

GRAMMATICALIZATION
THROUGH INCORPORATION
IN THE DIACHRONY
OF MIJE-SOKEAN VERBS
[WITH A FEW OTHER TIDBITS]

Terrence Kaufman

University of Pittsburgh

Pisces-Aries cusp 1997

prepared for the Rice Conference on Grammaticalization, 27-29 March 1997

lightly updated Oct 2016

Copyright 1997, 2016 by Terrence Kaufman

NOTE. The field work aspect of the PDLMA was in effect from 1993 through 2010. In the course of this field work 10 Mije-Sokean languages were documented lexically and via texts. Lexical databases had between 5000 and 20,000 entries. For several languages (MIG OLU SOT), descriptive grammars were written as doctoral dissertations.

ABSTRACT

The Mije-Sokean languages have what is essentially a radically right-headed morphology, with incorporation, and right-headed syntax. However, straightforward reconstruction of grammatical patterns and their morphological exponents to proto-Mije-Sokean reveals a few points where incorporated verb roots serve grammatical functions in apparent opposition to right-headedness. In each of these cases at least some Mayan languages (always some westerly ones) have an analogous pattern. Other cases of incorporation involving formerly independent verb roots are attested in various daughter languages, some of them in harmony with right-headedness, and some in harmony with the anti-right-headedness of the incorporations having Mayan analogs.

INTRODUCTION

I have carried out comparative work on Mije-Sokean from 1958 through 1963 and from 1991 through the present (1997). I have carried out field work on several M-S languages since 1993, and headed a project that is producing dictionaries and text collections for all the Mije-Sokean languages. My reconstruction of earlier stages has operated on the assumption that unless diffusion is involved features shared by two or more members of a group go back to the group's common ancestor, as well as that any feature found in one language of a group and one or more languages outside that group necessarily go back to the first group's common ancestor. In these languages, since they have essentially the same set of underlying segments, phonological facts rarely allow the distinction of similarity as resulting from cognacy or diffusion. Certain specific similarities between adjacent languages belonging to different subgroups are suspected of being due to diffusion. Some of what is here projected back to proto-Mije-Sokean, proto-Sokean, etc. may belong to a later stratum: but I am playing the cards dealt me, without at the outset trying to put a twist on the obvious.

The most important other work comparing Mije-Sokean languages is that of Søren Wichmann, carried out since 1989, yielding several monographs and articles.

NOTE:

Specialized use of lexical items as grammatical markers is not treated here: e.g. MAR /+k(@) 7it-w@/; COP /7in@k/; SOT /7ity7@k/, all = 'in the past'
pMijejan *7it-i AUX; SAY /+it/ 'NEG.IMV'; EMI /+@ty/ 'STATive'

Possible reclassification of adverbial NEGators as inflexional prefixes is not treated here.

Classification of the Mije-Sokean languages

MIJEAN BRANCH [Mijean]

- TAPachula
- OLUta (Popoluca)
- Mije group
 - SAYula (Popoluca)
 - Mije group proper [Mije]
 - Eastern/Lowland Mije language
 - San Juan GUIchicovi
 - JUQuila
 - Western/Highland Mije language
 - TOTontepec

SOKEAN BRANCH [Sokean]

- Epi-Olmec [EO]
- Soke group proper
 - Western/Oaxaca/Chimalapa Soke language
 - San MIGuel Chimalapa
 - Santa MARI*a Chimalapa
 - Eastern/Chiapas Soke language
 - COPainala*
 - TECpata*n
 - MAGdalena = Francisco Leo*n
 - RAYo*n
- Gulf Sokean [GS]
 - AYApa ("Zoque")
 - TEXistepec (Popoluca)
 - SOTeapan = Sierra Popoluca

The ***Epi-Olmec language*** is embodied in several texts from Southern Veracruz produced in a writing system deciphered by myself and John Justeson between 1991 and 1992. Of the four legible texts, there are two with dates: these date from the second century of our era (ca. 140-160 CE). Including the illegible texts, the writing system may span a period of from 200 BCE to 600 CE. The texts that have been read represent a Sokean language slightly more archaic than what can be reached reconstructing backward from the surviving Sokean languages. It has two or three grammatical traits that could not have been suspected given what is known about present-day Sokean languages. None of the recovered grammar traits of EpiOlmec relate to the grammaticalization issues being dealt with here.

All dates are educated guesses.

Proto-Mije-Sokean split into pSokean and pMijean ca. 1000 BCE

Proto-Mijean split into TAPachulteco, OLUta, and the MIJe group ca. 300 CE

Proto-Sokean split into Gulf Sokean and Soke proper ca. 500 CE.

The Mije group split into SAYula and MIJe proper ca. 1000 CE.

Mije proper split into Eastern Mije and Western Mije ca. 1200 CE.

Gulf Sokean split into SOTeapan, TEXistepec, and AYApa ca. 1200 CE

Soke proper split into Eastern Soke and Western Soke ca. 1200 CE

grammatical (and one phonological) boundary markers:

+X, X+	enclitic, proclitic
-X, X-	inflexional affix
>X, X>	syntactically-motivated class shifting affix
.X, X.	lexical derivational affix
=X, X=	postpound, prepound
%X	root

transcription:

p	t	ty	ch	tz	k	ʔ
		x	s			j
m	n				nh	
B	D					[voiced imploded]
		y			w	
i	@	u				
e		o				
	a					
length /V]V ₁ /						
stress /V]*/						

"morphophonemes"

/A/: underlying /@/ which becomes /a/ following a syllable containing a mid vowel.

/E/: underlying /e/ which becomes /i/ following a syllable containing a high vowel.

The verb in Mije-Sokean [here is dumped whatever will be discussed later].

The following section is designed to provide a typological outline/overview of verbs in Mije-Sokean, citing only morphemes found in at least two languages, spelling them as if they were proto-Mije-Sokean elements.

Reconstructed forms are proto-Mije-Sokean unless otherwise noted. Incorporated nouns and verbs, when their origins are known, are **bolded**.

