APPENDIXES

TESTING THE PROTO-MAYAN-MIJE-SOKEAN HYPOTHESIS

DAVID MORA-MARÍN

[*IJAL*, VOL. 82, NO. 2, APRIL 2016, PP. 125–80]

APPENDIX D

ADDITIONAL LEXICAL FORMS SHOWING SIMILARITY BUT PROBLEMATIC

Most of these forms are problematic due to likely sound symbolism; some are problematic because they are characterized by at least one unusual sound correspondence (e.g., **l* : **m*); a few because of unmatched segments; and several more because they are not semantically isomorphic. Some could prove promising (e.g., D7), others probably not. The footnotes provide the necessary commentary on these forms. Some of the forms in this addendum may be loans. NONE OF THESE FORMS HAVE BEEN INCORPORATED INTO THE ANALYSIS PRESENTED IN THIS PAPER. They are meant to stimulate further research, and also, in some cases, to illustrate the problems that arise when comparing proto-languages of the respective time depths of pMa and pMiSo.

#	pMa	pMa Gloss	Class	pMiSo	pMiSo Gloss	Class
D1	lahɬ’	squeezed, narrow	p, adj	maɬ	to grasp, rub, squeeze[1]	tv
D2	ʔaaq’ or ʔaʔq’	tongue	n	ʔaaka	cheek, edge[2]	n
D3	huul	to arrive here	iv	ʔoy	to leave and come back; to come back and leave	iv
D4	ŋeeh	tail	n	new-e (adj), pOMi *nehw	bent, twisted	adj, iv

[1] The only correspondences of root-initial pMa *l are with pMiSo *m, cf. D1, D11, D12. This is an unusual correspondence, and thus left out of the primary lexical data set. Also, one of the forms, D12, is likely sound symbolic.

[2] These forms could be related: it is plausible to suggest semantic shift due to metonymy (spatial proximity or contiguity, e.g., Spanish *mejilla* ‘cheek’ from Latin *maxilla* ‘upper jaw’). However, given that they are not semantically isomorphic, I prefer to defer their study to a later paper.

				~ *new (vi)		
D5	xul ~ xool	hole	n	hut; pOM *hut	to pierce; hole	tv
D6	šušub' ~ šuub'	to whistle	n	suus(ʔ)	to whistle	v
D7	č'ax	to wash	tv	pSo *¢eʔ	to wash	
D8	haʔb'	year		ʔaame	year	
D9	mič' ~ moč'	to clench, squeeze in hand		miɨʔks	to squeeze, extract	
D10	kab' ~ kaab'	earth, land	n	kama	field, cornfield[3]	n
D11	laq	clay plate, cup		miɨk-i	object made of clay	
D12	leq'	to lick		muuʔk	to suck	
D13	ŋihk	temblar	iv	miʔks	temblar[4]	iv
D14	luqum	worm	n	kumu	worm[5]	n
D15	ʔoxb' (better *ʔoxob'?)	cough	n	pMi *ʔoh(V), pMi *ʔohoʔ (vi)	cough; to cough	n
D16	pCMa *¢'u¢', also pCMa *¢'ub', WASw *tx'utx'b-* (vn) 'kiss, to kiss'	kiss, suckle	tv	¢uʔ¢	suckle, chew[6]	tv
D17	tʸuhb' ~ ¢uhb'	spittle	n	¢uh, ¢uh-i(C)	to spit; spittle	iv; n
D18	pCMa *tʸur	to cut fruit	tv	tuk	to cut fruit[7]	tv

[3] Elson (1992:578) reconstructs the gloss for this etymon as 'field', adding "any cultivated area."

[4] These forms are too similar, in overall shape, not to merit further study. However, the correspondence of pMa **ŋ* to pMiSo **m* is unusual.

[5] These forms are too similar, in overall shape, not to merit further study. However, at this time, the unmatched segments in the Mayan form preclude further scrutiny.

[6] The forms for 'to kiss, suckle' and 'to suckle, chew' are actually consistent with the sound correspondences described on the basis of the primary lexical data set. However, they are obviously onomatopoeic, as are the forms D6, D15, D17.

[7] These forms are similar enough to suggest a relationship of some sort, but the noncorrespondence of their second consonant precludes further scrutiny at this point.

D19	xiq' ~ xaq', xiʔq'	to choke; cough	v	pMi *hiʔkš	to choke[8]	iv
D20	pC *ʔuyux	kinkajou	n	pGSo	chango (monkey)	n
				*ʔuu¢uʔ[9]		
D21	WMa *wayuC	kinkajou	n	wi(ʔ)yuk(s)	kinkajou[10]	n
D22	hoonon (initial w- is attested in several languages from different subgroups)	wasp	n	weeʔnV(k)	wasp[11]	n
D23	LL+WMa *kitaam	peccary	n	pMi *ʔii¢ɨmɨ	peccary[12]	n
D24	siip	tick	n	pOMi *šiiy	tick	n
D25	pCMa *siip GK *siipoŋ	to swell	iv	pSo *siŋ, *siŋ-ɨ	to swell[9]	iv
D26	ʔil	to see	v	ʔis	to see[10]	iv, tv

[9] These forms are similar in shape, but they differ in meaning and are very restricted both within Mayan and within MiSo. They may not be related but are simply similar due to sound symbolism.

[10] This form is restricted to WMa within Mayan and thus not distributed broadly enough.

---

[8] These forms actually conform to the expected sound correspondences, but the form is onomatopoeic.

[9] It is likely that pMa **siip* 'tick' and pCMa **siip* 'hincharse' are related, in fact, the same root. Perhaps pOMi **šiiy* 'tick' and pSo **siw* too are related to one another. What are needed are further examples of a correspondence of pMi **y* : pSo **w*. At least one possible example can be offered: the adjective 'straight' is derived as **tɨy-ɨ* in pOMi from the root **tɨhw* ~ **tɨw*, traced back to pMiSo **tɨw* 'to be straight'. In this instance, a derivational suffix **-ɨ* appears to effect the change of **w* > **y*. Given this process, a parallel pattern can be observed for pSo **siŋ* 'to swell' and pOMi **šiiy* 'tick', especially if pSo **siŋ-ɨ* '(a) swelling' is taken into account. The pSo intransitive root **siŋ* 'to swell' is derived into a noun by means of the derivational suffix **-ɨ*, apparently the same one that derives the pOMi adjective **tɨy-ɨ* 'straight' from pOMi **tɨhw* 'to be straight'. And given that Wichmann's (1995) pSo **ŋ* is really /w/, it stands to reason that pSo **siŋ* (i.e., **siw*) can be traced back to pMiSo **siw* 'to swell'; the pOMi reflex **šiiy* would be the result of a derivational process.

[10] These forms show identity correspondences for the first two segments, in fact, but differ in their final consonant, and there is no recurrence for this correspondence of **l* : **s*. Many pMiSo verbs end in Cs consonant clusters, so many that it would make sense for /s/ in such clusters to be a fossilized suffix, perhaps a derivational suffix; also, the Cs cluster in MiSo does not appear to occur in other grammatical classes as often (if at all), mostly verbs. It is possible that pre-pMiSo could have had a form ***ʔiC$_2$s 'to see', which underwent deletion of *C$_2$, i.e., ***ʔiC$_2$s > **ʔis*. I suspect that these forms are in fact related, but at this moment it would be premature to compare them.

