SYNTAX FROM THE BOTTOM UP: ELICITATION, CORPUS DATA, AND THICK DESCRIPTIONS

George Aaron Broadwell
University at Albany,
State University of New York

Abstract: Syntacticians working in language documentation need to approach syntax in the context of an overall documentation program, which often includes creation of a lexicon and collection of texts. In such an environment, syntactic research needs to use both a text corpus and elicitation in a careful way. A deeper understanding of several unusual syntactic properties of Copala Triqui has emerged from syntactic research conducted in the context of such a language documentation program.

1 Top down and bottom up paradigms[1]

Linguists engaged in investigating the syntax of the world's languages might roughly be divided into two camps, which I will call top down syntacticians and bottom up syntacticians. Researchers in the top down paradigm are interested in some linguistic phenomena (such as control, anaphora, or the passive voice). As a result of this interest, the researcher chooses some languages of interest for research. The researcher then works alone or with language specialists or students on these topics, and authors or co-authors books, articles, and chapters with analyses of specific grammatical phenomena. Work from the top-down paradigm has been extremely influential in syntactic theory and this paradigm is the predominant model in both generative and typological approaches to syntax.

Another approach is the one that I will call the bottom up paradigm, and this is often the way of doing syntax that is most appropriate for those engaged in language documentation. In the bottom up paradigm, the researcher establishes a working relationship with members of some language community and commits to long-term work on language documentation. The research agenda in such a situation is partly determined by community needs and interests. A typical language documentation project needs a grammar, a text corpus, and/or dictionary. In such a situation, syntactic research needs to be integrated with the overall documentation agenda. For languages where there has been prior work by SIL, such as a dictionary or a Bible translation, it is also important to consider how academic linguists can use this material intelligently.

This paper focuses on syntactic research within the context of the Copala

---

Triqui language documentation project, and identifies some insights that the bottom up approach to this language has yielded.

2 The Copala Triqui language documentation project[2]

I and my students at University at Albany have been engaged in a language documentation project on the Copala Triqui language since about 1998. Copala Triqui (TRC) is a Mixtecan language, and Mixtecan languages are part of the larger Otomanguean stock.[3] There about 30,000 speakers in Oaxaca, Mexico (and in other parts of Mexico and the United States). Copala Triqui communities have been subject to external and intercommunal violence, which has resulted in a diaspora. There are now large Triqui communities in California and in the Albany, NY area.

2.1 Texts from our documentation project

We have produced two books and an online dictionary of Copala Triqui during this period. The first book is Broadwell et al (2009). This book is a traditional Triqui story of the origin of the sun and moon as twin brothers. The brothers have various adventures, slay a monster, and rise to the sky to become the sun and moon. The second book is Vidal-López (2012). This book was written by our primary language consultant, Román Vidal-López, after he had worked with our group for several years. Nana naguan' rihaan nij síí chihaan' 'Words of counsel for the Triqui people', describes the necessary steps for peace in the Copala Triqui community.

---

2 The orthography used in this paper is based on the practical orthography developed by Barbara and Bruce Hollenbach of the Summer Institute of Linguistics. I follow their usage in the representation of the consonants, including the following conventions: <x> = [ʃ], <xr> = [ʂ] (a retroflex alveopalatal sibilant), <ch> = [tʃ], <chr> = [tʂ], <c> = [k] (before front vowels), <qu> = [k] before back vowels, [v] = [β] and <j> = [h]. <Vn> represents a nasalized vowel. Copala Triqui has five level tones (1, 2, 3, 4, 5) and three contour tones (13, 31, 32), as discussed in Hollenbach (1984). This paper uses the predominant popular orthography, in which the high tones (4, 5) are marked with acute accent, the low tones (1, 2) are marked with an underscore, and the mid tones (3, 32) are unmarked. This tone-marking is sufficient to show all the relevant morphological distinctions of the language.
Glosses use the following abbreviations: acc = accusative, com = completive aspect, decl = declarative, f = feminine, m = masculine, n = neuter, neg = negative, p = plural, part = particle, pot = potential, s = singular.

3 In some places, the name of this language is spelled *Trique*, but speakers we have consulted prefer *Triqui*, and that is the form we use here.

Our primary software tool for most of this work is Fieldworks Language Explorer (SIL International 2014), which provides integrated lexicography and text corpus analysis.

2.2 Prior work on Copala Triqui

Barbara Hollenbach of SIL has written a popular grammar in Spanish (2004), a long grammatical sketch (1992), a dissertation on the phonology and morphology (1984), published collections of folktales (1977, 1988), and many articles about the grammar. There is also a very good Copala Triqui-Spanish dictionary (Hollenbach 2005).

In addition to the folkloric texts, there is also a version of the New Testament in Copala Triqui, translated by Barbara and Bruce Hollenbach. Because electronic versions of the text are available at scriptureearth.org, we were able to import these texts into Fieldworks Language Explorer, using the method described in Broadwell (2012). This has resulted in a text corpus that currently contains about 258,000 words.

2.3 Difficulties in Copala Triqui language documentation

Despite the large amount of literature from Hollenbach, speakers we worked with found it very hard to write Triqui, partly because of the tones. As a practical matter, we decided that we would like to concentrate on making an audio dictionary of the language, where speakers could hear the words to reinforce the link between spelling and sound. Because many speakers now live in the United States, we also wanted a trilingual (Triqu-Spanish-English) dictionary. Our online, searchable dictionary of Copala Triqui is http://copalatriqui.webonary.org, and it currently contains about 2200 words, with about 3000 audio files illustrating the headwords and example sentences.

2.4 Enriching the lexicography of Copala Triqui

Although Hollenbach's dictionary is excellent, its examples are limited and there is not much information about how words are used in context. We decided that we would like to develop a text corpus and supplement the dictionary with a much 'thicker' set of information about the way Copala Triqui words are used in texts, illustrating entries with both elicited and textual examples.

