya dakani.	CONJOINT	The elephant takes it to the forest.
46. Nḁou i a tshi dzhia i tshi ya dakani.	DISJOINT	"
47. Tshinoni tshi i dzhia tshi tshi ya dakani.	CONJOINT	The bird takes it into the tree.
48. Tshinoni tshi a i dzhia tshi tshi ya dakani.	DISJOINT	"

fhedzi:¹³

49. ? Tshinoni tshi fhufha fhedzi.	CONJOINT	? The bird only flies.
50. ? Tshinoni tshi a fhufha fhedzi.	DISJOINT	"
51. ? Nguluvhe i rembuluwa fhedzi.	CONJOINT	? The pig only turns around.
52. ? Nguluvhe i a rembuluwa fhedzi.	DISJOINT	"
53. ? Tshinoni tshi i dzhia fhedzi.	CONJOINT	? The bird only takes it.
54. ? Tshinoni tshi a i dzhia fhedzi.	DISJOINT	"
55. ? Nḁou i zwi ḁa fhedzi.	CONJOINT	? The elephant only eats it.
56. ? Nḁou i a zwi ḁa fhedzi.	DISJOINT	"

Controls:¹⁴

57. * Nguluvhe tshi bambela.		* The pig washes.
58. * Nḁou tshi u fema muya.		* The elephant breathes it, air.
59. * Nḁou tshi bambela mulamboni.		* The elephant washes in the river.
60. * Tshipuka i sinyuwa nga utavhanya.		* The animal grows angry quickly.
61. * Nguluvhe tshi rembuluwa fhedzi.		* The pig only turns around.
62. * Tshinoni i u dzhia a tshi ya dakani.		* The bird takes it into the forest.
63. * Tshinoni i zwi dzhia fhedzi.		* The bird only takes it.
64. * Tshinoni i zwi ḁa nga u ongolowa.		* The bird eats it slowly.
65. * Nḁou tshi mmbona ḁamusi.		* The elephant sees me today.
66. * Tshinoni i fhufha a tshi ya murini.		The bird flies into the tree.
67. * Tshinoni i a i dzhia thanga.		* The bird takes it, the seed.
68. * Nguluvhe tshi a bambela.		* The pig bathes.
69. * Nguluvhe tshi a bambela zwino.		* The big bathes now.
70. * Nḁou tshi bambela zwavhudi.		* The elephant bathes well.
71. * Tshinoni i a fhufha fhedzi.		* The bird only flies.
72. * Tshinoni i a fhufha tshi tshi ya murini.		* The bird flies into the forest.
73. * Nḁou tshi a tshi zwima.		* The elephant hunts.
74. * Nguluvhe tshi a zwi ḁa bulasini.		* The pig eats it in the river.
75. * Tshinoni i a i dzhia fhedzi.		* The bird only takes it.
76. * Tshinoni i a u dzhia tshi tshi ya dakani.		* The bird takes it into the forest.

Fillers:

77. Tshipuka tsho sinyuwa mulovha.	The animal got angry yesterday.
78. Nguluvhe i khou ḁa nemeḁeme.	The big is eating termites.
79. Nḁou i khou fema muya.	The elephant is breathing air.
80. Tshinoni tsho fhufha murini.	The bird flew in the tree.
81. Tshinoni tshi khou fhufha.	The bird is flying.
82. Zwino nḁou i khou rembuluwa.	Now the elephant is turning around.
83. Tshipuka tsho zwima mulovha.	The animal hunted yesterday.
84. Tshinoni tsho dzhia thanga.	The bird took the seed.
85. Nḁou i khou mmbona.	The elephant is watching me.
86. Tshipuka tsho mmbona mulovha.	The animal saw me yesterday.
87. Nguluvhe yo sinyuwa zwino.	The pig is angry now.

¹³All sentences in the *fhedzi* condition were intended to be grammatical and checked with one speaker prior to inclusion; however, in the results, all speakers rated them poorly.

¹⁴All control sentences are ungrammatical due to subject-verb agreement mismatch: When the subject is class 7, the agreement is class 9, and vice versa.