[11] This form is likely onomatopoeic and appears to show irregular correspondences within Mayan; interestingly, however, most of the Greater K'iche'an languages exhibit initial *w*.

[12] As Kaufman (personal communication, 2012) has observed, these two forms for 'peccary', if related, are more likely to be cognates than loans, for a Mayan listener would likely have heard the Mijean **¢* as *¢* (or perhaps *č*), but not as *t*, and vice versa. The reason for restricting them to this data set of problematic forms is their unmatched segments; it is feasible that **kitaam* could be a reanalysis of **k-itaam* 'my peccary', in which case a root **-itaam* would be more easily compared to the Mijean root **ʔii¢ɨmɨ*, but this is not likely.

D27	ʔaʔ¢'aam	salt	n	pMi *taama¢	salt[15]	n
D28	peč	nest	n	peʔn; pMi *peʔn-i 'nest'	to make a nest; nest[16]	iv; n
D29	ʔus	gnat	n	ʔuusu	mosquito	n
D30	kiis	fart	n	¢iis7	fart, to fart[11]	n, iv
D31	ERG-tuuk[-eel]	emphatic (reflexive) pronoun	pron	tum	one[12]	num
D32	kaq	red	adj	¢apa¢	red[13]	adj
D33	kaʔŋ	sky	n	¢ap(-hɨ)	sky[14]	n

[11] This instance, D30, is problematic for two reasons: first, it has been proposed to be POTENTIALLY onomatopoeic; and second, its initial consonant in pMa exhibits variation, given the multiple reconstructions as **kiis* ~ **¢iis* ~ **tiis* (Kaufman 2003:319). The evidence from most subgroups points to **kiis*: Yucatecan, Eastern Mayan, and most of Western Mayan. Within Western Mayan, three languages (Tzotzil, Tzeltal, Popti') attest to the form *¢is*, but given their close proximity to Sokean speakers, it is possible that they could have borrowed it from Sokean. It is also attested in Huastecan as *tiɵ*, which Kaufman (2003:328) points out could be traced to either **tiis* or **¢iis*. Both **kis* 'fishy-smelling' and **tis* 'fart' can be reconstructed to Proto-Ch'olan (Kaufman and Norman 1984:123). Given their narrow distributions, it is possible that the form *tiis and *¢iis could have been innovated much later, long after the breakup of pMa, and then diffused internally, the second one possibly from a Mije-Sokean language.

[12] The numeral one is reconstructed as pMiSo **tum* by Kaufman (2007). There is an unmatched final segment in these forms. Brown and Witkowski (1979) first suggested this comparandum (as a cognate).

[13] Fox (1978) and Brown and Witkowski (1979) both compared these etyma and assumed their cognacy. I did too in an earlier version of this paper. However, there are three sources of uncertainty that apply here: the correspondence of pMa **k* to pMiSo **¢*; the correspondence of pMa **q* to pMiSo **¢*; and the unmatched segments /pa/ of the MiSo form.

[14] Similarly to D32, Fox (1978) and Brown and Witkowski (1979) both compared these etyma and assumed their cognacy, and so did I in an earlier version of this paper. However, there are two sources of uncertainty that apply here: the correspondence of pMa **k* to pMiSo **¢*; and the unmatched segments.

[15] The pMa etymon for 'salt', **ʔaʔɕ'aam*, bears a general resemblance to the pMi etymon for 'salty', **taamaɕ*. However, it is necessary to assume metathesis in either pMa or pMi, as well as extraneous sequences of segments in both pMa *(*ʔaʔ)* and pMi *(*taa)* for a match to result. If these assumptions are made, then we find these correspondences: pMa **ɕ'* : pMi **ɕ*, pMa **aa* : pMi **a*, pMa **m* : pMi **m*. I suspect this coincidence is not merely accidental.

[16] I suspect these two forms for 'nest' are related; however, the unmatched segments preclude their comparison at present.

D34	sut (tv), suht (iv)	to turn over; to go back	tv; iv	pGSo *seet	to return[15]	iv
D35	LL *sakuʔn, pYu *sukuʔn	older brother	n	pOMi *ɕuku; ɕuku-m	aunt; uncle[16]	n
D36	LL *siHs[23]	cold	a	pOMi *šuhš ~ *šuš[24]	to make cold[17]	
D37	mi+	nothing, no one	part	ni+	negation[18]	part
D38	taaʔ	excrement	n	tɨʔn ~ tin	excrement	n
D39	ŋiiʔ	parent-inlaw; sonin-law	n	miʔut[19]	son-in-law (also parent-in-law across both branches)	n
D40	k'am	to receive, to take, to carry	tv	ɕɨm	to carry on the back	tv

APPENDIX E

[15] I suspect the pGSo form is a loan from some Mayan language, but this requires further study. It would have to be from the intransitive derivation, **suht*, but the difference in the vowel is unusual for a loan.

[16] The pOMi form is reconstructed by Wichmann (1995:291), and the Mayan forms by Kaufman with Justeson (2003:104). I suspect diffusion could be at work here, from a Mayan language into pOMi, but the semantics are not transparent. Perhaps the term for older brother was applied to uncle, and semantic shift was based on such use. [23] Reconstructed as a LL term by Kaufman (2003:408) [24] Reconstructed to pOMi by Wichmann (1995:291).

[17] I suspect this term may have diffused, but further analysis is needed.

[18] It is possible that pMiSo **ni+* is really a loan from Spanish *ni*.

[19] For D39, the relevant comparison would require two unmatched segments in the pMiSo form.

POSSIBLE ARCHAIC LOANS OR COGNATES

Next I list a few terms that are widely distributed within Mayan and Mije-Sokean, and could be either archaic loans or cognates. Few of these contain phonology that would point to cognacy as the MORE LIKELY explanation; at best, in some cases the phonology would seem to preclude a specific direction of diffusion (e.g., E1, if a loan, more likely from Mayan to MijeSokean; #171, if loan, also more likely from Mayan into Mije-Sokean; E5, if loan, more likely from Mije-Sokean into Mayan). It is possible that several of my proposed cognates (see Appendixes A and C in the text) could be archaic loans instead.

#	pMa	pMa Gloss	Class	pMiSo	pMiSo Gloss	Class
E1	¢'utux	corn tassel	n	pSo *¢utu	ear of corn[28]	n
E2	ʔam (Huastec ʔaam)	spider	n	ʔaamu or pSo *ʔamu[29]	spider[30]	n
E3	weet	fox	n	pSo *weetu[31]	fox	n
E4	poom	copal incense	n	poomɨ[20]	copal incense	n
E5	pCh *č'əb', GQ'+GMa *č'aʔb'iʔn	vine species (chipilín)	n	¢ɨpʔV	quelite (edible green)[21]	n
E6	ʔiš.i; ʔiš.iʔm[22]	to shell corn; shelled corn	tv, n	ʔɨks; ʔɨks-i[23]	to shell corn; shelled corn[24]	tv, n
E7	WMa+ *¢ak[25]	to grasp, to hunt, to chase after	tv	pSo *¢ɨk	to grasp	tv

[20] Kaufman (2003:1358) reconstructs this Mije-Sokean etymon as **poomɨ* and suggests it diffused into Mayan. Campbell and Kaufman (1976:85) first proposed its origin in Mayan as a loan from Mije-Sokean; Wichmann (1999*a*:306) agrees.