NEGATORS

*7u(y)+ [pSokean]
*jay +
*ni+

PERSON MARKERS

ABSolutive

X *7@+
I *t@+
2 *mi+
3 *∅

ERGative

X *na+
I *t@n+
2 *7in+
3 *7i+

VALENCY SHIFTERS

Reflexive-Reciprocal *na:y+ [<= pronoun 'self']
CAUsative *yak= [<= old verb]
PaSsiVe *yak= [pMijean: same as CAU]
ASsoCiative *m@:. [pMijean: <= *m@:t 'companion']
*n@. [pSokean];
INStrumental *tuk. [pMijean]

EVENT MODIFIERS

OTHER [elsewhere, someone else's] *ko:.
PuRPose *ni:.
AWAY *ku. [pSokean]
ON *n@k. [Soke]
"MOUTH" *7aw=
"HEAD" *ko7= <= *ko+ pak
FORWARD *win= [pMijean <= 'face']
UPWARD *yuk= [<= relational noun]
DOWNWARD *7ok= [Soke <= 'end']
BACKWARD [Mi]/AWAY [So] *j@s= <= '*back'
ON SURFACE/BODY *ni:7= [old root?]

V E R B R O O T

VALENCY REDUCERS

ANtiPassive *>7oy [pSokean]
 iNDefinite Subject *>An@m [pSokean]
 PaSsiVe *>At@j [pSokean]
 iNTransitivizer *>j@ [pMijean]

VALENCY INCREASER

INDirective [3 -> 2, G -> 2]
 *>ja:y7
 CAUsative₂ *>a7j [pSokean]
 [of movement intransitives]

OTHER POSTVERBAL adverbial ELEMENTS [more than one of these may occur]

ALL 'entirely, completely'
 * =p@7
 REPetitive *-REP (cf. -ke7t [Soke], -gak [GS],
 -7ok [Mije group], etc.)
 SuPerFactive *.ka7 [pSokean]
 DESideratve ***=wa:n7 [Mije group]**
 DESiderative *-to7 [Sokean: ESoke, SOT, TEX];
 FRuStrative 'wanted to but didn't' [Sokean]
***-t@7p [MAR, SOT] ~ *-t@7 [ESoke, TEX]**
 iNteNsive ***=ta:k7 [pMijean] <= *'to weave'**
 ReVerSive 'un(do)'
 =pu7 [WSoke]
 PeRSeverative 'to stand'
 =tena:y7 [WSoke]
 PRoBative 'to try'
 =ki7ps [WSoke]
 HaBiLitative 'can'
 =mus [WSoke]
 TeRMinative 'to finish'
 =tuk [WSoke], =yaj [TEX]
 iNCePtive 'to begin'
 =7ok.manh [WSoke], =ten [TEX]

DIRectionals

UP <= *'to go.up' * =ki7m [pSokean]
 DOWN <= *'to go.down' * =wanak
 IN <= *'to go.in' * =t@k.7@y
 OUT <= *'to go.out' * =put [pSokean]; * =pitz@m [pMijean]
 HITHER <= *'to come' * =mi:n7
 HENCE <= *'to go' * =n@ks
 BACK <= *'to return' * =witu7 [WSoke]: * =pit(u) [pMijean]
 OVER <= *'to cross.over' * =jak [pSokean], etc.
 PAST <= *'to pass' * =nas [pM-S]; =k@.tuk [Soke]
 ELSEWHERE <= *'to arrive' * =nu7k [pSokean], etc.
 THERE <= *'to stay' * =tz@7y [pSokean], etc.
 <= *'to leave' * =tzak [pSokean]

PLural

I-X-2 PLural *-ta7m
 3 PLural ***BE.FINISHED [identical with the lexical verb]**

*-yaj [pSokean]; -suk [WSoke];
 *-k@x [pMijean]

POTential *=**wa:n7** [pMijean] <= *`to want'
 PeRFect *-na:y7 [EO, SOT]
 ALREADY *-n@H [pMijean]
 Aspect-Mood
 INDependent INCompleTive *-pa
 INDependent CoMpleTive *-w@
 DEpendent INCompleTive *-e (> *-i after V_{high})
 DEpendent CoMpleTive *-ji [pMijean and EpiOlmec, therefore pM-S]
 IMperative *=@7 (> *-a7 after V_{mid})
 VETative [= DEP] *-w@₂
 OPTative *-7in [pSokean]
 IRRealis [= DEP] *=@ (> *-a after V_{mid}) [pSokean]
 LiNKer to AUX *=@ [WSoke];
 -i ~ -@ [OLU]

third person object *-e [pMijean]
 exclusive object *-k [pMijean]

AUXiliaries (preceded by LiNKer with no ERGative shift)
 MOVement [focus on actor]

ANDative *%n@ks
VENitive *%min
VoLTative *%7oy [pM-S], %tij [Soke]

OTHER

DESiderative *t@7 [WSoke]
 *7o:k [pMijean] lit. `to die'

#####

An overview of proto-Mije-Sokean grammatical structure.

pM-S was (at first glance) radically right-headed.

The predicate is preceded by its arguments, the object closer than the subject.

clause order:

S (O) P

The predicate contains a verb phrase or an equational predicate.

The verb phrase contains a lexical verb or an auxiliary governing a lexical verb. The verb phrase either has valency or marks valency; it must mark [a] aspect-mood, [b] subject person agreement; [c] object person agreement, when required lexically. The final word of the verb phrase bears markers of aspect-mood. The first verb (which can be the only verb) of the verb phrase bears subject and object agreement markers.

The **auxiliary construction** consists of a "main verb" marked for appropriate subject and object agreement and suffixed with a morpheme (called LiNKer) that shows that it is governed by the "auxiliary verb" that follows. The auxiliary verb is suffixed with the aspect-mood suffix that the particular clause calls for. It is not, however, marked for subject or object agreement. This shows that the auxiliary construction is a case of **clause union** but **not incorporation**. Neither the main verb nor the auxiliary verb is inflected for all the categories called for by the particular clause type: each in some way leans on the other. The auxiliary verb can be [a] a verb of movement, encoding VENitive, ANDative, and VoLTative functions; [b] a DESiderative. The MOVement AUXiliaries are lexically intransitive; the DESiderative is lexically transitive. Thus, it is not plausible to argue that the MOVement auxiliaries are inflected for 3ABS, which is phonetically ZERO; inasmuch as person marking is absent on the "transitive" auxiliary DESiderative, person marking on auxiliaries is simply not there.