Clifford Geertz (1994) has used the term 'thick description' to talk about the ways in which we ought to approach the interpretation of cultures – from the viewpoint of their rich interplay of symbols, histories, and contexts. Adapting his terminology to lexical knowledge, a lexicon that only lists part of speech and translation is a 'thin' model of speaker knowledge; speakers actually have rich knowledge of possible contexts for words, their

relationships to other words, etc. We can only get a sense of the true range of lexical syntax and semantics by observing words in natural, non-elicited contexts. This leads us to 'thick lexical descriptions'.

3 Syntactic documentation and thick lexical description

Our language consultants work long hours and have limited time to work with us (about two hours per week). We'd like to make the experience rewarding by focusing most of the time on activities that they have chosen (dictionary and text). Thus there is some time for syntactic elicitation, but we need to use it wisely. If possible, we try to use the text work to reinforce the syntactic work.

One consequence has been that much of our syntactic research has been focused on understanding the syntactic properties of different classes of verbs. For example, our corpus contains many textual examples of compound verbs like *me rá* 'want, love' (627 instances), *uun rá* 'do willingly' (343 instances), and *uun che'ę* 'begin' (130 instances). What can we learn about these verbs from our corpus, and what additional information needs to be investigated in elicitation?

Since our dictionary needs many example sentences that illustrate the use of words, we take two main approaches to finding the examples. First, we search the text corpus for instances of words in natural contexts. Second, since the texts don't always contain good examples of words, we also ask speakers to produce sentences that use words in context. Depending on the syntactic properties of these corpus examples and volunteered examples, we may try altering properties like word order or pronoun choice to see if this affects grammaticality or interpretation.

The remainder of this paper is organized in the following way. After discussing some basic properties of Copala Triqui syntax and inflection (§4), this paper discusses two overlapping parts of the system: the syntax of compound verbs (§5) and the syntax of control verbs (§6). Most control verbs in the language are compounds, so §5 focuses on the constituent structure of the verb class which includes compounds, and §6 focuses on the interclausal relationship between control verbs and their complements. The paper ends with a discussion of future challenges (§7) and a conclusion (§8).

4 Basic properties of Copala Triqui syntax

4.1 Word order and case marking

Copala Triqui is a head-initial VSO language, as seen in the following example:

1. A'níí   Mariá   chraa   rá   yoó   a.
    put   Maria   tortilla in   container   decl
    'Maria puts the tortilla in the tenate (straw container).'

Copala Triqui has differential object marking before the object. An accusative particle *man* is obligatory before pronominal objects and optional before other objects:

2. Que-ne'e   Mariá   (man)   Juán.
    com-see   Maria   (acc)   Juan.
    Maria saw Juan.

3. Que-ne'e   Mariá   man   so'.
    com-see   Maria   acc   3:s:m
    Maria saw him.

4. *Que-ne'e   Mariá   so'.
    com-see   Maria   3:s:m
    (Maria saw him.)

4.2 Pro-drop and gaps

Pro-drop is very limited, and found only in a few coordination contexts. Thus in contrast to (2), it is not grammatical to say either of the following, regardless of discourse context:

5. *A'níí   ___   chraa   rá   yoó   a.
    put   ___   tortilla in   container   decl
    '(S/he) puts the tortilla in the tenate (straw container).'

6. *A'níí   Mariá   ___   rá   yoó   a.
    put   Maria   ___   in   container   decl
    'Maria puts (it/them) in the tenate (straw container).'

Grammatical versions of both the sentences above would require pronouns:

7. A'níí   no'   chraa   rá   yoó   a.
    put   3:s:f tortilla in   container   decl
    'She puts the tortilla in the tenate (straw container).'

8. A'níí   Mariá   man yo'   rá   yoó   a.
    put   Maria   acc 3:s:n in   container   decl
    'Maria puts it in the tenate (straw container).'

While it is ungrammatical to omit arguments, gaps do appear when noun phrases and prepositional phrases have been dislocated from their positions after the verb. This can happen through Topicalization, Wh-fronting, Neg-fronting, and Relativization. Topicalization takes an argument or adjunct of the verb and puts it in a preverbal position, sometimes with a particle *ro'*.

In the following example the phrase *nu̱' rej siu̱u̱ tan'* 'all the bottoms of the ears of corn' is the subject of the verb 'stay behind', and has undergone the Topicalization rule. The ____ shows the expected position of the subject after the verb:

9. … ne̱	nu̱'	rej	siu̱u̱	tan'	ro'
ne̱	nu̱'	rej	siu̱u̱	tan'	ro'
and	all	place	bottom	ear of corn	topic

quináj ____		riaan	ya'anj	tu'vi̱i̱	xana̱	a
qui- náj		riaan	ya'anj	tu'vi̱i̱	xana̱	a
com- stay behind		to	god	thunder	woman	part

…as for the bottom parts of the ears, they went to the female thunder god.
(Thunder walked 16)

Wh-fronting also displaces noun phrases and prepositions to a clause-initial position, leaving a gap. Again, ___ shows the expected postverbal position for the subject:

10.	Me síí	c-aráán	___	chrej	rihaan	soj?
	who	com-prevent	___	road	to	2:pl
	'Who blocked your road?'					

Neg-fronting occurs with constituent negation, and is triggered by a few elements like *a̱ 'ó* 'not any' and *nuveé* 'not'. In the following example, the gap is shown by __ after the accusative *man*. The preverbal ne̱ and sentence final ma' show sentential negation, and the *a̱ 'ó* shows that *soj* 'you (pl)' is focused.

11.	'O̱ se	a̱ 'ó	soj	ne	qui-'ya̱j chi'i̱i̱	núj	man	__	ma'.
	because	not any	2:pl	neg	com-harm	1:pl	acc	___	neg
	'Because we did not harm any of you.' [2 Cor 7:2]								

4.3 Aspect inflection

4.3.1 High and Low Register Stems

In Copala Triqui there is a system of 8 tones – 5 in the high register, and 3 in the low register.[4] In the practical orthography, the low register tones are shown by the underscore on vowels. (Hollenbach 1984, 2004 has a much more extensive discussion of the tonal system.)