[21] This form is reconstructed by Wichmann (1995:280). It was probably entertained as a possible loan by Kaufman and Norman (1984:118), who listed their Mije-Sokean reconstruction with the Proto-Ch'olan entry but did not offer any explicit comments on the matter. Its Mayan distribution could point to a Proto-Central Mayan reconstruction.

[22] These forms are reconstructed by Kaufman (2003:1034).

[23] These forms are reconstructed by Wichmann (1995:241–42).

[24] I suspect pMa **ʔiš.i* and pMiSo **ʔɨks* are related, possibly through diffusion, from pMiSo into pMa. I assume the final *ks* cluster of pMiSo **ʔɨks* includes a frozen suffix *-s*, so that the lexical root itself would be *ʔɨk-*. It is unlikely that the final *š* of pMa **ʔiš.i* could be related to the final *k* of pMiSo **ʔɨk*. Instead, it seeems more likely that a Mayan speaker would have heard a cluster *ks* as *š* in final position. However, for the moment, I prefer to include it in the set for further analysis.

[25] Kaufman (2003:903) does not gloss this reconstruction but does provide glosses such as 'agarrar, cazar' and 'to chase after' for the cognates. The form is attested mostly in Western Mayan languages but also in Q'eqchi'; however, given the significant contact between Q'eqchi' and Lowland Mayan languages, it is quite possible that this term represents a loan into Q'eqchi'

[28] Kaufman (2003:1153) only remarks "cf. Soke tzutu," without explicitly suggesting either cognacy or diffusion. The disyllabic form could point to a Mije-Sokean origin; however, the phonology itself does not preclude a Mayan origin.

[29] Wichmann (1995:248) reconstructs it to Proto-Sokean.

[30] Kaufman (2003:698) remarks "[?diffused from Mije-Sokean *7a:mu]."

[31] Kaufman (2003:567) reconstructs this Sokean etymon as **weetu* and suggests it diffused into Mayan. Campbell and Kaufman (1976:86) originally proposed its diffusion into Mayan and were followed in this assessment by Justeson et al. (1985:23, table 11).

APPENDIX F

LIKELY LOANS BETWEEN MAYAN AND MIJE-SOKEAN

Most of the following forms are likely loans from a Mayan language into a Mije-Sokean language or vice versa.[26] A few are still unclear. Several (F63–F76, F79, and F80) are proposed here for the first time (see "This paper" under "Proposals and Discussion" column). A very few MAY be actual cognates. This list is meant primarily as a compilation of forms appearing in previous literature on the matter, to be used as a tool for testing a variety of hypotheses. IT IS NOT INTENDED AS A CONCLUSIVE AND DEFINITIVE LIST OF LOANS. All of the forms cited in literature from the 1960s through the 1990s have been checked against Kaufman (2003) and Wichmann (1995).

#	Ma	MiSo	Proposals and Discussion	Evaluation
F1	Colonial Yucatec calendric day name <oc> 'dog'	pMiSo *ʔuku 'agouti', pMi *ʔuku 'agouti', pOMi *ʔuk 'dog'	Campbell and Kaufman (1976:85), Justeson et al. (1985:23), Wichmann (1999*a*:306)	p(O)Mi > Yucatec (and Ch'olan) given semantic match
F2	Mayan *(h)iʔ(i)š 'fourteenth day (jaguar)'	pMi *ʔiiš'weasel', Northern Highland Mije 'cacomixtle (wild feline)'	Justeson et al. (1985:24), Wichmann (1999*a*:312)	p(O)Mi > Lowland Mayan/Classic Mayan

from Ch'olan. Otherwise, it would be feasible for the term to be of pCMa ancestry. It is possible that this term could be an archaic loan, although the direction of diffusion would not be obvious.

[26] ISO codes for the languages in the Mayan and Mixe-Zoquean families can be found at http://www.ethnologue.com/subgroups/mayan-0 and http://www.ethnologue.com/subgroups/mixe-zoquean-0, respectively.

F3	Yucatec *čuwen*, K'iche' *čowen* 'eleventh calendric day name (monkey)' (Ch'ol, Yucatec, K'iche')	pMiSo *¢aawi 'monkey' (Wichmann 1999:312), pMiSo *¢aawiʔ 'monkey' (Kaufman 2007)	Justeson et al. (1985:24), Wichmann (1999*a*:312)	Some MiSo > Lowland Mayan/Classic Mayan/K'iche'
F4	K'iche'an Proper *pus 'partir/cortar (con instrumento)'	pMi *puš 'to cut with a machete', pMi *puš-an 'axe'	Campbell and Kaufman (1976:85) and Kaufman (2003:877); Wichmann (1999*a*:307) favors Proto-Popolocan *pos 'hard stone'	pMi > K'iche'an

F5	Chiapas sphere (Tzo, Tze, Chj, Pop, Mch, Tuz) #tunuk' 'turkey'; native Mayan form is *ʔak'[27]	pSo *tuʔnuk 'turkey'	Campbell and Kaufman (1976:86), Kaufman (2003:631), Wichmann (1999*a*:308)	pSo > Chiapas sphere
F6	GLL or possibly WMa (Yu, Mop, Cht, Chl, Tze, Mch) *pehɉ' 'trampa (de piedras)'; Greater Mamean #patę 'trampa'	pMiSo *paɉ 'to squash', Cop. Soke paɉ 'to trap, to flatten', ko-paɉ-e 'trap'	Campbell and Kaufman (1976:86), Kaufman with Justeson (2003:932); Kaufman (2001-2009) is not certain about direction of diffusion; Wichmann (1999*a*:309) disqualifies this form as a loan	Some MiSo (given reconstruction of root to pMiSo) > Greater Mamean and WMa/GLL

[27] It is conceivable that *#tunuk'* (or *#tuluk'*, given Tzo and Tze data) could be a compound form incorporating Mayan **ʔak'* and having undergone vowel assimilation, e.g., *tun=ʔak'* > *tunak'* > *tunuk'*. If so, the hypothetical *tun=* could correspond to the form *tun* used with the names of male animals, e.g., Greater K'iche'an **tun ty'ok* 'male grackle' (Kaufman 2003:617), possibly based on a form *tuun* 'hermano mayor, tío (older brother, uncle)' attested in Mocho'; in fact, the Mocho' form of *#tunuk'* exhibits vowel length, *tuunuk'*, just like the term *tuun* for 'older brother, uncle'. The problem with this analysis is that the term *#tunuk'* refers to 'turkey' in general, not 'male turkey', and in fact, the Tuzantek form, *tunuk*, means 'turkey hen'. Nevertheless, if an earlier form *tuun=ʔak'* were to stand to further testing, the pSo form would be a loan from Mayan, not the other way around.

F7	Chiapas sphere (Tze, Toj) *¢am ‘bonito/bueno’	Soke *¢amɨ ‘fully grown’ (Kaufman with Justeson 2003:142); pMiSo *¢aamʔ ‘to ripen, grow’, pMi *¢aamʔ ‘ripe’ (Wichmann 1995:282)	Campbell and Kaufman (1976:86), Kaufman (2003:142); Wichmann (1999*a*:310) disputes this etymon as a loan	MiSo > Chiapas sphere
F8	GQ’ *ʔason ‘cloud’ (unMayan structure, limited distribution within Mayan)	Chiapas Soke *nas=ʔoʔna ‘fog’ (< *nas ‘earth’ + *ʔoʔna ‘cloud’); pMiSo *nas ‘earth’	Campbell and Kaufman (1976:86), Kaufman (2003:479); Wichmann (1999*a*:310) disputes this etymon as a loan	Chiapas Soke > GQ’(?)