AUXILIARIES (preceded by LiNKer with no ERGative shift)

ANDative	*%n@ks
VENitive	*%min
VoLTative	*%7oy [pM-S], %tij [Soke]
DESiderative	*t@7 [WSoke];
	*7o:k [pMijean]

Verb lexemes may consist of

a verb root;

a non-verb root with a verbalizing derivational suffix;

a radical or derived verb stem with **an incorporated lexical noun that serves as an internal object, instrument, or circumstance**; such verbs are often but not always intransitive when compared to the valency of the incorporating verb;

a radical or derived verb stem with an incorporated element that no longer serves as an independent noun lexeme in any surviving language, but is likely to have been such at a somewhat earlier stage: these occur in two functions; [a] **event modifier**, an incorporated noun that focuses on the place of an event; [b] **valency shifter**, which adds or subtracts arguments: while some of these are nouns, **others seem to have been verbs**.

IT IS NOT CLEAR HOW A VERB THAT CHANGES THE VALENCY OF ANOTHER VERB IS NOT THE HEAD OF THE CONSTRUCTION IN WHICH IT OCCURS. THIS IS THE FIRST PROBLEM WITH STRICT RIGHT-HEADEDNESS.

[a] EVENT MODIFIERS [incorporated nouns]

"MOUTH"	*7aw=
"HEAD"	*ko7= <= *ko+ pak
FORWARD	*win= [pMijean <= 'face']
UPWARD	*yuk= [<= relational noun]
DOWNWARD	*7ok= [Soke <= 'end']
BACKWARD [Mi]/AWAY [So]	*j@s= <= '*back'

ON SURFACE/BODY *ni:7= [old root?]

[b] VALENCY SHIFTERS

CAUsative	*yak= [\leq old verb!!]
PaSsiVe	*yak= [pMijean: same as CAU]
Reflexive-Reciprocal	*na:y+ [\leq pronoun 'self']
ASSoCiative	*m@:- [pMijean: \leq *m@:t 'companion']
	*n@- [pSokean];

Some event modifiers are not identifiable with any independently occurring noun or verb lexeme. Since they are preposed, if pM-S was indeed radically right-headed, they should be incorporees.

[b] other EVENT MODIFIERS

OTHER [elsewhere, someone else's]	*ko:.
PuRPose	*ni:.

Some westerly Mayan languages create causative auxiliary constructions by placing the verb *7aq' 'to put, give' to the left of a lexical "main" verb, and inflecting both with the appropriate person markers and aspect-mood categories, the main verb being marked dependent. Proto-Mijean has a verb *yak vt, glossed variously in the daughters as 'to give freely, place, put, leave'; a Gulf Sokean possible cognate means 'to push, remove'. If, as seems plausible, the common Mije-Sokean causative *yak is in origin this verb, the structure seems left-headed. It is curiously similar to the Mayan construction in order and in pronunciation. Is something going on here?

The verb word is terminated by an aspect-mood suffix, which heads the clause and governs the lexical verb. The verb word is begun with an absolutive person agreement marker, and this is followed by an ergative person agreement marker if needed by the verb. (Intransitive verbs under certain circumstances take an ERG marker instead of an ABS marker: this is called **ergative shift**).

A variety of **inflexional suffixes** may be attached to the verb: they include valency changers, and a variety of adverbial notions, followed by a pluralizer of either the subject or object of the verb.

VALENCY REDUCERS

PaSsiVe	*>At@j [pSokean]
iNTransitivizer	*>j@ [pMijean]

VALENCY INCREASER

INDirective [3 -> 2, G -> 2]	*>ja:y7
------------------------------	---------

OTHER adverbial POSTVERBAL ELEMENTS

ALL 'entirely, completely'

REPetitive *.p@7
 *-REP (cf. -ke7t [Soke], -gak [GS],
 -7o*k [Mije group], etc.)

The verb word ends with an aspect-mood suffix (or a LiNKer, if it is governed by an AUXiliary):

ASPECT-MOOD

INdependent INCompleteive *-pa
INdependent CoMpleteive *-w@
DEPendent INCompleteive *-e (> *-i after V_{high})
DEPendent CoMpleteive *-ji [Mijean and EpiOlmec]
IMperative *-@7 (> *-a7 after V_{mid})
VETative [= DEP] *-w@₂
LiNKer to AUX *-@ [WSoke];
 -i ~ -@ [OLU]

HOWEVER, BEFORE THE ASPECT-MOOD MARKERS, THERE CAN OCCUR AS INFLEXIONAL ELEMENTS INCORPORATED VERB ROOTS -- IN TWO DIFFERENT FUNCTIONS -- THAT, THOUGH THEY APPEAR AFTER THE LEXICAL VERB, DO NOT GOVERN IT AND DO NOT DETERMINE THE VALENCY OF THE WORD.

[A] Immediately before the aspect-mood markers are found markers of plural subject or object. Only one plural participant can be marked, but whether there are any preferences for person or function with transitives is not yet known. The two categories of plural participant are [a] X-I-2 PLural (or non-3PLural), [b] 3PLural. Non-3PLural is marked with *-ta7m. **BUT 3PLural is marked with whatever happens to be the current lexical verb meaning 'to be finished'**; this is an intransitive verb with a non-agentive subject. **This peculiarity is shared by all Mije-Sokean languages but one.**

[B] Before the plural participant markers and after other adverbial inflexional elements are found another kind of incorporated inflexional elements that do not affect the valency of the lexical verb which they serve to inflect: DIRectionals. As independent lexemes, directionals are intransitive verbs of motion (or rest). As inflexional incorporees, they serve to indicate the direction (or lack of it) imparted to the object of a transitive verb, or the movement engaged in by the subject of an intransitive verb. They are common with transitives, and not common with intransitives. The use of directionals in Mije proper is still not well understood, but they are flourishing in all other M-S languages.

The use of fully independent lexical verbs as post-lexemic inflexional elements within the verb word is an apparent violation of the radical right-headedness that otherwise characterizes pM-S. Unlike auxiliaries, they are not linked to the lexical verbs that they help inflect by any kind of linker that would show that they are (or ever were syntactically) separate

words from the preceding lexical verbs.

An analog of the use of BE.FINISHED as an inflexional element to encode plural participant is to be found in Tzeltal, Chol, and Yukateko, all Mayan languages spoken in the vicinity of Sokean languages.

