Yucatec Maya in SBCG: A Fragment

Justin Bai, Maksymilian Dabkowski,
Kalinda Pride, Nicholas Tomlin

Yucatec Maya (YM) is a Mesoamerican language spoken by about a million speakers primarily in the Yucatan Peninsula. In this paper, we present parts of a grammatical fragment with particular emphasis on the pronominal system, verbal morphosyntax, focus, deictic clitics and topicalization through the lens of Sign-Based Construction Grammar (Boas & Sag 2012), which allows us to account elegantly for disparate constructions which share many characteristics.

1   Predicativity and agreement

In our fragment, we adhere to the conventional analyses of Yucatecan word order and treat it as an underlyingly verb (or perhaps better: predicate) initial language. As for copular predicates then, the word order of the following sentences will be taken as most basic.[1]

	SINGULAR	PLURAL
1ST PERSON	**in** (**w**)*stem*	**k** *stem* **in** (**w**)*stem*-**o'on**
2ND PERSON	**a** (**w**)*stem*	**a** (**w**)*stem*-**e'ex**
3RD PERSON	**u** (**y**)*stem*	**u** (**y**)*stem*-**o'ob**

Figure 1: Set-A morphology

(1) Koolnáal-en   (tèen).
     farmer-B.1.SG I

     'I am a farmer.'

(2) Polok-ech (tèech).
     fat-B.1.SG you

     'You are fat.'

	SINGULAR	PLURAL
1ST PERSON	*stem*-**en**	*stem*-**o'on**
2ND PERSON	*stem*-**ech**	*stem*-**e'ex**
3RD PERSON	*stem*-**Ø**	*stem*-**o'ob**

Figure 2: Set-B morphology

There are two categories of agreement morphemes called Set-A and Set-B, terminology employed in the traditional Mayanist literature. The atypical nomenclature is used as the two categories do not quite completely overlap with other, typologically more common categories. Set-A, for example, perhaps best approximated as ergative-genitive, cross-references agents on transitive verbs, and subjects on intransitive incompletive verbs, but also functions as a possessor on nouns. Set-B, best approximated as absolutive, cross-references patients on transitive verbs and subjects on intransitive ones, and also functions as plural marking on nouns. Set-A and Set-B morphemes can attach to both nouns and verbs, but their behavior is unaffected by the lexical class of their host. That motivates a unified treatment, which we provide. We recast Set-B morphology via one morphological function and Set-A morphology using the same function accompanied by 3 additional lexemes.

All basic non-copular sentences in Yucatec Maya–that is, sentences containing active verbs–start with one of fifteen Aspect-Mood (AM) markers. Here, we adapt basic analytical insights from Bohnemeyer (1998) and Bohnemeyer (2002) and postulate that AM markers are in fact the predicates of those sentences. Example (3) would then be more literally translated as 'my reading of the newspaper is/was/will be near,' as *ta'itak* is the predicate. This allows for a simpler analysis and preserves coherence of VOS word order throughout the language.

---

[1]Glossing abbreviations: 1: first person, 2: second person, 3: third person, A: set A, APP: applicative voice, B: set B, CPL: completive status, D2: a deictic clitic, D3: a deictic clitic, DET: determiner, IMP: imperfective AM marker, INC: incompletive status, NEG: negation, OBL: obligative AM marker, ONGL: onglide, PL: plural, PROG: progressive AM marker, PROX: proximal AM marker, PRV: perfective AM marker, REC: recent past AM marker, SBJ: subjunctive status, SG: singular, TERM: terminative AM marker, TOP: topical.

(3) Ta'itak in   xok-ik-Ø   le   periyòodiko-o'.
     PROX   A.1.SG read-INC-B.3.SG DET newspaper-D2

'I /have/had/will have/ almost read the newspaper,'

'I /am/was/will be/ about to read the newspaper.' (taken from Bohnemeyer 2002)

2   Focus Constructions

YM sentences often contain a syntactically focused constituent that appears to the left of the other material in the sentence. The focused constituent is gapped in the remainder of the sentence. For example, in (4), *Maruch* is focused, creating a sentence identical in truth-conditional meaning to (5), but very different in emphasis. Focus constructions are also possible in copular sentences, where the predicate is an adjective or a predicative noun instead of an AM marker. For example, in (6), the argument of the predicative noun *Maruchech*, *tèech*, is focused. The difference between the two is, again, purely pragmatic.

