VERB COMPOUNDING IN YUKATEK MAYA:
A COMPLEX PREDICATE ANALYSIS

Session “Complex Predicates in the Languages of the Americas” at 2003 Annual Meeting of SSILA, Atlanta, GA, January 2-5.

J. Bohnemeyer, University at Buffalo and Max Planck Institute for Psycholinguistics
http://www.mpi.nl/world/persons/profession/bohnem.html mailto bohnem@mpi.nl

OVERVIEW

- Yukatek has a semi-productive template for forming compound verb stems (‘V-V compounds’) containing two verbs roots, V₁ and V₂, such that:
  - V₁ and V₂ must be transitive;
  - V₂ encodes the cause of the event denoted by V₁ (e.g. ‘break+hit’).

- V-V compounds have a number of peculiar properties, such as:
  - anti-iconic ordering (result before cause);
  - defeasible realization of the result.

- These are explained with reference to a superordinate X-V template:
  - of which the V-V template is an extension,
  - and in which X denotes the manner in which the event encoded by V is conducted.

I. INTRODUCING THE V-V AND X-V COMPOUND TEMPLATES

In V-V compounds, V₁ and V₂ are transitive roots:

1

(1) Le=x-ch’úupal=o’ t-uy=óop’+hàats’-t-ah le=plàato=o’.
DET=F-female:child=D2 PRV-A.3=burst+hit-APP-CMP(B.3.SG) DET=plate=D2
‘The girl, she burst-hit the plate.’ (ECR 07 NMP)

(2) T-uy=xíik+hàats’-t-ah y=éetel le=máartiyo=o’.
PRV-A.3=break.apart+hit-APP-CMP(B.3.SG) A.3=COM DET=hammer=D2
‘She crack-hit it with the hammer.’ (ECR 07 SBM)

Figure 1. Frame from ECR 07 clip

---

1 Yukatek belongs to the Yukatekan branch of the Mayan language family and is spoken by 700,000 to 800,000 people on the Yucatán peninsula in México and Belize. The language is exclusively head-marking, has rich valence-changing morphology, and shows productive incorporation of nouns and adverbs; in this sense, it exhibits polysynthesis. Argument marking in intransitive clauses is split, based on aspect-mood marking. The
Verb compounding in Yukatek

V-V compounds occur exclusively in transitive verb stems, yet, they are formally intransitive and require the ‘applicative’ (or applied object) suffix –t to license a ‘U(ndergoer)’ argument (in the parlance of Van Valin & LaPolla 1997).

V-V compounds (with transitive V₁s) have a characteristic pitch contour:

<table>
<thead>
<tr>
<th>V₁</th>
<th>V₂</th>
<th>V₁+V₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>bah ‘hammer’</td>
<td>lom ‘stab’</td>
<td>báah+lóom-t</td>
</tr>
<tr>
<td>buh ‘split’</td>
<td>ch’in ‘pelt’</td>
<td>būuh+ch’ín-t</td>
</tr>
<tr>
<td>nok ‘put upside down’</td>
<td>hats ‘hit’</td>
<td>nóok+hāats’-t</td>
</tr>
<tr>
<td>hen ‘demolish’</td>
<td>puch ‘crush’</td>
<td>héen+pùuch’-t</td>
</tr>
<tr>
<td>xik ‘break apart’</td>
<td>ts’eh ‘chip’</td>
<td>xiik+ts’ēeh-t</td>
</tr>
</tbody>
</table>

Table 1. Pitch contour in V-V compounds (with transitive V₁)

Transitive roots have a canonical CVC structure with a short nucleus unmarked for tone. In V-V compounds, V₁ occurs with a long syllable and rising pitch and V₂ with a long syllable and falling pitch. This matches the ‘anticausative’ (or ‘middle’) form for V₁ and the ‘antipassive’ form for V₂. This tonal pattern is certainly not coincidental; but it may be etymological. In the further analysis, it will be ignored.