Each verb in Copala Triqui has two stems – a high register stem and a low register stem. These two stems play an important role in the aspect and negation system of the language.

4.3.2 Aspect prefixes

Aspect is indicated by a prefix /k(V)-/. This prefix signals completive aspect with a high register stem. The same /k(V)-/ prefix signals potential aspect when used with a low register stem.[5] The stem without a prefix is the continuative aspect.

12. a. Ne'en   Juán   man   so'.
         see   Juan   acc   3:m:s
         'Juan sees him.'
     b. Que-ne'en   Juán   man   so'.
         com-see   Juan   acc   3:m:s
         'Juan saw him.'
     c. Que-ne'en   Juán   man   so'.
         pot-see:LOW   Juan   acc   3m
         'Juan will see him.'

The majority of verbs in Copala Triqui have this paradigm, which I label the **full paradigm**. A smaller number of verbs do not take the /k(V)-/ prefix and show a single high register form for the completive/continuative, contrasting with a low register stem in the potential. I call these **reduced paradigm** verbs.

4.3.3 Negation

When a negative particle appears – *ne* (in completive and continuative

---

4 Register here refers to pitch register, and is not used in with any sociolinguistic connotation.

5 In the practical orthography, /k/ is written <c> before <a,o,u> and <qu> before <i, e>. The vowel after /k/ is not (synchronically) predictable, and thus must be listed in the entry. For monosyllabic stems with an initial vowel, the prefix is /g-/ in place of /k-/.

aspects) or *se̱* (in potential aspect) – there is a surprising effect on the register of the following verb. As Hollenbach (1976) notes, the pattern of association between aspect and register is inverted after negative particles. In the negative context, the completive appears with low register and the potential with high register. Compare the affirmative sentences below with their negative counterparts.

13. a. Ne'en   Juán   man   so'.
         see   Juan   acc   3:m:s
         'Juan sees him.'
     b. Ne   ne'en   Juán   man   so'.
         neg   see   Juan   acc   3:m
         'Juan does not see him.'
     c. Que-ne'en   Juán   man   so'.
         com-see   Juan   acc   3:m:s
         'Juan saw him.'
     d. Ne   que-ne'e̱n   Juán   man   so'.
         neg   com-see:LOW   Juan   acc   3:m
         'Juan did not see him.'
     e. Que-ne'e̱n   Juán   man   so'.
         pot-see:LOW   Juan   acc   3:m
         'Juan will see him.'
     f. Se̱   que-ne'en   Juán   man   so'.
         neg:pot   pot-see   Juan   acc   3m
         'Juan will not see him.'

   Thus low register is not a consistent marker of aspect, since verb register is determined by a combination of aspect and polarity.

5 Compound verbs

   In addition to the simple verbs discussed above, Copala Triqui also has many compound verbs which are made up of two or three separate phonological words. Our current dictionary contains about 550 verbs, and more than a quarter of these are compounds. Consider *anóo̱ ra'á* 'touch with the hand' in the following example:

14. C-ano̱ ra'á so̱' cúú yave̱ so̱'
    pot-touch hand 2s head 2s
    'Touch your head.'

A lexicographic concern for compound verbs is that only some compositional in their syntax, and they do not all have the same syntactic structure.

The evidence for this difference can be found by looking at the occurrence possibilities of adverbs in sentences with complex verbs. A number of manner adverbs such as *uxrá* 'a lot' or *sa̱'* 'well' appear either at the beginning of the clause or after the verb of the clause.

15. a. Uxrá chá Juán rnee.
       much eat Juan bean
       'Juan eats beans a lot.'
    b. Chá uxrá Juán rnee.
       eat much Juan bean
       'Juan eats beans a lot.'

Ndo 'many, much, many times' and *tia'* 'much, many times' are also manner adverbs, but they cannot appear in clause-initial position. Thus *ndo* must appear after V in a VSO clause:

16. a. Chá ndo Juán rnee.
       eat much Juan bean
       'Juan eats beans a lot.'
    b. *Ndo chá Juán rnee.
       much eat Juan bean
       ('Juan eats beans a lot.')

Compound verbs fall into three types with respect to adverbs of this type. All compound verbs begin with a word which is a verb. The second element in a compound verb may belong to several different categories, as detailed below. Let us call the first part of a compound verb V and the second part X.

Using this terminology, the three types are differentiated by the degree to which VX forms a constituent. In the **incorporation** type of compound verb, V+X forms a single syntactic word. In the **adjunction** type of compound verb, X is head-adjoined to V to form a larger V. And in the **non-constituent** type, V and X are lexically linked to each other, but they do not form a syntactic constituent.

Adverbs in postverbal position are sensitive to the type of compound. For an incorporation compound verb like *toco' vaj* 'hang', the only possible position is after the X portion of the compound:

17. Toco' (*ndo)   vaj   (ndo)   xnii se   navẹ   so'   a.
     hang   (*much)   hang (much) boy   poss hat   3:s:m decl
     'The boy hung up his hat many times.'

 In contrast, an adjunction compound verb like *ru'maan che'ẹ* 'stomp'
allows a postverbal adverb after either V or X:

18. Ru'maan (ndo)   che'ẹ (ndo)   Juán man nij xcuaa.
     put   (much)   foot   (much) Juan acc pl   ant
     'Juan stomped on the ants many times.'

 Finally, for non-constituent compound verbs like *me rá* 'want', an adverb
must appear after V, but cannot appear after X:

19. Me   (ndo)   rá   (*ndo)   no'   gaạ   ta'nii no'.
     want (much) part (much)   3:s:f pot:exist child   3:s:f
     'She wanted very much to have children.'
     Lit. 'She wanted very much for her children to exist.'