(4) Maruch sáam u   yil-Ø-Ø   Jwàan.
     Maria   REC A.3.SG see-SUBJ-B.3.SG Juan

'Juan saw MARIA,' 'It's Maria that Juan saw.'

(5) Sáam u   yil-Ø-Ø   Maruch Jwàan.
     REC   A.3.SG see-SUBJ-B.3.SG Maria   Juan

'Juan saw Maria.'

(6) Tèech Maruch-ech.
     you   Maria-B.2.SG

'YOU are Maria.'

(7) Maruch-ech   (tèech).
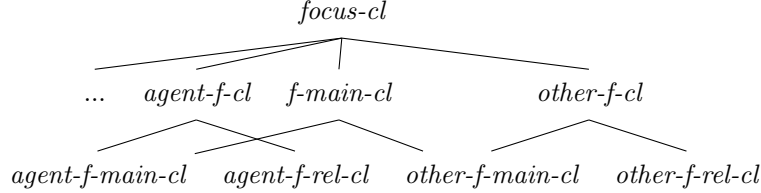   Maria-B.2.SG you

'You are Maria.'

Bohnemeyer (2002) argues that the focused constituent can also be treated as a sort of predicate. Under that analysis, the second translation given for (4) reflects its meaning more literally, although sentences like (4) should not be confused with syntactically distinct Yucatecan clefting. We account for this by identifying the left daughter with the head daughter. The right daughter (everything following the focused element) may be any type of predicate, so long as it has an appropriate gap. The mother is also stipulated to be predicative, despite the fact that the head daughter is not.

What is the proof for the predicativeness of Yucatecan focus constructions, though? There have been several tests proposed in the past, many of which are not unproblematic. Bohnemeyer (2002)'s *wáaj* test, for example, fails to recognize that *wáaj* can appear in more environments than first meets the eye. Nonetheless, we identify a new test: just like regular predicates, focused constituents are ungrammatical with definite morphology. We then suggest that the definite article *le* selects for [PRED −] items and thus explain the ungrammaticality of definite focus.

Relative clauses in YM are nearly identical to the focus constructions above. The only difference is that instead of standing alone as sentences, they have the same distribution as NPs. Importantly, this means they may be determined by *le*:

(8) wakax sáam u   yil-Ø-Ø   Maruch
     cow   REC A.3.SG see-SUBJ-B.3.SG Maria

'a cow that Maria saw'

(9) le   wakax sáam u   yil-Ø-Ø   Maruch-o'
     DET cow   REC A.3.SG see-SUBJ-B.3.SG Maria-D2

'the/that cow that Maria saw'

Due to the similarities between focus constructions and relative clauses, we define a multiple inheritance hierarchy where these two constructions differ only in the PRED values of the mother.

2

```
                          focus-cl
          ┌──────────┬───────┴──────────────┐
        ...   agent-f-cl   f-main-cl        other-f-cl
               ┌──────────┬──────────┐      ┌──────────┐
      agent-f-main-cl  agent-f-rel-cl  other-f-main-cl  other-f-rel-cl
```

We have extended this account to include coverage of agent focus, a distinct but related construction. Notably, our treatment of focus also captures the syntax of wh- questions. In YM, these kinds of questions are distinguished from declarative statements only by the presence of a focused element which is pragmatically uninformative, such as a focused wh-word.

3 Deictic Clitics

Deictic clitics is a term used traditionally in Mayanist literature to refer to four clause-final morphemes, whose occurrence must always be licensed by a clitic-triggering constituent. Their prototypical purpose is to orient the speaker deictically with regard to a constituent present in the clause. There is at most one clitic per clause, and it always appears in a clause final position.

The four clitics *a'*, *o'*, *e'*, and *i'* are glossed as D1, D2, D3, and D4, respectively. The first two clitics relate to the speech situation and are triggered by the definite article *le*. *A'* occurs if a referent is near the speaker; *o'* occurs otherwise. (They can be likened to English "this" and "that.") The semantics of *e'* and *i'* are somewhat harder to pin down. For example, *e'* is triggered by *way* "here," and serves the purpose of a topical marker. Negation and locatives frequently require *i'*. As already mentioned, deictic clitics appear not at the end of the phrase that triggers them, but rather at the end of the clause. In (10), we see the clitic appearing on *nool* even though it has been triggered by the previous constituent.