V-V compounds do not accept intransitive verbs in V₁ position, not even ‘externally-caused’ state change verbs (‘unaccusatives’) – otherwise the most likely class of verbs to occur in causative constructions:

(3) *T-u=\text{kim(-s-ah)}+ch’iin-t-ah \text{le=hàaleb}=o’.  
PRV-A.3=\text{die(-CAUS-ATP)+pelt-APP-CMP(B.3.SG)} \text{DET=tepezquintle=D2}  
‘(S)he die-pelted/kill-pelted the tepezquintle (i.e. killed it by pelting it).’ (V-V FEE RMC SBM)

(4) *T-u=\text{kul+hàats’-t-ah} \text{le=chan pàal}=o’.  
PRV-A.3=\text{sit+hit-APP-CMP(B.3.SG)} \text{DET=DIM child=D2}  
‘(S)he sit-hit the toddler (i.e. hit it so that it fell on its buttocks).’ (V-V RMC)

V₁ and V₂ “share” both their arguments, i.e. the actor of V₁ must be coreferent with the actor of V₂, and the same holds for the two undergoers. (5) is rejected with respect to the scenario in Figure 2 (girl throwing hammer to guy, who drops plate, which breaks), because “She didn’t throw the plate”; i.e. the plate is the theme of V₁, but not V₂, and the girl is the actor of V₂, but not V₁:

---

2 This goes with the exception of environments that force antipassivization, such as predicate focus constructions. Applicative stems do not form antipassives; their intransitive bases are used instead.


4 Intransitive verbs in Yukatek fall into four different form classes, which are distinguished on the basis of the aspect-mood (or ‘status’) suffixes they take (not unlike Latin/Romance conjugation classes). The semantic basis of this classification is the process-change distinction: there’s one class of (internally or externally caused) process verbs and three classes of state change verbs. See Lehmann (1993), Lucy (1994), and Bohnemeyer (2002; to appear) for details.
(5)  

\[\text{Le=x-ch’úupal=o’ t-uy=óop’+pùul-t-ah le=plàato=o’}.\]

\text{DET=F-female:child=D2 PRV-A.3=burst+throw-APP-CMP(B.3.SG) DET=plate=D2}

‘The girl, she burst-threw the plate.’ (ECR 01 FEE)

Similarly, (6) is rejected wrt. the scenario in Figure 3 (fist hitting the table next to plate, which breaks), because “The table didn’t burst”; i.e. the plate is the theme of \(V_2\), but not \(V_1\):

(6)  

\[\text{Le=x-ch’úupal=o’ t-uy=óop’+lòox-t-ah le=mèesa=o’}.\]

\text{DET=F-female:child=D2 PRV-A.3=burst+punch-APP-CMP(B.3.SG) DET=table=D2}

‘The girl, she burst-punched the table.’ (ECR 01 FEE)

Of course, (7) is equally bad, because the plate isn’t hit in the scenario:

(7)  

\[\text{Le=x-ch’úupal=o’ t-uy=óop’+lòox-t-ah le=plàato=o’}.\]

\text{DET=F-female:child=D2 PRV-A.3=burst+punch-APP-CMP(B.3.SG) DET=table=D2}

‘The girl, she burst-punched the plate.’ (ECR 01 FEE)

V-V compounds encode a causal relation between the subevents denoted by \(V_1\) and \(V_2\). Yet, event realization of the compound does not entail realization of the \(V_1\) event, i.e. the result:

(8)  

\[\text{T-u=túup+ùust-t-ah le=kib=o’}.\]

\text{PRV-A.3=extinguish+blow-APP-CMP(B.3.SG) DET=wax=D2}

\text{pero ma’ h-túup-ih.}

\text{but NEG PRV-extinguish\text\{CAUS\}(CMP)-B.3.SG}

‘(S)he extinguish-blew the candle (i.e. blew at it so as to extinguish it), but it didn’t extinguish.’ (V-V FEE NMP)

\[\text{\#}\]  

I’m using the hatch mark (\#) for forms or constructions which are structurally well-formed, but cannot be used in reference to a particular scenario.
Verb compounding in Yukatek

(9) T-uy=óop’+hàats’-t-ah le=tunich=o’,
PRV-A.3=burst+hit-APP-CMP(B.3.SG) DET=stone=D2

pero ma’ h-óop’-ih.
but NEG PRV-burst\ACAUS(CMP)-B.3.SG

‘(S)he burst-hit the stone (i.e. hit it so as to cause it to crumble), but it didn’t burst.’
(V-V SME)

In contrast, realization of the V₂ event is strictly entailed by realization of the compound event.