 Compound verbs raise questions of both a theoretical and practical nature.
From a theoretical perspective, what is the difference in syntactic structure
that accounts for this difference in behavior? From a practical point of view,
how much information about the constituency of compounds needs to be
included in a lexical entry?
 To address the theoretical question first, I suggest that these three types of
compounds correspond to structures like the following:
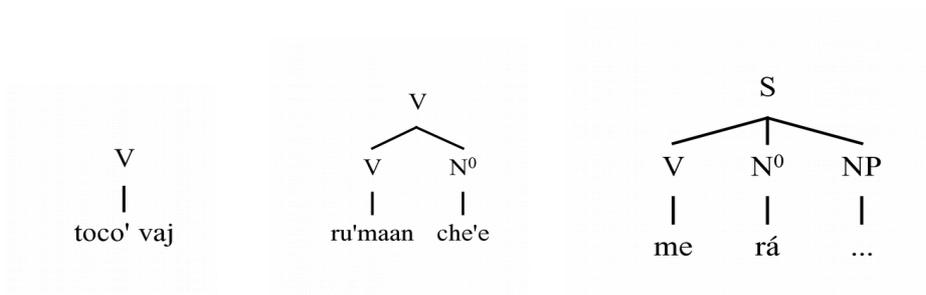
Three types of compound verb

I will assume that adjectives like *ndo* are adjoined to S, and that the word
orders observed are due to a principle of Copala Triqui grammar that places

verbs in a higher functional projection, here labelled Infl. For incorporating compounds, the entire compound appears in Infl.

```
                          IP
              ┌───────────┴───────────┐
            Infl                       S
            ↑=↓                       ↑=↓
             │              ┌──────────┴──────────┐
         Tocoj vaj         Adv                     S
           hang         ↓∈(↑ADJ)                  ↑=↓
                           │               ┌───────┴───────┐
                         ndo'o           (V)     NP         ...
                         much                 (↑SUBJ)=↓
                                                  │
                                                 xnii
                                                 boy
```
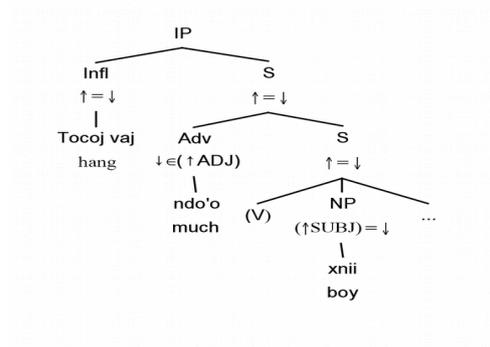
For adjunction compounds, there are two possibilities: positioning the entire [V X] in Infl or positioning just the V portion of the compound in Infl.

```
                          IP
              ┌───────────┴───────────┐
            Infl                       S
            ↑=↓                       ↑=↓
             │              ┌──────────┴──────────┐
         Ru'maan           Adv                     S
          stomp         ↓∈(↑ADJ)                  ↑=↓
                           │               ┌───────┴───────┐
                         ndo'o             V     NP         ...
                         much                 (↑SUBJ)=↓
                                    ┌────┴────┐    │
                                  (V)        N^  Juán
                                              │   Juan
                                            che'ẹ
                                             foot
```

```
                        IP
           ┌─────────────┴─────────────┐
         Infl                          S
         ↑=↓                          ↑=↓
      ┌───┴───┐              ┌─────────┴─────────┐
      V       N^            Adv                   S
      │       │            ↓∈(↑ADJ)              ↑=↓
      │       │              │            ┌───────┴───────┐
   Ru'maan  che'e̲         ndo'o        (V)      NP          ...
   stomp    foot          much              (↑SUBJ)=↓
                                                 │
                                               Juán
                                               Juan
```
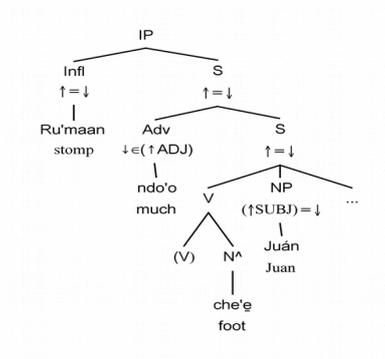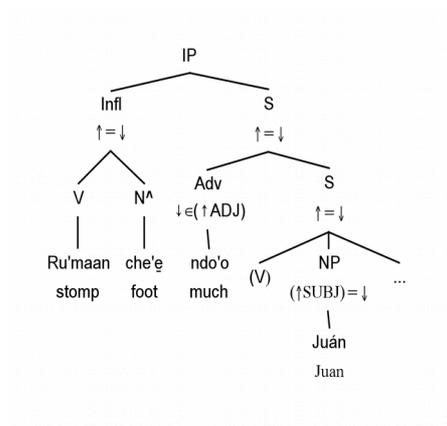
Finally, the non-constituent compounds have a structure like the following:

```
                     IP
          ┌───────────┴───────────┐
        Infl                       S
        ↑=↓                       ↑=↓
         │          ┌──────────────┴──────────────┐
         me        Adv                              S
        want     ↓∈(↑ADJ)                          ↑=↓
                    │              ┌────────┬────────┴────┐
                  ndo'o          (V)       N°      NP       ...
                  much                      │   (↑SUBJ)=↓
                                            │        │
                                           rá       no'
                                         (part)     3F
```
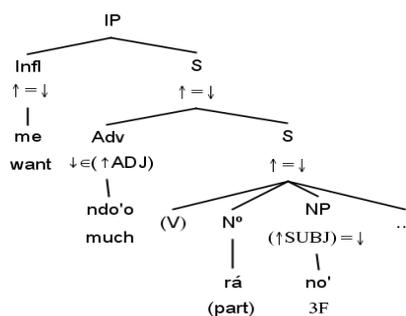
Thus the compounds have a structure that is useful to understanding the overall clausal organization of Copala Triqui, and the different types are important to understanding the syntax.

As a practical lexicographic question, how do we know whether a combination of a V plus another element counts as a compound which should be listed in the lexicon? Here the strict VSO order of Copala Triqui is very helpful – we have taken as a general principal whenever a V X NP$_{SUBJ}$ sequence in our texts, [V X] is a candidate compound verb if X is not an adverb. We select all candidates of this type in our text corpus and check them with our speakers. The corpus has revealed many such verbs that probably would not have arisen in elicitation. A harder question is determining which type a compound belongs to. Here we must turn from corpus to elicitation to check our speakers' judgments.