A fairly exact analog to the M-S directionals (in terms of function and lexical source) is found in most Mayan languages (though not Wastekan, Yukatekan, or Ch'olti7-Ch'orti7, and they are weakly developed in Q'eqchi7). Directionals in Mayan are not incorporated (Mayan lacks this grammatical device), but they follow the fully inflected verb they serve to modify, and have an invariant form -- though the lack of agreement amongst the various Mayan languages on just what suffix they bear, since a bare verb stem simply does not occur in Mayan any more than in Mije-Sokean, shows that originally directionals were inflected in some way, either as subordinate to the head verb, or in parallel with it, or in some as yet not understood way. In fact, Mamean languages show some mutability of the relative order of main verb and directional.

The development of directionals in both Mayan and Mije-Sokean can hardly be entirely independent, but a specific scenario is not yet in the cards. Nawa also has an analog of directionals, though not such an exact one, and I have argued [Kaufman 198x ms] that the Nawa pattern is modeled on the pattern found in Mije-Sokean; but this does not decide the issue of whether Mayan or M-S developed directionals first.

summary of verb word morpheme order:

- ABSolutive
- ERGative
- valency changers
- incorporated event modifiers
- verb root
- valency changers
- adverbial modifiers
- directionals
- plural
- aspect-mood

#####

NOUN WORD

The noun is preceded by its modifiers.

NP order:

Dem Num AdjP/RelCl/GenNP N (case/Loc)

RELATIVIZATION

Clauses are relativized by postposing a relativizer *+p@7k (with variants) or *+w@7 to the predicate; these relativized clauses serve as pre-nominal modifiers or as independent noun-like constituents. The whole structure and its nested parts are strictly right-headed.

POSSESSION

The possessed noun agrees with its possessor for person. If the possessor is a non-third person pronoun, it is usually omitted.

Possessive construction:

Possessor.NP # 3ERG-Possessed.NP.head

The possessor N may be itself possessed (and in Soke may optionally be marked for ergative case, ergative being the case of possessors). The possessed NP head may be marked for any case (except for Soke this is only LOCative).

LOCATIVES

The noun may be specified for its location by a **locative relational noun** which occurs to the right of the noun, forms a "loose compound" with the noun (both noun and relational noun are stressed, unless the relational noun is cliticized), and governs the noun. The noun that is compounded with the locative noun may be possessed or pluralized. The possession and pluralization relate to the noun, not the locative relational noun.

The noun may have "locative case" suffixes that specify its function in the clause. If the noun is governed by a locative relational noun, the relational noun is necessarily followed by locative case suffixes. These govern the rightmost noun, are the last morpheme in the phrase, and are the head of their phrase.

Locative construction

Pref (7aw=, ku.) - RN - case₁ (+m@7, +ji) - case₂ (+k@)

EXX [more exx from Mijejan languages need to be dug up]:

k@7+m@7 [SOT] 'at'

k@7+ji [COP] 'near'

7aw+m@7 [ESoke, JUQ]

7aw+ji [pSoke] 'near'

win+m@7 [M-S: MAR, OLU] 'on, above, over'
 win+pi [OLU] 'over'

 yuk+m@7 [SOT, SAY] 'on, above, over'

 k@s+jj [ESoke, EMije] 'on, by the top of, at'
 k@s+m@7 [COP, OLU] 'above, over, on the ground'

 k@7+m@7 [ESoke, SOT, Mije] 'under, below, at'
 ku.k@7+m@7 [SOT] 'under'
 ku.k@7+jj [MAG] 'under'
 pat+pi [OLU] 'under'

 joj+m@7 [pSokean, Mije] 'in'
 joj+jj [pSoke] 'in'

 joj+jj+k@ [MAR] 'from'

 7aw=joj+m@7 [SOT] 'between'
 7aw=kuk+jj [SOT, Mije] 'between, in the middle of'
 kuk+m@7 [COP, Sot] 'between, among, in the middle of'
 kuk=joj+jj [pSoke] 'between, in the middle of'
 kuk+jj [pSoke] 'between, in the middle of'

 7aw=win+jj [MAG] 'alongside'

 7aw=k@7+m@7 [OLU] 'in front'
 win=joj+jj+k@ [MAR] 'before'
 win=kuk+m@7 [OLU] 'in front of'

Relational Nouns

win 'face'
 yuk 'upper part'
 k@s 'top'
 kuk 'middle'
 joj [Sokean] ~ jo:t [Mije] 'inside' <= *heart
 k@7 'space beneath; nearness' <= *hand
 pa7t [pMije] 'space beneath' <= *liver?

locative case₁:

+m@7 'at, on, with; where'
 +jj 'where' [pSokean]
 +pi [pMije]

locative case₂:

+k(@) 'from' [pSokean]

LOCATIVE MARKERS in Sokean

The locational case markers are of four types and involve the morphemes listed below. The following combinations are attested:

1. *+m@7 [1 possibility]
2. *7aw + *+m@7 or *+ji [2 possibilities]
3. CORE + *+m@7 or *+ji [10 possibilities]
4. *7aw + CORE + *+m@7 or *+ji [10 possibilities]

Most of the possible combinations occur in one or another of the languages, but only a few occur in any one language. Many of these complex items also function as independent adverbials. It seems likely that the original status of all the locational markers was that of independent adverbial based on noun roots, but that some of the adverbials have decreased their distributions, being now only postposed to nouns, and that some noun roots have become obsolete except in these adverbials.

locative preposed element

***7aw=.** This is an incorporated form of the pMZ word for 'mouth', but no clear meaning can be discerned when it serves as part of a locative adverbial/adposition.

***=7aw+m@7** 'near, at' has reflexes in ESo and oMi.

***=7aw=kuk(+ji)** 'in the middle of' has reflexes on SOT and OM.

locative CORE

***=joj** 'inner part'. (Wichmann 1991: AF#061 pZ *-hoh-m@ 'in')

ESo -7oj-mo (#617) 'locative'

WSo -jo7 'en'

SOT -joo-m (#622) 'en'; -anh+joo-m 'entre'

***=joj+m@7** 'inside of' has reflexes in ESo, SOT, and OM.

***=kuk** 'middle part'. (Wichmann 1991: AF#044 pMZ *-kuk 'in the middle of')

ESo -kuk- (#628) 'in the middle of';

-kukoyj < -kuk=joj+ji (#627) 'in the middle of';

-kuk-m@ (#626) 'in the middle of'

-kukji < -kuk+ji (#625) 'in the middle of'

SOT -anh-ka7k < -7anh+kuk+ji (#627) 'en medio de'