(10) Yàan u   bon-ik-Ø   le   naj   in   nool-o'.
      OBL A.3.SG paint-INC-B.3.SG DET house A.1.SG grandfather-D2

      'My grandfather has to paint the (that) house.'

When there are multiple clitic-triggering elements in a clause, the most "highly-ranked" one wins. Speech-situation clitics are the most highly ranked, followed by *e'*, and right after *e'* comes *i'*. Trivially, a clitic always outranks the absence of a clitic. The hierarchy can therefore be stated as follows:

$$a', o' > e' > i' > \varnothing$$

In (11), for example, the clitic *i'* (triggered by *ma'*) is outranked by *e'*.

(11) Le   paax-e'   ma' t-u   yu'ub-a'al-Ø   way-e'
      DET music-D3/TOP NEG IMPF-A.3.SG feel-PASS.INC here-D3

      The music is not heard over here. [7]

All this data is straightforwardly accounted for via a non-maximal construction, from which most maximal constructions inherit.

$$
\textit{deictic-q-cxt} \Rightarrow
\begin{bmatrix}
\text{MTR} & \begin{bmatrix} \text{ENQ-D} & \text{F}_{\max}(\boxed{1}, \boxed{2}, \ldots \boxed{n-1}, \boxed{n}) \\ \text{DEQ-D} & \boxed{m} \end{bmatrix} \\[3ex]
\text{DTRS} & \left\langle \begin{bmatrix} \text{ENQ-D} & \boxed{1} \\ \text{DEQ-D} & none \end{bmatrix}, \ldots \begin{bmatrix} \text{ENQ-D} & \boxed{n-1} \\ \text{DEQ-D} & none \end{bmatrix}, \begin{bmatrix} \text{ENQ-D} & \boxed{n} \\ \text{DEQ-D} & \boxed{m} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

The above construction formalizes all the relevant insights of the current section. It introduces two new features: DEQUEUE-DEICTIC (or DEQ-D) and ENQUEUE-DEICTIC (or ENQ-D). ENQ-D contains the information of which clitic is triggered by each word or, in other words, which clitic is "enqueued" by a given word. DEQ-D

contains information about which clitic is actually attached to a given word, or which word "dequeues" it. The ENQ-D of the mother is the highest-ranked clitic of all those which are triggered by the daughters. The hierarchy of clitics is formalized with the F_{max} function.

Of course, all that is to little avail unless we also specify what the top node looks like. All this machinery is set up to ensure the ENQ-D and the DEQ-D of the top node are identical. Intuitively, all it says is that the highest-ranked clitic which some constituent requires is also the one which the sentence eventually realizes. Thus, we arrive at the sentential node presented on the right.

$$S = \begin{bmatrix} SYN & \begin{bmatrix} CAT & \begin{bmatrix} PRED & + \end{bmatrix} \\ VAL & \langle\rangle \\ GAP & \langle\rangle \\ ENQ\text{-}D & \boxed{1} \\ DEQ\text{-}D & \boxed{1} \end{bmatrix} \end{bmatrix}$$

4 Topicalization

YM sentences are frequently prefaced by one or more topic phrases,
which introduce background information and behave idiosyncratically with respect to clitics. Topics may precede any grammatically independent utterance of YM, including AM phrases, sentences consisting of a focus and AM phrase, and predicates. The topicalized constituent is typically gapped from the main clause. It necessarily occurs with the enclitic *e'*, or with another higher-ranked deictic clitic if one is appropriate. Unlike with focus, there is no requirement for the presence or absence of the determiner *le* in the topic phrase. Topics always appear to the left of a focus if there is one, which is captured through the MRKG feature. Specifically, focus clauses prevent their daughters from being marked *topical*, while the topic construction ensures a topical marking on the mother node. Since relative clauses are focus clauses, this also correctly predicts that topics cannot occur within relative clauses.

(12) Ten-e' in   k'àaba'-e' Maks-Ø.
      I-D3   A.1.SG name-D3 Maks-B.3.SG

   'As for me, as for my name, it is Maks.' (or, a little less awkwardly: 'My name is Maks.')

(13) Le   ts'akyah-o' u   k'àaba'-e' Pedro-Ø.
      DET doctor-D2 A.3.SG name-D3 Pedro-B.3.SG

   'As for that doctor, as for his name, it is Pedro.' (or: 'The doctor's name is Pedro.')