F9	Huastec *koy* ‘rabbit’	pMiSo *koya ‘rabbit’	Campbell and Kaufman (1976:86), Wichmann (1999*a*:311)	some MiSo language > Huastec
F10	Greater K’iche’an *ʔa¢ ‘same sex elder sibling’, Greater Mamean *ʔa¢ik ‘same sex elder sibling’	Sokean *ʔa¢i(ʔ) ‘same sex elder sibling’ (Kaufman 2003:103); pMiSo *ʔa(h)¢i ‘elder brother’ (Wichmann 1995:243)	(Campbell and Kaufman 1976:87), Kaufman (2003:103), Wichmann (1999*a*:312)	(p)MiSo language > pEMa (or pGK’, pGM)
F11	LL *huuj and EMa+ *ʔoohon (Kaufman 2003:642)	Mije <juun> ‘iguana (day name)’	Campbell and Kaufman (1976:87), Justeson et al. (1985:73), Wichmann (1999*a*:312)	Some Mijean language (possibly pMi according to Wichmann 1999:312) > LL and EMa+
F12	K’iche’ *sak* ‘coral bean tree’, Kaqchikel <sac> ‘dados; habas con que juegan los indios dados’ (Varela 1600) (Kaufman 2001–2009)	pMiSo *sɨk ‘bean’	Kaufman (20012009)	Some MiSo > K’iche’an

F13	Tzotzil *ʔowen* (Kaufman 2001–2009)	pSo *ʔow, CopSo *ʔowiʔ* ‘avocado (aguacate de tierra fría)’ (Kaufman 2001–2009)	Kaufman (2001–2009)	Some So > Tzo
F14	Q’eqchi’ *koyow* ‘avocado species’ (large) (Kaufman 2001–2009), Mopan *koyow* ‘wild avocado’ (Hofling 2011:245)	Soke *koyonh: Chiapas Soke koyonh ‘avocado species’ (chinín/chinini, aguacate chupón, aguacate de mono)	Kaufman (2001–2009)	Some So > Q’eq
F15	Yucatec <cumku> #kumk’u name of eighteenth month (semantically corresponding to Mije month #1, “big town”) (Kaufman 2001–2009)	Proto-Soke *kumkuyʔ ‘town’	Kaufman (2001–2009)	(p)So > Yu
F16	EMa+GQ’ #kaayu(m)paʔ	SmiChimalapa Soke keyo-pa ‘it	Kaufman (20012009)	So > EMa

	‘relámpago’ (K’iche’ koyopaʔ ‘lightning’)	shines, it flashes’ (Kaufman 2001–2009)		
F17	WMa *how ‘loco, enojo, cólera, espantaje, lobo, coyote, guerrero, soldado’; pTz *how ‘crazy, rabid’	pSo *how-ɨ ‘crazy, stupid’ ← pSo *honh ‘to become deranged’ (Kaufman 2001–2009), Soke *hoʔw.eʔ	Kaufman (2003:148), Kaufman (2001–2009)	(p)So > (some) WMa
F18	LL+ *yuum ‘padre, sacerdote Maya, Dios, sol, amo, dueño, señor’	From Sokean *yumi ‘master, owner’[28]	Kaufman (2003:110); Wichmann (1999*a*:313) has different Sokean reconstruction	So > LL

[28] Kaufman and Justeson (2004:1080) argue that Yucatecan borrowed the term from prepSo **yuumi*, prior to the loss of contrastive vowel length in Sokean. Therefore, it would have

			(*ʔoomi) but agrees with loan analysis	
F19	POQ *poh ‘month, moon’	From MiSo *poyʔa	Kaufman (2003:502)	pMiSo > Poq
F20	EMa *šamal ‘fuego’	pMiSo *sam ‘to heat’	Kaufman with Justeson (2003:511)	pMiSo > EMa
F21	WMa #peč ‘tipo de perdiz; pato’ (diffused)	Borrowed into Soke, Chiapaneko	Kaufman (2003:617)	
F22	EMa *wah ‘tortilla’	pMiSo *way ‘moler/to grind’, pMiSo *way-e ‘molido/ground’	Campbell and Kaufman (1976:85), Kaufman (2003:1198)	pMiSo > EMa
F23	WMa+LL *xuč’ ‘moler’ (if not loan, then LpMa)	pMi *hɨ¢ ‘to grind’ and *hɨ¢-i ‘dough’	Campbell and Kaufman (1976:85), Kaufman (2003:1060), Wichmann (1999*a*:305)	pMi > WMa+LL; Wichmann (1999*a*:305) argues the reverse, but he did so assuming the Mayan form was pMa, not a contact region (WMa+LL)
F24	K’iche’, Kaqchikel, Q’eqchi’ *may* ‘twenty, twenty years (of 400 days each)’ and Kaqchikel *may q’iix* ‘the calendar’, Q’eqchi’ *mayex* ‘sacrifice, offering’	pMiSo *may ‘to count’, but Wichmann (1999*a*:306) prefers the homonym pMiSo *may ‘much, a lot’ for the comparison	Campbell and Kaufman (1976:85)	Some MiSo > GK’

been a relatively early loan into Yucatecan. However, it is also possible that the term was originally **yumi*, and that Yucatecans could have borrowed it as **yuum*.

F25	pMa #kakaw (diffused in Mayan languages, not actually pMa)	pMiSo *kakawa ‘cacao’	Campbell and Kaufman (1976); Justeson et al. (1985:23, table 11), Wichmann (1999*a*:300–302), Kaufman (2007)	(p)MiSo > (p)Ma (early loan)
F26	pMa *¢ima(ʔ) ‘gourd’	pMiSo *¢ima ‘gourd’	Campbell and Kaufman (1976:84), Wichmann (1999*a*:302), Kaufman (2003:993)	(p)MiSo > (p)Ma (very early loan)
F27	Huastec ¢iw, pEMa *č’imaah, Ch’orti’ č’iwan	pMi *¢iʔwa ‘ayote (squash)’	Campbell and Kaufman (1976:84), Justeson et al. (1985:24), Wichmann (1999*a*:303), Kaufman (2003:1129),	pMi > Huastec, Eastern Mayan, Ch’orti’
F28	EMa+ *ʔiš-kooyaaʔ ‘tomato’	pSo *koya ‘tomato’ (Wichmann 1995:360)	Campbell and Kaufman (1976:84), Justeson et al. (1985:23, table 10), Wichmann (1999*a*:303), Kaufman (2003:1133)	p(Mi)So (or some later Sokean) > EMa and Ch’ol
F29	GQ’ *saʔm, GLL *sam.eht ‘comal (griddle)’	pMiSo *sam ‘calentarse/to heat’ (Wichmann 1995:449)	Campbell and Kaufman (1976:86), Kaufman (2003:925)	Some MiSo > WMa or > WMa, GLL
F30	WMa *ʔumaaʔ ‘deaf’	pMiSo *ʔuuma ‘mudo (mute)’, pMiSo *ʔuumaʔ (Kaufman 2003:728)	Campbell and Kaufman (1976:86), Justeson et al. (1985:23, table 11), Kaufman (2003:728)	pMiSo > pWMa (relatively early loan)