***=kuk-m@7** 'in the middle of' has reflexes in ESo and SOT.

***=k@s** **'upper part'**. (Wichmann 1991: AF#042 pMZ *-k@s-i 'on, at';
also AF#043 pMZ *-k@s-m@
 'position with respect to the ground')

ESo -k@s@ (#621) 'on'; -k@s-m@ (#622) 'above'
OLU -k@x-m@ 'en el suelo'
*LM -k@x 'at, on, by the top of' [SW]

***=k@s-m@7** 'above/over' has reflexes in ESo and OM.

***=yuk** **'upper part'**.

***=yuk+m@7** 'above/over' has reflexes in SOT and SAY.

***=k@7** **'lower part'**. (Wichmann 1991: AF#090 pZ *-k@7-m@ 'under')

ESo -k@7-m@ (#624) 'under'
SOT -k@7@-m (#624) 'a'

***=k@7+m@7** 'under/below' has reflexes in ESo, SOT, and OM.

locative suffix position +1

****m@7** **'locative₁'**. (Wichmann 1991: AF#045 pMZ *-m@)

WSo +m@7
ESo -m@ 'place-name marker or locative'
SOT -m@ 'instrumental, locative'
SAY -m ~ -Vm 'location'
OLU -m@ 'position'

****+ji** **'locative₂'**. (Wichmann AF#046 pMZ *-i)

ESo -ji ~ -y (#619) 'locative'
WSo -i ~ -j 'lugar'
*LM -y 'locational: in, to'

*+m@7 and **+ji both mean simply 'location' and have no obvious distinction in meaning.

locative suffix position +2

***-k** **'ablative'**. Meaning: 'from' -- can follow stems with simplex
*+m@7 only; this includes place names. It can safely be assigned only to

demonstratives occurring with *+m@7, but quite possibly it had a wider range in the protolanguage.

cf. SOT tzu*7m@k 'last night' < tzu7 'night' + +m@7 'locative' + -k 'from'.

The noun lexeme may be preceded by a possessive agreement marker, a modifier; it may be followed by a pluralizer which refers either to the noun or the possessor. The noun may be governed by a case marker at the extreme right.

noun word morpheme order [simplified]:

		noun root	noun derivation	
possessor			OR	plural
case/loc.expression		verb root + nominalizer		

SUMMARY ON GRAMMATICALIZATION BEFORE THE BREAK-UP OF PROTO-MIJE-SOKEAN

To summarize on grammaticalization before the proto-Mije-Sokean stage. This is recognized in grammatical affixes that clearly (or apparently) have their origin in what was originally a lexical item.

Apparently, before pM-S broke up, ca. 1000 BCE, it had innovated a device for incorporating lexical verbs as inflexional markers on verbs. This innovation, with respect to postposed material, has analogs in Mayan, which cannot be accidental, though an explicit hypothesis would probably be too daring at this particular point. But it should not be ruled out that the M-S event was later than the break-up of the family and spread within the family by internal diffusion.

Later M-S languages, especially Sokean, made use of this innovative wedge to increase considerably the set of verbs incorporated suffixoidally as verbal inflexions.

VALENCY SHIFTERS

Reflexive-Reciprocal	*na:y+ [<= pronoun 'self']
CAUsative	*yak= [<= old verb]
PaSsiVe	*yak= [pMijejan: same as CAU]
ASsoCiative	*m@:- [pMijejan: <= *m@:t 'companion']
	*n@- [pSokean];

EVENT MODIFIERS

"MOUTH"	*7aw=
"HEAD"	*ko7= <= *ko+ pak
FORWARD	*win= [pMijejan <= 'face']
UPWARD	*yuk= [<= relational noun]
DOWNWARD	*7ok= [Soke <= 'end']
BACKWARD [Mi]/AWAY [So]	*j@s= <= '*back'
ON SURFACE/BODY	*ni:7= [old root?]

DIRectionals

UP <= *'to go.up'	*=ki7m [pSokean]
DOWN <= *'to go.down'	*=wanak
IN <= *'to go.in'	*=t@k.7@y
OUT <= *'to go.out'	*=put [pSokean]; *=pitz@m [pMijejan]
HITHER <= *'to come'	*=mi:n7
HENCE <= *'to go'	*=n@ks
BACK <= *'to return'	*=witu7 [WSoke]: *=pit(u) [pMijejan]
OVER <= *'to cross.over'	*=jak [pSokean], etc.
PAST <= *'to pass'	*=nas [pM-S]; =k@.tuk [Soke]
ELSEWHERE <= *'to arrive'	*=nu7k [pSokean], etc.
THERE <= *'to stay'	*=tz@7y [pSokean], etc.
<= *'to leave'	*=tzak [pSokean]

PLURAL

3 PLural	*BE.FINISHED [identical with the lexical verb]
----------	---

*-yaj [pSokean]; -suk [WSoke];
*=k@x [pMiJean]

The directionals in the various languages.

	WSoke	ESoke	SOT	TEX	SAY	OLU
UP	yuk= =ki7m	=ki7m	=ki7m	=ke7 (m)	=ku7t	=pet
DOWN	7ok= =wanak	=wanak	=ket	=ket	=manak	=ka7
IN	joj= =t@k.7@y	=t@k.7@y	=ku7m	=ka7 (m)	=t@g@y	=t@kiy
OUT	j@s= =pitz@m	=put	=put	=put	=pichin	=pitz@m
HITHER	=min	?	?	=Beny	=min	=min
HENCE	=n@ks	=n@.manh	=n@kk	=D@k	=n@x	=n@kx
BACK	=witu7	=witu7	=seet	=seet	=pituk	=win.pit
ACROSS	?	=jak	=jak	=jak	?	?
PAST	=k@.tuk	=k@.t@k	=nas	=Das	=nax	=nax
ELSEWHERE	=ye7tz	=nu7k	=ku+nu7k	?	?	=jamat
THERE	=tzak =tz@7y	=tz@7y	=tz@7y	=tzak =tz@7y		=tij

The grammaticalization of these incorporees consists in that not only their functions but also their meanings are different from what they are as independent lexical items; their meanings are in fact sometimes metaphorical; they form a set; and the combination they make with the lexical verb they modify is occasionally lexicalized.