In the above examples, three of the the four topic-accompanying clitics are *e'*. Only in the first topic in (13), *e'* is outranked by *o'*. Significantly, those clitics appear within a sentence, so the construction responsible for topics cannot inherit from *deictic-queue-cxt*. We need a new one, which stipulates the clitic-related mechanisms somewhat differently:

$$topical\text{-}cl \Rightarrow \begin{bmatrix} MTR & \begin{bmatrix} MRKG & topical \\ ENQ\text{-}D & \boxed{1} \\ DEQ\text{-}D & \boxed{1} \end{bmatrix} \\ \\ DTRS & \left\langle \boxed{2}\begin{bmatrix} ENQ\text{-}D & \boxed{3} \\ \mathbf{DEQ\text{-}D} & \mathbf{F_{max}}(\textbf{\textit{e'}}, \boxed{3}) \end{bmatrix}, \boxed{4}\begin{bmatrix} CAT & \begin{bmatrix} PRED & + \end{bmatrix} \\ VAL & \langle\rangle \\ GAP & \langle\boxed{2}\rangle \oplus L \\ MRKG & mrk \\ ENQ\text{-}D & \boxed{1} \\ DEQ\text{-}D & \boxed{1} \end{bmatrix} \right\rangle \\ \\ HD\text{-}DTR & \boxed{4} \end{bmatrix}$$

The most important part of the construction is in boldface. The clitic dequeued by the topic has to be *e'* or a higher ranked clitic, if one is enqueued by a constituent of the topic phrase. To formalize that, we use the F_{max} function again. The ENQ-D and DEQ-D of the mother are identified with those of the head.

Selected References

[1] Scott AnderBois. Focus and uninformativity in Yucatec Maya questions. *Natural language semantics*, 20(4):349–390, 2012.

[2] Grant Armstrong. On Copular Sentences in Yucatec Maya. In *Proceedings of the Conference on Indigenous Languages of Latin America-IV.*, 2009.

[3] Hans C. Boas and Ivan A. Sag, editors. *Sign Based Construction Grammar*. Center for the Study of Language and Information Publications, 2012.

[4] Jürgen Bohnemeyer. *Time Relations in Discourse: Evidence from a Comparative Approach to Yucatec Maya*. PhD Dissertation, Center for Language Studies Nijmegen-Tilburg, 1998.

[5] Jürgen Bohnemeyer. *The Grammar of Time Reference in Yukatek Maya*. Lincom Europa, 2002.

[6] Jessica Coon. When Ergative = Genitive: Nominals and Split Ergativity. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, pages 99–107, 2008.

[7] Yucatec Maya example Nº 110. (Available online at http://valpal.info/languages/ yucatec-maya/examples/110. Accessed on 2016-12-28). In: Lehmann, Christian. *Yucatec Maya Valency Patterns*. In: Hartmann, Iren & Haspelmath, Martin & Taylor, Bradley (eds.). *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.

[8] Colette Grinevald and Marc Peake. Ergativity and voice in Mayan: A functional-typological approach. *Ergativity, valency and voice*, pages 15–49, 2012.

[9] Rodrigo Gutierrez-Bravo and Jorge Monforte y Madera. On the nature of word order in Yucatec Maya. *Information structure in indigenous languages of the Americas*, pages 139–170, 2010.

[10] Edward Keenan and Bernard Comrie. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1):63–99, 1977.

[11] Christian Lehmann. *Possession in Yucatec Maya. Second, revised edition*. Universität Erfurt, 2002.

[12] Philip H. Miller and Ivan A. Sag. *French Clitic Movement without Clitics or Movement*. 1995.

[13] Elisabeth Norcliffe. *Head Marking in Usage and Grammar: A Study of Variation and Change in Yucatec Maya*. Stanford University, 2009.

[14] Elisabeth Norcliffe. *Revisiting Agent Focus in Yucatec*. Stanford University, 2009.

[15] Mike Reape. Domain union and word order variation in German. In J. Nerbonne, K. Netter, and C. J. Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, page 151–197. Stanford University: CSLI Publications, 1994.

[16] Elisabeth Verhoeven. *Experiential constructions in Yucatec Maya*. John Benjamins Publishing Company, 2007.