- | | |
|---|---------------------------------------|
| 88. Nguluvhe i khou fema. | The pig is breathing. |
| 89. Tshipuka tsho rembuluwa dakani. | The animal turned in the forest. |
| 90. Nḁou i khou bambela. | The elephant is bathing. |
| 91. Tshinoni tshi khou bambela. | The bird is bathing. |
| 92. Nguluvhe yo bambela mulovha. | The pig bathed yesterday. |
| 93. Nḁou yo bambela mbamulovha. | The elephant bathed 2 days ago. |
| 94. Tshinoni tsho dzhia thanga tshi tshi ya dakani. | The bird took the seed to the forest. |
| 95. Nḁou yo la nemeṇeme mbamulovha. | The elephant ate termites 2 days ago. |
| 96. Tshinoni tsho dzhia thanga murini. | The bird took a seed into the tree. |

Practice:¹⁵

- | | |
|----------------------------|-------------------------|
| 97. Ndo vhona mukalaha. | I saw the old man. |
| 98. * Mukalaha ri la nama. | * The old man eat meat. |
| 99. * Ndi khou lwa na. | * I am arguing with... |
| 100. Ndi bambela. | I bathe. |

APPENDIX B: PENULTIMATE LENGTHENING EXPERIMENT DESIGN

The production experiment on penultimate lengthening was part of a larger experiment collecting information on Tshivenda prosody. The composition of the four experimental conditions is described below; these were intermixed with sentences from other experiments which served as fillers. Within each condition, a variety of lexical and morphological conditions were manipulated such that the conditions were balanced for factors such as the length and tone of the verb.

.1 In situ object condition:

The *in situ* object condition consisted of transitive verbs followed by an object. The following factors were manipulated:

1. Verb: lexical High or Low tone
2. Verb: short (1 syllable) or long (4 syllables)
3. Object: all lexical Low tone or High-tone bearing
4. Tense: Past (/o-/) or present progressive (/khou-/)
5. **Total number of stimuli:** 2 x 2 x 2 x 2 = 16

All the objects were at least 3 syllables long. Sentences followed the schema in (47), where V and O were the combinations in the table below.

- (47) (ndo / ndi khou) V O
(1S.PST / 1S.PROG) V O

		SHORT		LONG	
		High	Low	High	Low
High		néa nwanányána 'give the girl'	gwa muḁáva 'dig the sand'	mbíndimedza hoṅndo 'destroy the oven'	vhanyuludza vhánányána 'go through girls'
	Low	lá nemeṇeme 'eat termites'	gwa ludambwana 'destroy the dam'	mbíndimedza ludambwana 'destroy the dam'	vhanyuludza ludambwana 'go through dam'

¹⁵The first four questions of the survey were practice questions; the experimenter discussed how to rate these four with the subject.

.2 *Adverb condition:*

The adverb condition consisted of intransitive verbs followed by an adverb. The following factors were manipulated:

1. Adverb type: Temporal / Locative / Manner
2. Adverb: High tone bearing or all Low tone
3. Verb: short (1 syllable) or long (4 syllables)
4. **Total number of stimuli:** $3 \times 2 \times 2 = 12$

Sentences in this block followed the schema in (48), where the verbs were draft from (49) and the adverbs from the table below.

(48) ndo V Adv
1s.PST V Adv

(49) a. námáila ‘stagger’
b. wa ‘fall’

	TEMPORAL	LOCATIVE	MANNER
High	namúsi ‘today’	háyáni ‘at home’	u távhanya ‘quickly’
Low	mulovha ‘yesterday’	vhunameloni ‘at the bus stop’	u ongolowa ‘slowly’

.3 *Right dislocated condition:*

The right dislocated condition consisted of transitive verbs bearing an object marker and an overt object. The following factors were manipulated:

1. Verb: lexical High or Low tone
2. Verb: short (1 syllable) or long (4 syllables)
3. Tense: past (/o-/) or present progressive (/khou-/)
4. **Total number of stimuli:** $2 \times 2 \times 2 = 8$

All the dislocated nouns were low-toned and at least four syllables long. All sentences in this block followed the schema in (50), where V and O were selected from the table below (with an object marker matching the class of the object):

(50) (ndo / ndi khou) OM V O
(1s.PST / 1s.PROG) OM V O

	Short	Long
High	i lá nemeṅeme ‘eat it, the termite’	lu mbíndimedza ludambwana ‘destroy it, the dam’
Low	lu gwa ludambwana ‘dig it, the dam’	lu vhanyuludza ludambwana ‘go through it, the dam’

.4 *Secondary predicate condition:*

The secondary predicate condition consisted of intransitive verbs of motion followed by a clausal adjunct indicating the destination or origin of motion. The following factors were manipulated:

1. Main verb: lexical High or Low tone
2. Secondary predicate: denoting destination (/ya/ ‘go’) or origin (/ḍa/ ‘leave’)

3. Secondary predicate: Singular subject (/ndi/) or plural subject (/ri/)
4. **Total number of stimuli:** $2 \times 2 \times 2 = 8$

The matrix subject was held constant, so that there was either a match or mismatch between the matrix and embedded subjects. Sentences in this block followed the schema in (51), where V was one of the two verbs in (52):

(51) ndi V (ndi / ri) tshi (ya / da) haya -ni
1sg V (1sg / 1pl) DEP (go / leave) home LOC

- (52) a. **reila** 'drive'
b. **gidima** 'run'