F31	WMa *waʔš ‘fox’	pOMi *waa[h]š‘fox’	Campbell and Kaufman (1976:86), Justeson et al. (1985:23, table 11), Kaufman (2003:568)	possibly from some WMa language > pOMi prior to ca. 1000 C.E. (pOMi is of much shallower time depth than WMa)
F32	Proto-Ch’olan *¢uk ‘mouse’	pMiSo *¢uuk ‘rat, mouse’ (Wichmann 1995:290)	Campbell and Kaufman (1976:86), Kaufman and Norman (1984:133), Justeson et al. (1985:23, table 10)	Some MiSo language > (Western) Ch’olan
F33	Tzotzil poket ‘widemouthed, large handless bowl’, Chol pok’ ‘gourd bowl’	pMiSo *pok(ok) ‘gourd’ (Wichmann 1995:433)	Campbell and Kaufman (1976:87), Justeson et al. (1985:27, table 14)	(Mi)So > Ch’ol and Tzotzil
F34	LL *xaʔan ‘man’s brother-in-law’	pMi *hɨyɨ ‘man’s brother-in-law’	Justeson et al. (1985:23, table 10)	Unlikely in present author’s opinion
F35	pYu *čiʔk ‘pizote (coati)’	pMiSo *¢iku ‘coati’	Justeson et al. (1985:23, table 10)	“from protoMije-Sokean *tziku” (Kaufman 2003:581)
F36	Yucatecan, Eastern Ch’olan *sahp ‘armspan, fathom’ (Justeson et al. 1985:23, Table 10)	pMiSo *sah ‘armful’, pSo *sah(a) ‘wing’, Cacalotepec šAhpk, Sayula Popoluca šáhpak (Wichmann 1995:448)	Justeson et al. (1985:23, table 10)	Possibly from (p)So > Yu and Eastern Ch’olan
F37	pYu *mač ‘to take, grasp’	maɕ ‘to grasp’	Justeson et al. (1985:23, table 10)	pMi > pYu, unless pMi speakers heard pYu *č as *ts
F38	Mopan šaak-t ‘to examine’	Totontepec Mixe šaak ‘to guard’	Justeson et al. (1985:23, table 10)	Perhaps ToMi > Mopan (given use of -t ‘transitivizer’)

F39	Mopan suʔuk ‘grass’, Yucatec suʔuk ‘grass’, Itzaj suʔuk ‘grass’	pSo *soʔk ‘grass’	Justeson et al. (1985:23, table 10),	pSo > pYu
			Kaufman (2003:1155)	
F40	Mopan sum ‘rope’, Kaufman with Justeson (2003:965) reconstruct pYu #sum (Itzaj, Mopan); also present in Yucatec as sùum ‘rope’	pMi *ȼum ‘to tie’	Justeson et al. (1985:23, table 10), Kaufman (2003:965)	pMi > pYu
F41	pYu #tuts ‘corozo’ (Itzaj, Mopan) Kaufman (2003:1119)	pMiSo *tuuȼ=kuy ‘palm tree’ (Wichmann 1995:477)	Justeson et al. (1985:23, table 10), Kaufman (2003:1119)	Some MiSo language > pYu
F42	Yucatec makeʔč ‘escarabajo, camaleón’	Sayula Popoluca mageʔč ‘cockroach’ (Justeson et al. 1985:23, table 10), pMiSo *makoko ‘cockroach’ (Wichmann 1995:374)	Justeson et al. (1985:23, table 10)	SaP > Yucatec
F43	Yucatec *mok* ‘gata de mar’	Sayula Popoluca múguy ‘topote’	Justeson et al. (1985:23, table 10)	Unable to confirm data
F44	GK’ *ʔiš+paʔač, Yucatec *-pač* ‘male iguana’	pMiSo *paȼi ‘lizard’ (Wichmann 1995:418)	Justeson et al. (1985:23, table 10), Kaufman (2003:641)	pMiSo (or several MiSo languages) > several Mayan languages
F45	Yucatec *sok* ‘snail’	pSo *soki ‘snail’ (Wichmann 1995:456)	Justeson et al. (1985:23, table 10)	pSo (or some Sokean language) > Yucatec
F46	Ch’ol *meʔ* ‘deer’	pSo *mɨʔa ‘deer’ (Wichmann 1995:367)	Justeson et al. (1985:23, table 10)	Some Sokean language > Ch’ol

F47	GLL *kuket ‘trunk (of body)’	pMiSo *kuk ‘middle’ (Wichmann 1995:350)	Justeson et al. (1985:23, Table 11), Kaufman with Justeson (2003:303)	Some MiSo > GLL
F48	GLL *qo(h)m ‘short’	pMiSo *kono ‘short’	Justeson et al. (1985:23, table 11), Kaufman with Justeson (2003:1375)	pMiSo > GLL

F49	pMa *páta ‘guayaba’, contrast with native pMa *(ʔi)kaq’ ‘guayaba’	pSo *patahaC (Kaufman with Justeson 2003:1101), Sierra Popoluca pátaŋ; pSo *pataha? (Kaufman and Justeson 2008)	Campbell and Kaufman (1976:84), Justeson et al. (1985:23, table 11), Wichmann (1999*a*:304), Kaufman (2003:1101)	pre-pSo > (p)Ma (early loan)
F50	LL+WMa *ʔune... ‘child, baby, boy, etc.’	pMiSo *ʔunV(k) ‘baby’, pSo *ʔune, pMi *ʔunak	Campbell and Kaufman (1976:86), Justeson et al. (1985:23, table 11), Wichmann (1999*a*:310), Kaufman (2003:117)	pSo > Lowland Mayan+WMa
F51	“highland Mayan day name to:x (and the Quiche deity name <Tohil>)” (Justeson et al. 1985:24)	pMiSo *tuuh ‘rain’ (Wichmann 1995:478)	Justeson et al. (1985:24)	Some MiSo language > K’iche’
F52	Yu ¢apaʔ ‘dwarf’ (Justeson et al. 1985:25)	“cf. Sayula Popoluca ¢apa ‘short’” (Justeson et al. 1985:25)	Justeson et al. (1985:25), also make a comparison with Nahua	Could not confirm data
F53	pMa *¢i¢ ‘aguacatillo (nectandra)’ (Nectandra spp.)	pOMi *šihts ‘aguacatillo’ (Wichmann 1995:438)	Justeson et al. (1985:28, table 15), Kaufman (2003:1112)	(p)Ma > pOMi

F54	Yucatec (h+)yùuk ‘fawn’ (Bricker 1998:318), Itzaj yuk ‘brocket deer’ (Hofling 1997:696)	pMi *yuuhk ‘forest’ (Wichmann 1995:438)	Justeson et al. (1985:28, table 15)	Semantic connection is perhaps too indirect in my opinion, but plausible; not very strong candidate at this point
F55	WMa *ʔamay ‘flute, whistle’	MAR ʔamaʔy ‘carrizo; flauta’	Kaufman (2003:753)	Sokean > WMa
F56	Chiapas sphere (Tzo, Toj) *nutiʔ ‘net’	“from Soke”	Kaufman (2003:976)	Soke > Chiapas sphere
F57	Mam /naayaj/ ‘izquierda (mano)’	Soke *ʔaŋ-nahiʔ	Kaufman (2003:246)	Soke > Mam

F58	Mochó, Tuzanteco *pots* ‘to get tired’	Oaxaca Soke *pots ‘to get tired’	Kaufman (2001–2009)	Oaxaca Soke > Mochó, Tuzanteco
F59	CT *k’el ‘to see, look’[29]	pSo *ken ‘to be seen, appear, look’	Kaufman (2001–2009); I am not convinced by this example	(p)So > CT
F60	Chamula Tzotzil winah ‘to appear’ > winah.el ‘sky’, San Bartolo Tzotzil winah ‘to resemble, to appear as’	pSo *win.ʔah ‘to get out in front’ < pMiSo *win ‘face, front, surface, eye, body, self’	Kaufman (20012009)	(p)So > CT
F61	pYu #k’ek’en ‘pig (shift since Spanish introduction of pigs)’ (Kaufman 2003:564)[30]	pMi *keem(A) (?) ‘pig’ (originally ‘peccary’)	Kaufman (2001–2009); I am not convinced by this example	pMi > Yu

[29] It seems to me unlikely that Mayan speakers would have heard Sokean *ken* as *k’el*.