Plural participant marking on verbs

	3pl subj	1/2pl subj
TOT	-tV	-tV
Tla	-k@*hx 'all'	-t@ 'plural'
Col	<-ta>	<-ta>
Coa	-k@x	??
SAY	=g@x obj; -ga7 subj	-ga7
OLU	-k@x	-:t < *jate
SOT	-yaj subj/obj	-ta7m
TEX	-yaj	-ta7
ESo	-yaj subj/obj	-tam
WSo	-suk	---

In pM-S, plural marking on verbs for non-third plural person categories was apparently achieved with the suffix *-ta7m, surviving in Chiapas Soke, Gulf Sokean, and Mije. Oaxaca Soke has plural independent pronouns that it uses when plural subject needs to be expressed. Sayula has innovated a suffix -ga7, of unclear origin, which it also uses for marking third plural subject.

Third plural marking still needs to be sorted out by subject and object, since Sayula at least makes such a distinction. Otherwise the data I have assembled suggests that the subject-object distinction is not relevant [maybe I have overlooked something]. In any case, except for Colonial Mije and Tlahuitoltepec Mije, which apparently use the reflex of pM-S *-ta7m to pluralize third person participants agreement-marked on verbs, all the other languages do this by incorporating into the verb word a verb root that means 'to be finished'. In Mije this is *k@x, in Oaxaca Soke this is **suk**, and in Gulf Sokean and Chiapas Soke this is *yaj. This device is reminiscent of the Tzeltal construction wherein third plural subject of intransitive verbs may be encoded by incorporating the intransitive verb **laj** 'to be finished' in front of the lexical verb. These facts indicate a regional linguistic trait that straddles genetic boundaries.

Among the areal traits that connect Mije-Sokean and Mayan are:

	Mije-Sokean	Mayan
causative based on "to put"	all	Tzeltalan [optional], Chuj, Qanjobalan, Mamean
third plural absolutive	all but Mije "to be finished"	Tzeltal, Yukateko "to be finished" [optional]
dependent intransitive Ergative shift	all	Chuj, Qanjobal lg ar, Mam, Teko
directionals	all: incorporate vi of motion after lexical verb before AM	all but Wasteko, Yukatekan, Qeqchi, Chorti

Traits found in proto-Mije-Sokean logically date to before 1000 BCE. The Mayan analogs are clearly innovations within Mayan. The earliest is probably directionals, dating from between 500 BCE and 600 CE. After that are ergative shift and the use of "put" as a causative former, after about 500 CE. Last is the use of "finish" as a third plural marker, after about 600 CE. Thus, in each case the most plausible explanation would have the innovation originating in Mije-Sokean and spreading to Mayan; cumulatively, the evidence is even more indicative. At the same time, it should be noted that each Mije-Sokean trait, if a borrowing, has a different area of spread within Mayan. It should not be forgotten that, except for ergative shift, these grammatical traits go against the grain in Mije-Sokean. In Mayan, it can be argued that ergative shift and "third plural = finish" go against the grain, but that the periphrastic causative and directionals are structurally congenial. A suspicion arises of a heavy-duty language contact zone at least between some Sokean languages and the more westerly Mayan languages over a period of several centuries.

It may be that the borrowings in script inventory in both directions between EpiOlmecs and Lowland Mayans [reported on briefly in Justeson & Kaufman 1993] relate to these times and places for linguistic cross-fertilization. The word-order shift to predicate-first in Mije-Sokean languages is also more consistent with Mayan than with Oto-Mangean influence, since the noun phrase structures in Mayan and Mije-Sokean are somewhat similar, while between Mije-Sokean and Oto-mangean there is essentially no similarity in the structure of the noun phrase.

This whole topic merits further study.

LATER INNOVATIONS

Bound grammatical markers that are judged to be later innovations by virtue of their incomplete distribution within the family are treated here only insofar as a good reason exists for supposing that they were once separate words. New affixes of unknown origin are NOT assumed in principle to have come from originally independent words.

Around 1000 BCE, proto-Mije-Sokean split up into proto-Sokean and proto-Mijejan.

GRAMMATICALIZATION IN SOKEAN

Various Sokean languages underwent the following grammatical innovations:

POSTVERBAL adverbial/modal ELEMENTS

DESiderative	*-to7 [ESoke, SOT, TEX];
FRuStrative 'wanted to but didn't'	*-t@7p [MAR, SOT] ~ -t@7 [ESoke, TEX]

A frustrative *-t@7(p) was innovated, possibly to proto-Sokean. A desiderative *-to7 was innovated, possibly to proto-Sokean. Both of the two above suffixes suspiciously resemble the auxiliary /t@7/ found in Western Soke. Apparently this verb, not as an auxiliary, and without the linker -A in front of it, was made the source of two different new inflexional suffixes. Together, they point to the existence of a pSokean verb *t@7 vt 'to want'.

ReVerSive 'un(do)'	=pu7 [WSoke]
PeRSeverative 'to stand'	=tenay [WSoke]
PRoBative 'to try'	=ki7ps [WSoke]
HaBiLitative 'can'	=mus [WSoke]
TeRMInative 'to finish'	=tuk [WSoke], =yaj [TEX]
iNCePtive 'to begin'	=7okmanh [WSoke], =ten [TEX]

Western Soke has added a sizable set of incorporated lexical verbs as inflexional modal modifiers, as shown above. In other Sokean languages, the equivalent structures are usually of the AUX + main verb type, and the WSoke incorporees are auxiliary verbs.

Unless these are survivals from pM-S, proto-Sokean has innovated two valency-reducing suffixes, *>An@m 'indefinite subject of intransitives', and *>At@j 'mediopassive'. These look suspiciously like auxiliaries {n@m} and {t@j} after a verb with the linker suffix -A. This is not, however, a correct analysis. Auxiliaries do not affect the valency of the main lexical verb: these suffixes remove an argument. It is not clear how pM-S *n@m vi 'to say' could become an auxiliary or suffix meaning 'indefinite subject'. A verb {t@j} as the putative origin of {>At@j} 'mediopassive' has not been uncovered.