6 Control in Copala Triqui

6.1 Control verbs

Copala Triqui has a type of control which is very infrequent cross-linguistically, a pattern known as ***copy control*** (Polinsky and Potsdam 2006). Let us refer to the matrix clause containing a control verb as C1, and its complement as C2. Using this terminology, the normal pattern in Copala Triqui is that the subject of C2 must be a copy of the subject in C1.[6]

20.   *Nó xcúún*   *Juán ca-'na'*   *Juán*
          obliged   Juan pot-come:LOW   Juan
          'Juan needs to come/Juan must come.'

Nearly all of the control verbs in Copala Triqui are also compound verbs. In addition to the problem of distinguishing the three compound types, control verbs present an additional set of complication, since they may impose different register restrictions on their complements.

What are the general properties of control in Copala Triqui and what aspects of the syntax of these verbs are lexically specific?

6.2 Register restrictions on control complements

There are no infinitival forms of verbs in Copala Triqui, but there is a special relationship between the aspect of the verb of C1 and the verb of C2. One class (the majority) of verbs of control requires the aspect of C2 to be in the potential aspect (low register). The other class of verbs requires that the tone register of the verb of C2 match the tone register of the verb of C1 (Hollenbach 1992:218-220).

The verb *uun rá* 'act voluntarily' is in the class that requires potential aspect (low register) on the verb of C2. Consider the following example:

---

6   As an aid to understanding the structure, I have used red text to show the coreferential argument of C1 and C2. Green text shows the verbs of C1 and C2.

21.

Dan me se		gaa	guun	rá		Pedró
dan me ze		gaa	g-	uun ráa		Pedró
(new paragraph)		when	com	act:volunt		Peter

catuu	Pedró	rá	ve'	，	ne	nari'
c- atuu	Pedró	rá	ve'		ne	na- ri'
pot enter:LOW	Peter	in	house		and	rep meet,

tuvi'	Cornelió	ga	so'
tuvi'	Cornelió	ga	so'
friends, family	Cornelio	with	3:m:s

'As Peter entered the house, the family of Cornelius met him …' (Acts 10:25)

A verb like *uun che'e* 'begin' is in the second class of control verb, and requires that the register of verb of the subordinate clause match the main verb. In (22), the register of *cataj xna'anj* 'testify' must be high register because *guun che'e* is in the completive aspect and is thus high register.

22.

Gaa ne	guun	che'e		Pedró	cataj	xna'anj	nica
gaa ne	g-	uun che'e		Pedró	ca-	taj xna'anj	nica
then	com	begin		Peter	com	testify, say	straight

so'	rihaan	nij so'
so'	rihaan	nij[3] so'[3]
3mSg	to	they

'Then Peter began to testify to them…' (Acts 11:3)

But if the verb uun che'e 'begin' changes register, the register of its complement also changes, as in (23). In this example, the verbs *go', cha, co'o,* and *guun* are in the low register because *guun che'e* is low register due to its potential aspect.

23. ...gaa ne guun che'e   so'   go'   so'   man
     gaa ne   g-   uun che'e   so'   g-   o'   so'   man
     then   pot. begin:LOW   3mSg   pot hit:LOW   3mSg   acc

yo'ó   nij   mozó   do'   ,   man   nij   mozó   chana
yo'ó   nij   mozó   do'   man   nij   mozó   chana
other   plural   servant and   acc   plural   servant   woman

do'   ,   ne   cha   so'   co'o   so'   ne
do'   ne   cha   so'   co- 'o   so'   ne
and   and   eat:LOW 3mSg   pot drink:LOW   3mSg   and

guun   xno   so'   a   ·
g-   uun   xno   so'   a
pot. become:LOW   drunk   3mSg   part

'… and will begin to beat the other servants, both men and women, and to eat
and drink and get drunk.' (Luke 12:45)

Why call this register agreement rather than aspect agreement?   We can see
the difference when the verb of C1 appears with a negative:

24. Ne   guun yucuan'   rej   chej   so'
    ne   g-   uun yucuan'   rej   chej   so'
    neg   com be able:LOW   father   in-law   3mSg

c-unanj   rej   chej   so'   rique   lléé   ma'   ·
c-unanj   rej   chej   so'   rique   lleé   ma'
com-run:LOW   father   in-law   3mSg   in   clearing neg

'His father-in-law didn't have time to run from the clearing.' (Thunder god
fights 24)

In this example, although the aspect of C1 is completive, its register is low,
due to the preceding negative. Here the verb of C2 shows low register
agreement; not the normal high register of the completive. Thus many
control contexts show register agreement, not aspect agreement.[7]

_____

7 Hollenbach (1992:210) notes the same descriptive phenomenon, though her
   account does not use the idea of register agreement.

6.3 Types of control verbs

6.3.1 Inventory of verbs

Copala Triqui has verbs with subject control, object control, and oblique control by the object of a preposition. The following is a partial list of control verbs, grouped by control type and register type of the complement

Control type	Aspect/register of complement	Copala Triqui	Gloss
Subject control	potential (low register)	*me rá*	want
		nó xcúún	should
		a'vej rá	be willing, consent
		na'vej rá	refuse
		uun rá	want, do intentionally
		síj rá	dare
	matching register	*uun che'e̲*	begin
		uun nucua̲j	have the strength, be able
		'yaj canaán	succeed in
		uun yucuan'	hurry to, have time to
		navij rá	decide
Object control	potential (low register)	*a'néé*	send, order
		nago' … chrej sa̲'	encourage
Oblique control	potential (low register)	*aráán … chrej*	prevent
		vaa ya'núj	be free to

6.3.2 Control by object and oblique

Previous examples have shown control by the subject of C1. In the following examples, the subject of C2 must be identical to a.) the object of C1 or b.) an oblique (the object of a subcategorized preposition) in C1.