[30] Bricker (1998:151) analyzes *k’éek’en* ‘pig’ as based on *k’en*, an onomatopoeic root. However, reduplicated forms based on roots of the shape CVC typically exhibit a plain vowel or a vowel with low tone, not high tone, in the first, reduplicated syllable, e.g., *sasak* or *sàasak* ‘very white’.

F62	pK' *čakač 'basket' (Kaufman 2003:979)	pMi *kaȼE 'chiquihuite, cesta, canasta sin asa, pecho' (basket, chest) (Kaufman 2001–2009), pOMi *kač 'basket' (Wichmann 1995:340)	Kaufman (2001–2009)	pMi > pK'
F63	CT *čič 'older sister; female relative', Yucatecan *kiik 'man's (older) sister' (Kaufman with Justeson 2003:108)	pMi *ȼiʔ 'older sister, aunt' pSo *ȼiȼi 'older sister, aunt' (Wichmann 1995:276)	This paper	Possibly CT *čiič > So (but not clear)
F64	GLL *muq' 'strength' (Kaufman 2003:1391)	pMi *mikik 'strength' (Wichmann 1995:368)	This paper	pMi > GLL
F65	pMa *nooq' 'cotton; clothes' (Kaufman 2003:1015)	pGSo *nokkoy 'pantalón (pants)' (Wichmann 1995:403),	This paper	Ma > pGSo
		anachronistic definition due to semantic shift		
F66	LL+WMa (diffused) *kitaam 'peccary' (Kaufman 2003:563)	pMi *ʔiitsimi 'peccary' (Kaufman 2001–2009), pOMi *ʔitsimi (Wichmann 1995:235)	This paper	pMi > LL
F67	pYu *maan 'to pass (by/across)' (Kaufman 2003:1310)	pSo *ma[ŋ] 'to go' (Wichmann 1995:378)	This paper	pSo > pYu

F68	WMa (pCh, Mch) *xuy ‘to stir’ (Kaufman 2003:811), pYu *xuhy ‘to stir’, hence possibly LpMa (and if so, perhaps cognate candidate)	pMi *hoʔy ‘to whip’ (Wichmann 1995:327)	This paper	pMi > LL+Mocho’ (if loan; could be cognate)
F69	LL *čuy ‘to sew’ (Kaufman 2003:1010)	pMiSo *suy ‘to sew, fish with hook’ (Wichmann 1995:454)	This paper	pMiSo > LL
F70	Yucatec lèek ‘frog’	GuMi lek ‘frog’	This paper	Form is onomatopoeic in both languages; unclear direction of diffusion
F71	pYu #jok’oʔ cheʔ ‘hook’, presumably /j/ on basis of Q’eqchi’ form (Kaufman 2003:809)	pOMi *hoʔk-n ‘hook’, probably based on pMi(So) *hoʔk (tv) ‘to hook, to pick fruit with a hook’ (Wichmann 1995:323)[31]	This paper	Likely loan, perhaps p(O)Mi > pYu, given pMi(So) etymology of the root *hoʔk

F72	pMa *q’iiŋ ‘sun, day’, EMa *q’iix ‘sun, day’, Huastecan *k’ij[32] (Kaufman 2003:461)	pMiSo *kih ‘to give light’ (Wichmann 1995:328)	This paper; semantics would require significant shift during process of diffusion and is thus somewhat questionable	Possibly EMa > pMiSo

[31] Wichmann’s (1995:323) ‘esperar/to wait’ gloss is problematic: first, none of the Mijean varieties he lists bears such a meaning, showing instead ‘enganchar (to hook)’, ‘to pick fruit with a hook, to pull with a hook; to form a tortilla in the hands (the first stage)’and, second, none of the Sokean entries is provided with a gloss.

[32] The Huastecan form **k’ij* is shown here with <j> instead of <x> or <h> to indicate that Huastecan does not distinguish /x/ from /h/.

F73	WMa *pukuux ‘diablo’ (Kaufman 2003:1366)	pMiSo *pok ‘hacer brujería’ (Wichmann 1995:433)	This paper	pMiSo > WMa
F74	EMa *muuŋ ‘nube, sombra’ (Kaufman with Justeson 2003:475)	pSo *[muŋ] ‘sombra de persona’ (Wichmann 1995:384)	This paper	pSo > EMa
F75	GK’ *tiix ‘oración, rezo’ (Kaufman 2003:181)	pOMi *tih ‘decir’ (Wichmann 1995:459)	This paper	Perhaps GK’ > pOMi
F76	Ch’orti’ kisiš ‘chinche’ (Kaufman and Norman 1984:123)	pMiSo *¢isi(k) ‘chinche, bedbug’ (Wichmann 1995:271)	This paper	MiSo > Ch’orti’, if loan; could be sound symbolic and independently innovated
F77	Hua ʔokow ‘wasp’s nest’, Tze ʔáko ‘wasp’s nest’, Toj ʔáko ‘wasp’s nest’ (Kaufman 2003:675); cf. pCMa *ʔahqaanh (not loan)	pSo *ʔokɨwVC ‘wasp’s nest’ (Kaufman 2003:675)	Kaufman (2003:675); Wichmann (1999a:308) disputes this form as loan	pSo > Mayan (Hua, Tze, Tzo)
F78	pMa *b’uhq ‘clothes’ (Kaufman 2003:1014)	pGSo *puuki ‘cotton’ (Wichmann 1995:428)	Brown and Witkowski (1979:38)	(p)Ma > (p)(G)So
F79	Qp (Q’anjob’alan Proper) *ŋik’ ‘arrancar (to pull out)’ (Akatek, Popti’) (Kaufman 2003:821)	pMiSo *nuk (vt) ‘arrancar/to pull out’ (Wichmann 1995:400)	This paper	(p)MiSo > Qp (given distribution within Mayan)

F80	LL #pujuy (Kaufman with Justeson 2003:687), Yucatec *puʔjuy* 'nighthawk', Ch'olti' <puhuy> 'pájaro nocturno', Tuzantek pujuy 'tapacamino', Ch'ol *xpujyu* 'tapacaminos (nighthawk/roadrunner)' (Aulie and Aulie 1998:152, 220, 269)	pMiSo *puʔhuyu 'tapacaminos' (Wichmann 1995:428)	This paper	(p)MiSo > LL+
F81	Ch'ol and Chontal -b'ä 'relativizer' (Zavala 2007; Kaufman and Justeson 2009)	Sokean *-pʉ(ʔ) 'relativizer'	Zavala (2007)	Some So > Western Ch'olan

APPENDIX G

REVIEW OF PREVIOUS APPROACHES TO MAYAN AND MIJE-SOKEAN COMPARISONS

Next I offer some details on the approaches by Fox (1978), Brown and Witkowski (1979), and Kaufman (1964*a*), as well as the criticisms by Campbell and Kaufman (1980; 1983).