Indefinite, passive and (reflexive-)reciprocal in Sokean.

vi(3) means an intransitive verb that occurs with third person subject only

	proto-Sokean	SOT	COP	MAG	MAR
	*-A-n@H@m	-V...n@:m	-An@m	-n@m	-An@m
added to:	vt/vi/(a)	vt/vi	vt/vi/a	vt/vi	vt/vi/a
resultant class:	vi(3)	vi(3)	vi(us.3)	vi/vt	vi(3)
meanings:	indef subj generalized action	ind sub	gen act weather	ind sub nature	ind sub
	*T-taj	T-taH	T-Ataj	T-t@j	(<u>T-7@m</u>)
added to:	vt	vt	vt	vt	vt
resultant class:	vi	vi	(<u>vt</u>)	vi	vi
meanings:	agentless passive	a-less pass	(<u>mandat</u>)	pass	a-less pass
	*nay-T-A-taj	na-T-V-taH	nay-T-At@j	nay-T-t@j	nay-T-Ataj
added to:	vt	vt	vt	vt	vt
resultant class:	vi	vi	vi	vi	vi
meanings:	reflex- recip	refl- recip	recip	recip	recip

Comment:

Corresponding to Sokean *-taj 'passive' and *nay-VT-Ataj 'reflexive-reciprocal' are Mijejan *-j@H 'antipassive/inverse (or passive)' and *nay-VT-j@H 'reflexive-reciprocal'. These Mijejan parallels suggest that Sokean *-taj and *-Ataj are mutual variants. Mijejan does not seem to have a functional equivalent to Sokean *-An@H@m 'indefinite subject'.

What is the meaning/function/reading of *-(A)taj or *-j@H that allows it to serve to mark the reflexive-reciprocal? *nay- seems to mean 'self reference'. If *nay- were an incorporated noun, that ought to be sufficient to form a reflexive-reciprocal without further derivational morphology. Since additional morphology occurs, perhaps it is adverbial or adjectival. Then *-(A)taj and *-j@H need have no more than some generic intransitivizing (valency-reducing) function. This would relate to the fact that in Sayula -j@ marks some kind of generic antipassive, but is not used in inverse constructions, while in the other Mije languages, *-j@H marks inverse, which could be analyzed as an agentless passive. In Sayula, another use of -j@7 is in combination with preceding -'o*7k 'often, a lot' to mark the desiderative function. The semantic factoring (or combinatory reading: 'to feel a lot like VBing'??) is fairly opaque here. If, however, -'o*7k is the verb 'to die', then the meaning is easy to compute, since the use of 'to die' for 'to have [strong physical or psychic: redundant!!] cravings' is fairly common in MesoAmerican languages.

RELATIONAL CASES ON NOUNS IN SOKEAN

In the Soke group only, a set of relational cases on nouns has been developed:

core cases

+7is ERGative case

+k@ ABSolutive case

peripheral cases

+ko.toy.a7 BENefactive case

+jE7..Vnh SOCiative case

+pi7t INStrumental case

+sV SiMiLative case, 'like'

The extent to which these may be nouns in origin is unclear; however /+ko.toy.a7/ can clearly be seen to be a nominalization of the verb {%toy}, 'to burn, to love', with the derivational prefix {ko.} 'else'.

These morphemes not only relate nouns to verbs but also (except for {+7is}) serve, in some language or other, to subordinate clauses.

*+7k@ 'generic subordinator [MAR]; 'when' [ESoke, EO]; ending of several adverbial words [SOT]

+ko.toy.a7 'in order that' [WSoke]

+jE7..Vnh 'beside the fact that'

+pi7t 'because of the fact that; if [Soke]'

+se 'in the same way that, like/as, how [Soke]'

The locative markers are also clause subordinators:

+ji 'where' [WSoke]

+m@7 'where' [ESoke]

+m@7+ji 'where' [ESoke]

At the end of a clause is where you might expect to find clause subordinators in a radically right-headed language; however, in present-day M-S languages such elements have been observed only in various forms of Soke: these clause-final subordinators often, even typically, co-occur with subordinators at the beginning of the clause. The left-located subordinators are often, though not always, borrowed from Spanish. The new left-located subordinators tend to supplant the native right-located subordinators. Additional right-located subordinators may yet be uncovered in Mijejan languages.

[*+7k@ 'when' is found in EO as well as in the Soke group, and should therefore be reconstructed; it is not clear whether the other cases/subordinators will turn out to have Gulf Sokean or Mijejan cognates.]

PREDICATE IN FIRST POSITION AS A DIFFUSED INNOVATION

EpiOlmec (ca. 150 CE) has strictly OV word order, but intransitive and equational predicates may have the predicate before the subject, especially when the intransitive subject is not an agent [Kaufman & Justeson 1997].

At some point in the history of the M-S family, the fact that in most Mesoamerican languages predicates occupy first position in the clause, a trait found in both its Oto-Mangean and Mayan neighbors, exerted a strong influence from the outside. Except for Western Soke and Mije proper, and somewhat also in OLUta, all M-S languages are now predicate-initial. Even Western Soke and Mije have many verb-initial phenomena, but verb-final syntactic traits persist as well, even a few in OLUta.

The superimposed verb-initial patterns have caused the original auxiliary constructions to become switched around, so that the auxiliary now precedes the main verb.

protoSokean apparently had the following kinds of auxiliary phrase:

[1] AUX₁-AM # (ABS+) ERG+ VERB-w@₂

The auxiliaries used here are

sun DESIDERATIVE ('want')
mus POTENTIAL ('can')
n@ks ANDATIVE ('go')
min VENITIVE ('come')
7oy [~ tij (Soke)] VOLTATIVE ('go&come')
n@mV PROGRESSIVE [Soke only]
7u VETATIVE

This construction requires **ergative shift**, that is, intransitives mark their subject with ERG rather than ABS person markers. SOT, however, for **intransitive** main verbs has AUX-AM # ABS+ VI-i7, which seems to be the pM-S auxiliary structure switched around, with no other change.

This structure, as reconstructed for [1], with **ergative shift and suffixed -w@₂**, is used also when an intransitive verb of movement is subordinated to any preceding lexical verb. [WSoke, SOT].

[2] SUBORD # (ABS+) ERG+ VERB-@

The subordinators used here mean

'perhaps'
'although'
'lest'

This construction requires ergative shift.

The pSokean marker *-@ could be glossed 'irrealis, subjunctive'.

The innovated auxiliary construction did not cause the demise of the original auxiliary construction in Western Soke, but it did make its analysis somewhat opaque, since these phrases are amenable to being analyzed as single verb words, with a linker vowel -A- before the incorporated auxiliary; in MIG, in fact, this linker vowel is dropped, and the vowel in the preceding syllable is lengthened (according to a general rule), thus making the phrasal nature of the construction even more opaque. SOT as well preserves a relic of the original Mije-Sokean auxiliary construction, as noted above. The innovated Sokean auxiliary pattern has not been found in available EpiOlmec texts, but neither has the old Mije-Sokean one.