25.

... gaa	ne̱	ca'néé		so'	man	nij so'	ca'a̱nj		nij so'
gaa	ne̱	c-	a'néé	so'	man	nij so'	ca-	'a̱nj	nij so'
then		com	order	3mSg	acc	they	pot	go:LOW	they

chuman'	Jope	·· ·
chuman'	Jope	
town	Joppa	

'… and he sent them to Joppa.' (Acts 10:8)

ei.

emph

'…but Satan prevented us from coming to you' (Lit. '...prevented *to* us...) (1 Thes 2:18)

26.

Tza̱j	ne̱	sese	gaa ya'núj		rihaan	mozó	gu̱un	
tza̱j	ne̱	sese	gaa ya'núj		rihaan	mozó	g-	u̱un
but	if		pot:be free to		to	servant	pot	become:LOW

yaníj	so'		rihaan	ru'vee	rii ta'nga'		rihaan	so'	···
yaníj	so'		rihaan	ru'vee	rii ta'nga'		rihaan	so'	
apart	3mSg	to	boss	be in charge of	to	3mSg

'But if the servant is free from to leave the master who rules him ...' (1 Cor 7:22)

6.3.3 The form of the second subject

The coreferential subject of C2 must be a repetition of the subject of C1. The options are a.) Total repetition, b.) Pronominal repetition, or c.) a gap repetition.

The distribution of the three kinds of repetition is complex and will require more research. However, the most usual pattern is that total repetition and pronominal repetition are both acceptable for light, non-pronominal subjects. Pronominal subjects and heavy pronominal subjects allow only pronominal repetition. Gapped subjects normally allow only gap repetition.

6.3.3.1 Total repetition

Total repetition is frequent when the subject of C1 is a.) a proper name, b.) a generic, or c.) a pronoun.

In the following example, the proper noun *Diose̱* 'God' is repeated in both C1 and C2, illustrating the most frequent pattern for proper names:

27.   … guun nucua̱j   Diose̱   ti̱nanii
           g-   u̱un nucua̱j   Diose̱   ti̱nanii
           pot have strength, be able:LOW   God   pot:save:LOW

Diose̱   man   ní'   rihaan   sayuun   ado̱nj   ·

Diose̱   man   ní'   rihaan   sayuun   ado̱nj
God   acc   we (inclusive)   to   problem intens
'…God is able to save us from problems.' 1 Cor 1:18c

Generic nouns like chana̱ 'woman' and yuvi̱i 'people' are also frequently repeated in both C1 and C2

28. Ne   nó xcúún   chana̱   ca'mi̱i   chana̱   ma'   ·
     ni   nó xcúúńn   chana̱   c-   a'mii   chana̱   ma'
     neg   be obliged   woman   pot speak:LOW   woman   neg

'Women should not speak.' (1 Cor 14:33)

29. Se̱   guun yucua̱n'   nij   yuvi̱i   xca̱j
     se̱   g-   u̱un yucuan'   nij   yuvii   xcaj … cuentá
     not   com be able:LOW   plural   people   pot:learn:LOW

nij   yuvi̱i   cuentá   da̱j vaa   rá   so'   ma'   ·

nij   yuvi̱i   cuentá   da̱j vaa   rá   so'   ma'
plural people   learn   in this way   think   3mSg   neg
'People are not able to learn this, according to him.' (1 Cor 2:11)

The final type of full repetition is found when the subject of C1 is a pronoun, and in this case, pronominal repetition is the only possibility.

30. … | me rá | núj | nata' | núj | se nananana | Diose |
| --- | --- | --- | --- | --- | --- |
| me rá | núj | nata ' | núj | se nana | Diose |
| want | we (excl) | pot:explain | we (excl) | poss word | God |

rihaan	soj
rihaan	soj
to	2pl

'…we want to explain the word of God to you.' (1 Thes 2:16)

6.3.3.2 Pronominal resumption
Pronominal resumption occurs when the subject of C1 is a non-pronominal, while the subject of C2 is a coreferential pronoun.

31. | Nana | nihánj | me se | › | nó xcúún | nij |
| --- | --- | --- | --- | --- | --- |
| nana | nihánj | me ze | | nó xcúúń | nij |
| word | this | cleft marker | | have the obligation | plural |

xnii	cuno xrej	nij so'	vaa	ne	·
xnii	c- uno xrej rej	nij so'	vaa	ne	
youth	pot listen:LOW	3:m:p	part	and	

'Boys have the obligation listen to these words.' (Nana nuguan' 1:13)

An interesting subclass of pronominal resumption is found when the subject of C1 is a pronoun + appositive. There are two possibilities. The first is resumption solely by the pronoun.

32. | Ne | nó xcúún | núj | | sno'o | caráán |
| --- | --- | --- | --- | --- | --- |
| ne | nó xcúúń | núj | | sno'o | c- aráán |
| negative | be obliged | we (exclusive) | | male | pot cover:LOW |

núj		raa	núj	ma'	·
nuj[5]		raa	núj	ma'	
we (exclusive)		head of	we (exclusive)	neg	

We men should not cover our heads. (1 Cor 11:7)

But a more complex type of resumption is seen in the following example. Here the logical subject of C1 is 'you, the authorities', but the appositive portion alone appears in C1, with the pronominal portion in C2.

33. Nanạ nihánj me se ⟩ nó xcúún nij

nanạ nihánj me se nó xcúúńn nij

word this cleft marker be obliged plural

síí nicạj suun cunọ xrẹj nij soj dọ'

síí nicạj suun c- unọ xrẹj nij soj do'

authority pot listen:LOW 2:pl and

'You who hold office have the obligation to listen to these words.' (Nanạ nuguạn' 1:20)

The reverse pattern, where the pronominal is in C1 and its appositive is in C2 is unattested:

34. *Nanạ nihánj me se ⟩ nó xcúún nij soj

nanạ nihánj me se nó xcúúńn nij soj

word this cleft marker be obliged 2:pl

cunọ xrẹj nij síí nicạj suun dọ'

c- unọ xrẹj nij síí nicạj suun do'

pot listen:LOW plural authority and

(You who hold office have the obligation to listen to these words.)