G.1. Fox (1978). Fox (1978) compiled 42 comparanda for Mayan and Mije-Sokean, consisting of data from a few, selected individual languages, but in some cases a single language, in the case of Mije-Sokean; his data for Mayan were comprehensive. Fox (1978:90–92) utilized these data, specifically a subset of 20 comparanda, to argue that Mayan *k/k'/q/q'* generally correspond to Mije-Sokean *¢* or *k*, to support his hypothesis for the presence of a velar/labiovelar stop contrast instead of a velar/uvular stop contrast in Proto-Mayan. He did not attempt to account for the regular correspondences through the formulation of sound changes, nor did he attempt to reconstruct the lexemes he compared. The main problem with the data and methodology utilized by Fox (1978) consists of the narrow distribution within Mije-Sokean characteristic of some of the sets compared. There were also a few instances of onomatopoeia and pan-Americanisms and panMesoamericanisms, which he did not isolate as potentially problematic. Out of Fox's 42 original comparanda, 32 remained potentially viable and constituted the basis of original set for the present effort; several of these original comparanda have been culled since then.

G.2. Brown and Witkowski (1979). Not long after, Brown and Witkowski (1979) focused on the same types of correspondences also highlighted by Fox (1978), but they further proposed reconstructions and sound changes to account for them, using a data set of 62

comparanda. The following sound changes constitute their proposed developments for MijeSokean (their "Zoquean"): **k/*k'* > MiSo *k*/#__V[+round, +back], **k/*k'* > MiSo *ts*/#__V[-round], **q/*q'* > MiSo *k*. However, their paper was soon critiqued, heavily and extensively, by Campbell and Kaufman (1980); Witkowski and Brown (1981) followed with a rebuttal, after which Campbell and Kaufman (1983) put forth a more comprehensive and definitive response.

G.3. Campbell and Kaufman (1980; 1983). These critiques outlined two types of problems: one type of problem pertains to the methodological framework, the other to the compilation of the data set. Regarding the former, Campbell and Kaufman (1980; 1983) identify three methodological problems with the approach by Brown and Witkowski (1979): no discussion of the respective Proto-Mayan and Proto-Mije-Sokean sound systems; no description of the assumed subgrouping models for Mayan or Mije-Sokean; and no outline of the reconstruction methodology. Regarding the latter, Campbell and Kaufman (1980; 1983) point to problems with most of the 62 cognate sets proposed by Brown and Witkowski (1979), such as: the adjustment of Mije-Sokean forms to the phonotactic structure, phonetic shape, and semantics of Mayan forms; the lack of semantic matches in roughly half of the data set (30/62); etyma exhibiting monoconsonantal correspondences (9/62); several onomatopoeic examples (14/62); several cases of erroneous analysis, particularly of the morphological makeup of the forms (12/62); several cases of erroneous analysis of cognacy within Mayan or within Mije-Sokean (8/62); several cases of "widespread forms" or "pan-americanisms" (16/62); a significant number of cases that are not sufficiently distributed within Mayan to constitute Proto-Mayan etyma (29/62), or within MijeSokean to constitute proto-Mije-Sokean etyma (30/62); and several instances of possible loanwords between Mayan and Mije-Sokean (5/62). After applying the necessary corrections based on the more accurate and complete data sets by Wichmann (1995) and Kaufman (2003), and excluding almost all of the sets refuted by Campbell and Kaufman (1980; 1983), only 16 of the 62 sets in Brown and Witkowski (1979) remained potentially viable; these were added to the 32 sets from Fox (1978).

The total number of comparanda for the earliest stage of this project, consequently, was 48 sets, several of which have been discarded since. In fact, during the entire period of time during which the present project has evolved, approximately 180 items have been considered, more than half of which have been discarded.

G.4. Kaufman (1964*a*; 1964*b*). Kaufman (1964*a*:131–35) discusses the problem of distinguishing loans from cognates in the comparison of Mayan and Mije-Sokean languages, assuming the correctness of the Macro-Mayan hypothesis by McQuown (1942; 1956). Specifically, Kaufman suggests that some phonologically similar forms could be cognates, including among these the terms for 'spider', 'fox', 'ripe', 'fly', and 'incense', several of which included in Appendixes E and F. He also proposed a set of 13 loans, most of which have stood further testing by that author and several of his colleagues (Cambell and Kaufman 1976, Justeson et al. 1985, and Kaufman 2001–2009; 2003; 2007), and are included in Appendix F.

Kaufman (1964*b*) carried out a test of the Macro-Mayan hypothesis, which would relate Mayan, Mije-Sokean, and Totonakan. This work has been cited by Campbell and Kaufman (1980:852), who state that "the hypothesis has not proven productive," and Campbell (1997:165), who adds that "the evidence presented to date is inadequate to support it, though sufficient to

suggest that the proposal merits further investigation" and concludes "Perhaps the main problem in this case is to distinguish possibly inherited similarities among these languages from diffused traits widespread in the Mesoamerican linguistic area." To my knowledge no details of Kaufman's (1964*b*) investigation have been published.

APPENDIX H

BASIC CRITERIA FOR COMPARISON OF LEXICAL ITEMS

In order to minimize the risk of chance similarities as well as the risk of inadvertently sampling loanwords, the following restrictions and desiderata, outlined in Campbell and Kaufman (1980; 1983) and further elaborated on in Campbell (1997:224–40) and Campbell and Poser (2008:163–212), have been considered.

H.1. Distribution of comparanda. One of the most important criteria for comparing proto-languages is the use of forms that are widely distributed within each of the language families. Failing to do so can lead to the presence of loanwords in a data set or even increase the risk of chance similarities, for narrow distribution within the family tree also equates to shallow time depth of the forms in question. This paper uses only data reconstructible to very early stages of Mayan and Mije-Sokean that are of comparable time depth: pMa, LpMa, CMa for Mayan; pMiSo, pMi, and pSo for Mije-Sokean. No comparanda of shallower time depth and narrower distribution than these were considered except as possible loans.

H.2. Semantic isomorphism. Just as important as the distribution of the comparanda is the criterion of semantic isomorphism: the requirement that the comparanda being considered be as close in meaning as possible; otherwise, the chance of accidental similarity increases dramatically. If the forms are not as semantically isomorphic as possible, then the situation becomes complicated by the fact that different scholars may have different views about what degree of semantic latitude is allowed and what types of semantic shift can be cited to account for it. Here, semantic isomorphism has been observed as closely as possible. In a few instances, plausible semantic shift is taken into account, and such cases are noted below. In some such cases, the approach followed was to survey the range of meanings for the relevant etymon in both language families, in order to assess whether similar lines of development were documented. Nevertheless, it has to be recognized that acceptable semantic latitude is relative and ought to be justified on a case-by-case basis.