The order of V₁ and V₂ is strictly anti-iconic (caused event before causing event); reversals are unacceptable.⁶

(10) a. T-u=xíik+pùuch’-t-ah.
PRV-A.3=break.apart+crush-APP-CMP(B.3.SG)
‘(S)he crack-crushed it (i.e. caused it to crack by crushing it).’

b. *T-u= pùuch’+xíik-t-ah.
PRV-A.3=crush+break.apart-APP-CMP(B.3.SG)
‘(S)he crush-cracked it.’ (V-V NMP)

At this stage, the following questions arise – which the further analysis will have to try to answer:

- Why go to the trouble of detransitivizing two transitive roots, only to yield a compound that must be obligatorily re-transitivated and then has all the formal properties of a simple transitive verb?
- If V-V compounds encode causal chains, then why exclude externally caused state change verbs, of all things, from the V₁ position?
- Why the restriction to complete argument sharing? If you have to have two transitive roots anyway, why not exploit the full expressive power of such a structure by admitting complex causal chains with multiple actors (as in Figure 2)?
- How is the encoding of causality reconciled with the defeasibility of V₁ realization? What does it mean to have caused an event that nevertheless didn’t happen???
- And why the anti-iconic order Yukatek is (almost) consistently head-initial, and as Talmy (2000: 153) argues, constructions that conflate causing and resulting events tend to construe the result as the ‘main’ and the cause as the ‘subordinate’ event.

The answers to these questions, it is argued here, comes from recognizing V-V compounds as an extension of another compound verb form, called ‘X-V compounds’ here, on account of the fact that the X in this structure doesn’t even have to be a verb. It may be an adverb, such as táan-il front-REL ‘early’ in (11a) and pàach-il back-REL ‘late’ in (11b):

---

⁶ To be more precise, the reversal is acceptable provided the reversed interpretation can be made sense of; which is rarely ever the case. One exception is káach+wùuts’ snap-bend ‘snap sth. by bending it’ and wùuts’+kàach bend-snap ‘bend sth. (permanently) by snapping it (w/o. complete separation of the parts)’.
(11) a. H-\text{táan-il+pak'al-nah-ih}. \\
PRV-front-REL+plant\text{ATP-CMP-B.3.SG} \\
‘(S)he planted prematurely.’

b. H-\text{pàach-il+pak'al-nah-ih}. \\
PRV-back-REL+plant\text{ATP-CMP-B.3.SG} \\
‘(S)he planted belatedly.’ (FN)

In many cases, X is an ideophonic particle:

(12) Pablo=e’ t-u=\text{wiich'+hàats’-t-ah} le=\text{pèek’ y=étel àak’=o’}. \\
Pablo=D3 PRV-A.3=\text{[whish]+hit-APP-CMP(B.3.SG)} DET=dog A.3=CON vine=D2 \\
‘Pablo, he \text{whish}-lashed the dog with a vine.’ (V-V RMC)

(13) Pedro=e’ t-u=\text{búus+lòox-t-ah} uy=\text{íitsin}. \\
Pedro=D3 PRV-A.3=\text{[buss]+punch-APP-CMP(B.3.SG)} A.3=younger.sibling \\
‘Pedro, he \text{buss}-punched his younger sibling.’ (V-V RMC)

V is generally a (derived or underived) activity (i.e. an agentive process) verb and X encodes the manner of this activity. Like V-V compounds, X-V compounds may be applicativized; however, with X-V compounds, this process is optional:

(14) a. Antonio Agilar=e’ h-\text{áawat+k’àay-nah-ih}. \\
Antonio Agilar=D3 PRV-\text{cry+sing\text{ATP-CMP-B.3.SG}} \\
‘Antonio Agilar, he cry-sang (i.e. sang very loudly/hollering).’