It is important to notice that in examples like (34) above, the subject of C1, 'authorities', is [PERSON 1], while the subject of C2, 'you (pl)' is [PERSON 2]. This feature incompatibility seems to indicate that an approach to control via subsumption, as in Sells (2006), cannot work for Copala Triqui. The features of C2 cannot be a superset of those for C1, due to the potential incompatibility of the PERSON feature.

6.3.3.3 Subject gaps

If the subject of C1 is a gap, then the subject of C2 must also be a gap. For instance, in example (35), the relative subject *nij síí* 'the ones who' precedes the control verb *na'vej rá* 'refuse' and is associated with a gap after the verb of control (C1). Accordingly, the subject of C2 is also a gap:

35. Dạnj   catạj   soj   rihaan   nij   síí   na'vej rá __
    dạnj   ca- tạj   soj   rihaan   nij   síí   na'vej rá
    thus   pot say   2pl   to   pl   one(s) who   refuse

cunọ   __ nanạ   ca'mii   soj   a   ·
c-   unọ   nanạ   c- a'mii   soj   a
com hear   word   com speak   2pl   decl

'Thus you will say to the ones who refuse to hear the words that you speak.'
(Luke 10:11)

Similarly, in (37), the topicalized subject *roj* 'you (pl)' precedes the control
verb *nó xcúún* 'should' and leaves a gap. A gap is also found after the subject
of *nari'* 'meet'.[8]

36. Sanẹ   ạ   ›   asino ñạạn   soj   ro'   ›   nó xcúún __   doj
    sanẹ   ạ   asino ñạạn   soj   ro'   nó xcúún   doj
    but   now   first   2pl   top   should   more

nari'   __ sạ'   ñanj   gạ   xna'ánj xlá   ...
na- ri'   sạ'   ñanj   gạ   xna'ánj xlá
rep meet   well paper and Spanish language
'But first you should learn your subjects and Spanish well...' (*Nanạ nuguạn)*

 Consider example (38) where the subject *yo'ó tinúú ní'* 'other brothers of
ours' appears twice. When it is in-situ (post-verbal) in the first sentence, a
pronominal resumptive appears. When the same noun phrase is clefted in the
second and third sentences, a gap appears after the verb in both C1 and C2.

_____

8   *Nari'...ñanj,* literally 'meet paper' is idiomatic for 'learn'.

37. …ne̱ uun nucua̱j yo'ó tinúú ní' tinahuun
and be able other brother (of a man) we (incl) heal

so' man síí ran' … Yo'ó tinúú
3:mːs acc one(s) who suffer other brother (of a man)

ní' me síí uun nucua̱j __ 'yaj suun __ sa̱' noco̱o ›
we (inclusive) who be able work good big

ne̱ yo'ó tinúú ní' uun nucua̱j __ nata' __
and other brother (of a man) we (incl) be able explain

nana̱ …
language

'And another of our brothers is able to heal those who suffer.… Another of our brothers is able to work miracles and another of our brothers can translate languages.' (1Cor 12:9-10)

Substitution of a pronoun for the second gap in sentences of this sort is judged possible (39), though perhaps somewhat less natural. This contrasts with the ungrammatical pattern where a pronoun substitutes for the first gap (40):[9]

38. ? Da̱nj cata̱j soj rihaan nij síí na'vej rá __
da̱nj ca- ta̱j soj rihaan nij síí na'vej rá
thus pot say 2pl to pl one(s) who refuse

cuno̱ nij so' nana̱ ca'mii soj a ·
c- uno̱ nij so' nana̱ c- a'mii soj a
com hear 3:mːpl word com speak 2pl decl

'Thus you will say to the ones who refuse to hear the words that you speak.' (Luke 10:11)

_____

9   We might account for the ungrammaticality of (36) as a "condition D" effect (Lasnik 1986), where a pronoun may not c-command a coreferent r-expression.

39. *D̲a̲nj catạj   soj   rihaan nij síí   na'vej rá nij so'

    da̲nj   ca- tạj soj   rihaan nij síí   na'vej rá nij so'
    thus   pot say 2pl   to   pl   one(s) who   refuse   3:m:pl

cuno̲   __ nana̲   ca'mii   soj   a   ·

c-   uno̲   nana̲   c-   a'mii   soj   a
com hear   word   com speak   2:pl decl

'Thus you will say to the ones who refuse to hear the words that you speak.'
(Luke 10:11)

7. Challenges and future research questions

Though we understand much of the syntax of compound verbs and control in
Copala Triqui, there are still some problematic areas. For compound verbs, it
is difficult to document all the possibilities and to predict the type of a
compound. For the overlapping domain of control verbs, there are also some
challenges. These include a.) the possibility of covert copies, b.) the length of
the copy, and c.) the relation of normal control in Copala Triqui to the
emotion auxiliary construction.

7.1 Covert copies

In our texts, the subject of C2 is almost always an overt copy. Of several
thousand textual examples, we have found less than ten where an in-situ
subject in C1 lacks a copy in C2. Such examples have a syntax similar to
English or Spanish, where the subject of C2 is covert:

40. Me rá xrej   ca-'a̲nj Ø.

    want priest pot-go

'The priest wants to go.'

Despite their textual rarity, our language consultants readily accept
constructed examples of this sort, though they do not usually volunteer them.
We suspect that the acceptability of these sentences may reflect language
change in progress. Our language consultants are about 50 years younger
than the consultants for the New Testament (who worked with the
Hollenbachs in the 1960s). Our speakers are also more strongly bilingual or
trilingual in Spanish and English than earlier generations of Copala Triquis.

7.2 Length restrictions

Our textual examples also show an unusual restriction on the length of the subject of C2 which is difficult for most syntactic theories to capture. The subject of C1 may be of any length, but only short subjects are accompanied by a full copy. In our texts, a full copy is almost always 1-2 words. We have no examples of a copy that is longer than three words.