#####

PROTO-MIJEAN GRAMMATICALIZATION

proto-Mijejan underwent the following grammatical innovations involving incorporation:

POSTVERBAL adverbial ELEMENTS

iNTeNsive	*=ta:k7 [pMijejan] <= *'to weave'
DESideratve	*=wa:n7 [Mije group]

PRE-AM

POTential	*=wa:n7 [pMijejan] <= *'to want'
ALREADY	*-n@H [pMijejan]

Mijejan added *=ta:k7 INTENSIVE to the set of postverbal adverbial elements by incorporating the verb meaning 'to weave' without a linker (NOTE: 'to weave' can mean 'for rain to come down hard' in Sokean languages).

Mijejan created a new aspect-mood-like position immediately before the inherited one, and incorporated *=wa:n7 'to want' as POTENTIAL aspect/mood; Mijejan also incorporated *-n@H 'already', an item of uncertain origin, but probably originally an enclitic, as a PERFECT tense-aspect marker in the new pre-aspect-mood position.

Mije proper incorporated *=wa:n7 DESIDERATIVE into the set of postverbal adverbial elements: OLU keeps the original structure of
ABS-ERG-VERB-LINKER # 7o:k-AM.

VALENCY SHIFTERS

PaSsiVe	*yak= [pMijejan: same as CAU]
ASSoCiative	*m@:- [pMijejan: <= *m@:t 'companion']

Mijejan added a new function, passive, to the preverbally incorporated verb *yak= CAUsative.

The Mijejan ASSOCIATIVE voice, marked by *m@:-, seems to be an incorporation of a reduced form of pMijejan *m@:t 'companion; with'.

Mijejan also shifted the relative positions of the person markers and the negators. If the person markers are inflexional affixes, Mijejan has in effect turned the negators also into inflexional affixes.

The Mijean POTENTIAL

This is an incorporated form of the proto-Mijean verb *wan7 (pre-proto-Mijean *wa:n7) 'to want'; Western Mije replaces this with -ot, whose origin is still unclear; since it does not seem to begin with a laryngeal, it is probably a vowel-initial suffix.

	pMijean	Olu	Say	Coa	Col	Tot
indep	=wan7-pa/-e	-a*m(a)	-a*m- ~	-a*:m(-p(y))		
direct	[-ama]		-a*j#			
WM	-ot-pa/-e			-@-p(y)	#-ob	-up
indep	=wan7-j@H-pa	-an@p	(-am/-aj)	-(w)a:n-@-p		
inverse						
WM	-j@H-ot-pa (Coa, Tot)			-H@::p	??	-jup
	-ot-pa-j@H (Tla)					
dep	=wan7-e	-a*7n(e)	-wa*7n	-(w)a*7ny	#-ot	-@t
direct	[-wa7ne]					
WM	-ot-e				#-ot	-@t
dep	=wan7-e-j@H	-a7nej	(-am/-aj)	??		
inverse						
WM	-j@H-ot-e (Tot)				??	-j@t
	-ot-e-j@H (Tla)	[prob. more original]				

The combinations of the future marker and the Aspect-Mood markers undergo a fair amount of phonological simplification/reduction, even though incorporated *wan7 is stressed.

*-wan7-pa is generally reduced to [a*ma].

*-wan7-e is reduced to [(w)a*7ne].

*-ot-pa is reduced to [-op].

Wichmann [1991ms] saw that the future marker ends underlyingly in /n7/, but did not notice that it begins with /w/, or that is simply the Mijean verb 'to want', incorporated. It can therefore NOT be a proto-Mije-Sokean form. Wichmann corrected his account in Wichmann 1995.103-105, apparently after having this problem brought to his attention by reading TK 1993ms.

What does it all mean?

Other literature.

Kaufman 1963 reconstructs a fair number of the grammatical markers cited here. Wichmann 1993 and 1995 reconstruct several more. The present writing reports on several elements reconstructed by me beginning 1991 and reported in unpublished manuscripts. In a revised version of this paper a careful accounting of the sources of the reconstructions will be provided. Wichmann 1993 discusses several cases of grammaticalization in M-S languages. Mostly they are not the same as the ones discussed or presented here. Wichmann's study is more fine-grained and detailed than the present one, but deals with fewer categories.

Bibliography

- Justeson, John and Terrence Kaufman. 1993. 'A decipherment of EpiOlmec hieroglyphic writing'.
- Kaufman. 1963ms. Mixe-Zoque Diachronic Studies
- Kaufman. 1989ms. 'The prehistory of the Nawa linguistic group'
- Kaufman. 1993ms. Mije-Sokean Comparative Grammar
- Kaufman & Justeson. 1997 in preparation. 'The Epi-Olmec language'. To appear in Encyclopaedia of Ancient Languages, Roger Woodard, editor.
- Wichmann, Søren. 1993. 'Grammaticalization in Mixe-Zoquean languages'. Sprachtypologie und Universalienforschung 46.45-60.
- Wichmann, Søren. 1995. The Relationship among the Mixe-Zoquean Languages of Mexico. Utah. [circulated in ms since 1991]

Project for the Documentation of the Mije-Sokean and Sapotekan languages:
lexical databases and text collections for
 OLUta [Roberto Zavala 1994-1997]
 SAYula [Dennis Holt 1994, Panuncio Isidoro Rafael 1995, Rich Rhodes
1994-present]
 GUIchicovi [George Yonek 1995, Giulia Oliverio 1996]
 TOTontepec [Dan Suslak 1996]
 San MIGuel Chimalapa [TK, Heidi Johnson 1994-1996]
 Santa MARI*a Chimalapa [TK 1994, Loretta O'Connor 1995, Daniel Lo*pez
Za*rate 1995-1996, Randy Rightmire 1996]
 COPainala* [Mary Swift 1995, Cliff Pye 1996]
 AYApa [Jim Fox 1995, 1996]
 TEXistepec [Cathy Berezna 1994, 1995]
 SOTEapan [TK 1993-1994, Valerie Himes 1993-1995]

The research lying behind this paper has been supported since 1993 by the National Geographic Society and the National Science Foundation.

NGS #4190-92 (Sum 1993-Spr 1994)
NGS #5319-94 (Sum 1994-Spr 1995)
NSF SBR-9411247 (Sum 1994-Spr 1995)
NSF SBR-9511713 (Sum 1995-Spr 1998)