b. Antonio Agilar=e’ t-u=\text{áawat+k’àay-t-ah} le=ranchèera=\text{=o’}. \\
Antonio Agilar=D3 PRV-A.3=\text{cry+sing\text{ATP-CMP-B.3.SG}} DET=\text{ranchera=D2} \\
‘Antonio Agilar, he cry-sang the \text{ranchera} (i.e. sang it very loudly/hollering).’ (V-V EMB RMC SBM)

X-V compounds do admit both ‘unaccusative’ verbs ((15)-(16)) and internally caused process verbs (17) in the X slot. The meaning of these is adapted, be it by ‘coercion’ or by lexicalization of the compound, to the requirement of specifying the manner of V:

(15) Pedro=e’ \text{chéen wa’tal+tsikbal} t-u=\text{mèet-ah}. \\
Pedro=D3 only \text{stand:INCH+chat} PRV-A.3=\text{make-CMP(B.3.SG)} \\
‘Pedro, stand-chatting is all he did (i.e. he didn’t take the time to sit down and have a proper chat).’ (VV RMC)

(16) T-in=\text{máan+òok-s-ah-t-ah} le=\text{bòola te=gòol=\text{o’}.} \\
PRV-A.1.SG=\text{pass+enter-CAUS\text{ATP-APP-CMP(B.3.SG)}} DET=\text{ball LOC:DET=goal=D2} \\
‘I pass-entered the ball in the goal (i.e. I missed trying to score).’ (VV RMC SME)

(17) Ko’x \text{báaxal+ts’òon / kanbal+ts’òon!} \\
EXHORT \text{play+shoot / learn+shoot} \\
‘Let’s play-shoot/learn-shoot (i.e. let’s pretend/practic shooting).’ (VV RMC)

The analysis developed in section V argues that V-V compounds are adapted to the X-V template by treating V\textsubscript{2} as specifying the manner in which the event denoted by V\textsubscript{1} is conducted: this event is conducted in such a manner as to cause the event denoted by V\textsubscript{1}. This explains both the anti-iconic ordering of V\textsubscript{1} and V\textsubscript{2} and the fact that V-V realization doesn’t entail V\textsubscript{1} realization – all that is required for V-V realization is that the V\textsubscript{2} event is indeed conducted in such a manner as to cause the V\textsubscript{1} event; whether external circumstances nevertheless thwart realization of the V\textsubscript{1} event is immaterial.
II. REMARKS ON PRODUCTIVITY, COMPOSITIONALITY, AND USE

- V-V and X-V compounds are attested in Yukatekan (cf. e.g. Hofling 2000 for Itzá) and Ch’olan languages (cf. Vázquez Alvarez 2002 on Ch’ol), but not elsewhere in the Mayan language family. It seems plausible that this distribution is due to contact with Mixe-Zoquean languages (Roberto Zavala p.c.).

- Across natural genres of Yukatek, occurrences of X-V compounds vastly outnumber occurrences of V-V compounds. Especially in narratives and procedural texts, V-V compounds are quite infrequent.

- The most productive domain for (non-V-V) X-V compounds is that of goal-directed activities that in one way or another miss their goals, by being conducted too early, too late, too quickly, too slowly, too hurriedly, to roughly, etc.

- A semantic “hot spot” for V-V compounds is the domain of physical separation and destruction. Yukatek has an abundance of break-type verbs which specify a particular “style” in which a theme breaks and cut-type verbs which specify a particular manner of breaking a theme, e.g. by using a particular kind of instrument. Whenever there’s a mismatch – i.e. when a theme is broken in some atypical way – then V-V compounds are the way out.

- As (15)-(17) illustrate, compositionality in X-V compounds is severely reduced. This extends to V-V compounds as well, as is discussed in section V. This lack of predictability, in tandem with the relative narrowness of the semantic domains in which verb compounds are mostly used, poses significant limits on the productivity of the X-V and V-V templates. Nevertheless, it is quite possible to coin compound verbs on the spot, and this does occur in natural conversations.