Copala Triqui pronouns are one or two word long, with an optional preceding plural marker or quantifier (e.g. all, each, other). Thus pronominal echos are also 1-3 words long. Similarly, in elicitation our speakers usually reject constructed examples with an copy longer than 2 words:

41. *?Me rá chana̧ chrȩ' nihánj ca-'a̧nj chana̧ chrȩ' nihánj.

     want woman short that   pot-go woman short that

(Intended reading: That short woman wants to go.)

7.3 Control and the emotion auxiliary construction

Copy control in Copala Triqui shares many traits with the emotion auxiliary construction (Broadwell 2013). In this construction, subjects of emotion predicates (C1) optionally control a second verb *ne'e* 'see' or *ni'ya̧j* 'look' (C2):

42. Nachri' nii   ni'ya̧j nii   man núj.

    hate   indef look   indef acc   we (excl)

    'People hate us.' (≈ 'They hate to regard us')

All the constraints on the register of C2 and on the shape of the copy in C2 are same as in copy control. In particular, the register of *ne'e* 'see' or *ni'ya̧j* 'look' must match the register of the verb in C1. However, the semantic contribution of the second verb in emotion auxiliary constructions is problematic, and the emotion predicates in C1 do not seem to be standard control verbs. Control and emotion constructions thus need an analysis that captures their shared morphosyntactic properties while distinguishing their syntax.

7.4 Copala Triqui control within a typology of control

The kind of control seen in Copala Triqui is probably best analysed as obligatory anaphoric control (Bresnan 1982, Dalrymple 2001, inter alia). However, previous descriptions typically use equations like

43.   (↑ COMP SUBJ PRED) = 'PRO'

((↑ COMP SUBJ)σ ANTECEDENT) = (↑ SUBJ)σ

(Dalrymple 2001:327)

This clearly will not work for Copala Triqui, since the subject of C2 need not be a gap. However it is not a trivial matter to revise these equations to model the Triqui data. In particular, it is difficult to capture the generalizations that

A.) The PRED of the COMP SUBJ can be either PRO or the PRED of the controller

B.) In appositives, the COMP SUBJ may be quantified or contain PERSON features distinct from its controller.

C.) The COMP SUBJ obeys a length restriction that limits its size.

8. Conclusion

Though there are still many unanswered questions, we have begun to get closer to a 'thick' description of Triqui compound and control verbs, and the range of syntactic contexts associated with them. To the extent that we have made good progress, it has been through a mix of elicitation and corpus work. We would not have figured out the full range of data solely from the corpus.

On the other hand, we probably would not have discovered the full range of facts from elicitation; they only came to light in a large corpus of the language. The combined power of the two methods of work can yield results that are more powerful than either method alone.

We have also benefitted from sharing the research direction with the community. Dictionary-making and text collection were the parts of linguistic research that members of the local Copala Triqui community valued as a way to preserve their language for future generations. We might not have focused as much energy on these parts of our language documentation project without the guidance of our community language partners.

References

Bresnan, Joan. (1982). Control and complementation. In *The mental representation of grammatical relations* (J. Bresnan, ed) Cambridge: MIT Press.

Broadwell, George A; Matsukawa, Kosuke; Martín del Campo, Edgar; Scipione, Ruth; and Susan Perdomo (eds). (2009). *The origin of the sun and moon: A Copala Triqui legend* (Román Vidal López, narrator). Munich: Lincom Europa.

Broadwell, George A. 2012. Un método para añadir textos religiosos rápidamente a una base de datos lingüística: un estudio del Triqui de Copala. Presented at Coloquio sobre lenguas otomangues y vecinos, Oaxaca, Mexico 2012.

Broadwell, George A. 2013. An emphatic auxiliary construction for emotions in Copala Triqui. *Proceedings of LFG13,* (M. Butt and T. Holloway King, eds). *Stanford: CSLI Publications*.

Dalrymple, Mary (2001). Lexical Functional Grammar. San Diego: Academic Press.

Geertz, Clifford. (1994). Thick description: Toward an interpretive theory of culture. Readings in the philosophy of social science, 213-231.

Hollenbach, Barbara E. 1976. Tense—negation interplay in Copala Trique. International Journal of American Linguistics 42:126–32.

Hollenbach, Barbara. 1984. The phonology and morphology of tone and laryngeals in Copala Trique. Ph.D. thesis, University of Arizona.

Hollenbach, Barbara E.. 1992. A syntactic sketch of Copala Trique. Pages 173–431 in C. Henry Bradley and Barbara E. Hollenbach, editors. Studies in the syntax of Mixtecan languages, Volume 4. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

Hollenbach, Barbara. 2004. Gramática popular del triqui de Copala. ms. (Posted at http://www.sil.org/~hollenbachb/Posted.htm)

Hollenbach, Barbara. 2005 Vocabulario preliminar de triqui de San Juan Copala, Oaxaca. ms. (Posted at http://www.sil.org/~hollenbachb/Posted.htm)

Lasnik, Howard. 1986. On the necessity of binding conditions. *Essays on anaphora,* H Lasnik (ed.). Dordrecht: Kluwer

Polinsky, Maria, & Potsdam, Eric. 2006. Expanding the scope of control and raising. *Syntax*, *9* (2), 171-192.

Sells, Peter. 2006. Using subsumption rather than equality in functional control.*Proceedings of LFG06,* (M. Butt and T. Holloway King, eds). *Stanford: CSLI Publications*.

SIL International. 2014. Fieldworks Language Explorer (FLEx). SIL Fieldworks. http://fieldworks.sil.org/flex/.

Vidal-Lopez, Román. (2012). Nana naguan' rihaan nij síí chihaan' | Consejos para la gente Triqui | Words of counsel for the Triqui people. Edited by George A Broadwell, Ashley LaBoda, Sharone Horowit-Hendler, and Gabriela Aquino Dehesa. IMS Occasional Publication No. 16. Albany NY: University at Albany.