III. ASSUMPTIONS ABOUT COMPLEX PREDICATES

Cf. in particular Ackerman & Webelhuth (1998):

- Complex Predicates (CPreds) are morphologically complex:
  - they may consist of multiple morphemes;
  - they may have parts which are targets of morpho-syntactic operations.

- CPreds have a single argument structure, projected from a single semantic ‘event structure’ representation – the latter may be a merger of the event structures of the components.

- CPreds instantiate a morpho-syntactic ‘template’ structure which has the properties of a ‘construction’ in the sense of Goldberg 1995. In particular:
  - the CPred template is a conventional form-meaning pairing, and thus encodes a meaning that is not encoded by any part of the template;
  - but the meanings of the parts are mapped into the template meaning compositionally;
  - there are ‘inheritance’ relations among CPred templates, such as (cf. Goldberg 1995: ch. 3):
    - ‘instance’: one CPred template may be a special case of another CPred template

---

7 Cf. also Fillmore 1988; Fillmore & Kay 1992; Fillmore, Kay, & O’Connor 1988.
IV. A CPRED ANALYSIS OF X-V COMPOUNDS

Adapting Goldberg’s (1995) Construction Grammar diagrams to CPred templates:

A further prerequisite – the applicative form:

---

---

---
V. THE V-V TEMPLATE AS AN EXTENSION OF THE X-V TEMPLATE

The intransitive V-V template is an extension of the X-V template. The transitive V-V template is formed by applicativization of the intransitive V-V template; i.e. the intransitive V-V template and the Applicative template are subparts of the transitive V-V template. The inheritance relations among the various templates are diagrammed in Figure 6 (I_M represents an extension link and I_S a subpart link):

![Diagram of inheritance relations among CPred template in V-V compounds](image)

The semantic link that licenses extension from X-V to V-V is ‘manner-as-to-cause’. The V_1 event is understood as a “manner goal” of V_2, so to speak: the V_2 process is conducted in a manner such as to cause the V_1 event.

However, since the X-V template does not encode a causal relation, causality has to be encoded in both V roots, in a lexically “encapsulated” representation, as it were:

![Diagram of extension inheritance between X-V and V-V templates](image)

For instance, in view of the analysis in Figure 7, (1) above, repeated here as (18), should be glossed as ‘She hit the plate in such a manner as to cause it to shatter’:
The semantic relations among the subevents of $V_1$ and $V_2$ are not made explicit in Figure 7, except for the coindexation of their participants – the themes of the two CHANGE subevents must be coreferent, and the same holds for the agents (or ‘effectors’, cf. Van Valin & Wilkins 1996) of the two PROC subevents. Effectively, the PROC subevent specification of $V_2$ (which is tapped into by the V-V template) semantically specifies the PROC subevent of $V_1$, while the CHANGE subevent specification of $V_1$ semantically specifies the CHANGE subevent of $V_2$. In Figure 8, these specification relations are represented by dotted arrows: 

Figure 8 relies on the assumption that $V_1$ is a verb with a lexically underspecified (i.e. generic) PROC subevent, whereas $V_2$ denotes an underspecified CHANGE subevent. This assumption is in line with $V_1$ being an anticausativizing verb and $V_2$ an antipassivizing verb (cf. Bohnemeyer 2001; to appear). V-V compounds on this account that licenses the formation of transitive verbs with two lexically specified subevents.

Note that the intransitive V-V template in Figure 8 has one unlinked thematic relation (or two coreferent ones – namely the themes of the two CHANGE subevents). This makes the intransitive V-V template an ill-formed construction, on account of a principle such as Van Valin & LaPolla’s (1997: 325) Completeness Constraint:

**Completeness Constraint:** All of the arguments explicitly specified in the semantic representation of a sentence must be realized syntactically in the sentence, and all of the referring expressions in the syntactic representation of a sentence must be linked to an argument position in a logical structure in the semantic representation of the sentence.

This problem is taken care of by inserting the V-V$_{itr}$ structure in an applicative stem, which introduces a U argument with the relevant linking specifications, as depicted in Figure 9:

---

The account presented in this section is essentially compatible with intransitive activity-denoting V2s as well (lexical “encapsulation” of a causal chain in V1 would seem sufficient to meet the requirements of manner-as-to-cause), apparently contrary to fact. This remains to be investigated.
Verb compounding in Yukatek

Figure 9 depicts the caused state change introduced by the applicative form as tapping into the corresponding CHANGE subevents of the two V roots. This is a simplification. All that is required e.g. for (18) to apply is that the undergoer (the plate) is somehow causally affected by the action of hitting it in such a manner as to cause it to shatter. Crucially, the plate doesn’t have to actually shatter – (18) according to Figure 9 doesn’t entail an unbroken causal chain leading from the activity component of hitting to the shattering subevent, but merely a manner link between these subevents.

The analysis as presented so far accounts for the following properties of V-V compounds:

- The anti-iconic order of V₁ and V₂: this is inherited from the X-V compound (V₁ filling the X slot) (and it isn’t really anti-iconic, since V₁ doesn’t really denote a causal result).
- Defeasibility of V₁ realization: likewise inherited from the X-V template (V₁ encodes the manner in which the V₂ activity is conducted, not the result of V₂).
- Basic intransitivity of V-V compounds: again inherited from X-V – V-V compounds encode activities in a manner-as-to-cause, leaving causality “encapsulated” in the V roots.
- Restriction to applicative stems: to satisfy Completeness Constraint on linking (manner-as-to-cause does require two distinct subevents).
- Restriction to transitive roots: to represent causality at least in the lexically “encapsulated” form – the X-V template does not afford the encoding of causality at all.
- Complete argument sharing: probably ultimately an artifact of general semantic constraints on transitive verbs, which the applicative form cannot violate.

One additional piece of evidence for the V-V-as-X-V analysis: it predicts – correctly! – that V-V compounds should admit ambiguous X-V readings:

(19) \( T-u= \text{ch’iik}+\text{ch’iin-}t-\text{ah} \quad \text{le}=\text{bòola} \quad t-\text{uy}=\text{ét+}bàaxal=0' \).

PRV-A.3=stick.in+pelt-APP-CMP(B.3.SG) DET=ball LOC-A.3=CON+play=D2
‘(S)he stick-pelted the ball to his fellow player’; i.e.

V-V reading: pelted F. with the ball and caused the ball to get stuck in F.’s body

X-V reading: hit the ball to the ground forcefully, such that it bounced off towards F.

(VV RMC)
VI. CONCLUSIONS

- The V-V template is parasitic on the X-V template, as an exploitation of the X-V template for the encoding of causal chains.
- The semantic link that enables accommodation of V-V under X-V is “manner-as-to-cause”.
- The pragmatic function of V-V compounds is to allow lexical specification of both subevents denoted by a transitive verb form.

REFERENCES

México.

**APPENDIX: GLOSSES AND ORTHOGRAPHICAL CONVENTIONS**

The orthographic representation in this paper is morphemic rather than morpho-phonemic. The orthography applied is based on Lehmann (1998). In the interlinear morpheme glosses, the following conventions are used: ‘-’ for affixes; ‘=’ for clitics; ‘+’ for compounding; ‘/’ for subsegmental realization or infixation. Abbreviations in the glosses include the following: 1 – 1st person; 2- 2nd person; 3 – 3rd person; A – set-A cross-reference clitics; ACAUS- anticausative derivation; ALL – universal quantifier; ALT – ‘alternative’ particle (question focus, conditional protasis, disjunctive connective); APP – applicative derivation; ATP – antipassive derivation; B – set-B cross-reference suffixes; CAUS – causative derivation; CAUSE – causal preposition; CMP – completive status; D2 – distal-deictic/anaphoric particle; DEF – definite determiner; EXIST – existential/locative/possessive predicate; IMPF – imperfective aspect; INC – incompletive status; INCH – inchoative derivation; IRR – irrealis modality; LOC – generic preposition; PASS – passive derivation; PL – plural; PROG – progressive aspect; PRV – perfective aspect; REL – relational derivation (nouns); RES – resultative derivation; SG – singular; TERM – terminative aspect; TOP – topic marker.