H.3. Phonological distance. Phonetic or phonological distance is also a relative criterion. As Meillet (1925/1954) has remarked, similarity on its own may not be the crucial criterion, but whether or not a comparison of forms exhibiting semantic isomorphism shows that they abide by the regular correspondences identified elsewhere. A myriad of changes in one or both of the descendant languages being compared may have conspired to render true cognates barely recognizable as similar. It is the regularity of sound correspondences that can be used to ascertain whether semantically isomorphic forms are actually related. Perhaps phonological similarity or distance, on its own, can be said to be more useful in identifying possible loanwords. The primary

lexical data set, however, consists of sound correspondences that are regarded as "plausible," that is, where it is not a stretch to conceive of a scenario that can account for the sounds in the correspondence as reflexes of an ancestral sound. A few sound correspondences in this data set require the assumption of a sequence of successive changes in either Mayan or Mije-Sokean; I only compiled etyma for which only one of the biconsonantal correspondences exhibited significant phonological distance. The chances of accidental similarities would probably increase dramatically if one allows for significant phonological distance on both parts of the biconsonantal correspondence.

H.3. Factors that can increase risk of similarity not due to cognacy. Diffusion of lexical and grammatical morphemes will lead to close resemblances and the appearance of cognacy. In principle, the Comparative Method can discern the difference between true cognates and loanwords: the former should exhibit regular correspondences, the latter do not. However, "archaic loans," i.e., vocabulary diffused very early in the history of two languages, can over time acquire the appearance of regular sound correspondences and confuse the distinction between actual cognates, if there are any, and loanwords. Section **3** (in the text) addresses the sound correspondences evident in a large set of likely loanwords diffused between Mayan and MijeSokean languages, a few of which are proposed for the first time in this paper. Many of these are suspected to be loanwords because of their narrow distribution within one of the language families, as well as other factors, such as phonological similarity and incongruous phonotactic shapes, as discussed in more detail below.

Sound symbolic segments or sequences can resist regular sound change, as noted by Campbell and Kaufman (1983:364), and may also exhibit irregular alternations, as carefully illustrated by Kaufman (1994) for Huastecan, who cautions that certain rare sounds in a language may be restricted to onomatopoeic roots (iconic sound symbolism) and sound-symbolic roots (synesthetic sound symbolism). In a sense, this is analogous to the way in which different grades are differentially preserved among the Indo-European language families, occasionally leading to irregular correspondences among the vowels for a particular etymon. Yet, overall, a regular pattern of correspondences is in evidence or extractable through internal reconstruction. The difference lies in the fact that the Indo-European grades pertain to grammatical categories (e.g., verb tenses) and derivation (e.g., derivation of nouns from verbs), whereas sound-symbolic alternations typically distinguish degree (e.g., intensification) or size (e.g., diminutive), among other possible semantic distinctions. But, in the end, differential preservation of sound-symbolic forms can lead to irregular correspondences of a similar nature to those due to differential preservation of IndoEuropean grades. Sound-symbolic forms may also abide by the same changes evident in other types of vocabulary and, if so, they could exhibit the same regularly recurring correspondences attested therein. Thus, on their own, sound-symbolic forms should not be used to argue for a special relationship between languages, but there is no reason to exclude them from a study; instead, the correspondences present among such forms should be checked for consistency with the correspondences attested in other types of vocabulary not deemed to be sound symbolic.

The issue of widespread forms, the so-called pan-Americanisms (Campbell and Kaufman 1980; 1983, Campbell 1991, and Kaufman 1990:25–26), and also pan-Mesoamericanisms (Campbell et al. 1985 and Smith Stark 1996), is a somewhat nebulous one, but it is an important topic to consider, especially in long-range comparison. These could be the result of ancient

inheritance or areal diffusion, or perhaps even independently innovated and only accidentally similar forms as a result of diffused principles for sound symbolism. Particularly risky items to compare are the so-called nursery forms; two forms that typically fall within this realm, 'mother' and 'father', are in fact quite widespread within Mesoamerica and have been excluded altogether from the comparison. In general, widespread forms should not be used, on their own, to argue for a special relationship between any two languages. However, at least some of them should not be ignored altogether: they should be isolated from the rest of the vocabulary but checked for consistency with regular sound correspondences found among vocabulary items not consisting of pan-Americanisms or pan-Mesoamericanisms. If they exhibit innovations characteristic of each of the language families, and if they are consistent with regular sound correspondences found between the two language families, then they should not be excluded from the set of possible cognates.

Chance similarities are not at all uncommon when only short forms are compared or when the morphological composition of the comparanda have not been considered (Campbell and Kaufman 1980:853 and Campbell 1997:229). A common problem attested in the literature is the comparison of short sequences within larger forms, in which a part, sometimes a significant component of the entire forms under comparison, bear no resemblance whatsoever; often such "extraneous" components cannot even be analyzed, and this makes it unclear whether the short sequences that are being compared are even comparable morphologically and semantically in the first place. For this reason, morphological composition has been taken into account. While most of the forms in the data sets are lexical roots, a few are compounds, and in some such cases, only one of the terms of the compound may be of relevance for the comparison: A25, pMa **tuuk* : pMiSo **tuu¢*[-kuy] 'palm[-tree]', illustrates a case where a pMa root is compared with the first term of a pMiSo compound. For some comparanda, it was not possible to determine the morphological composition in detail, and they were excluded from both the primary and secondary lexical data sets, e.g., pMa **šikin* 'ear' and pMiSo **taa¢ɨk* 'ear', part of Fox's (1978) comparison, contain a sequence that is comparable, PMa **šik-* and PMiSo **-¢ɨk*, but until it is possible to account for their entire composition in the respective languages, these forms cannot be compared further.

As explained below, this paper uses the criterion of comparing biconsonantal correspondences for lexical morphemes: both consonants in a *CVC form in PMa are required to exhibit regular correspondences to both consonants in a *CVC or *CVCV form in PMiSo. This is the case of the primarily lexical data set, which constitutes the basis for the conclusions drawn in this paper. The secondary lexical data set includes a few items where one of the two consonantal correspondences does not exhibit recurrence, but at the very least they exhibit similarity (e.g., the non-recurring correspondence of pMa **ŋ* : pMiSo **n* in one lexical item), and their morphological composition is not in question. The risk of chance similarities is greater for the grammatical comparanda, which are by nature short forms, usually involving only one consonant and one vowel, sometimes simply a vowel, and are in general composed of the least marked segments of the phonemic inventory of the language (Campbell 1997:219–22 and Campbell and Poser 2008:189).[33]

[33] Interestingly, grammatical morphemes are regarded by some authors to be potentially stronger evidence, on their own, of a historical connection of shared descent. Also, grammatical morphemes are more prone to analogical reshaping and phonetic erosion than lexical morphemes.

These two facts combine to make chance resemblances much more common than some scholars have thought. Also worth considering, scholars have utilized the criterion of "shared aberrancies" (Meillet 1925/1954) in order to impose constraints in their comparison of grammatical comparanda, and some even argue that such unusual correspondences of morphological form:function patterns can provide stronger evidence of relationship than the Comparative Method. Nichols's (1996:46) "multidimensional paradigmaticity" approach calls for the recognition of cognate morphemes and cognate paradigmatic patterns, not necessarily of idiosyncratic or unusual patterns. This is probably what Kaufman (1990:25) implies by "distribution" when he states that "Grammatical correspondences involving agreements in sound, function, and distribution should be found." For these reasons, grammatical forms are relegated to their own data set, rather than incorporated into the primary lexical data set, and their evaluation is largely postponed to a future paper, with the exception of a couple of the grammatical morphemes that are discussed in some detail.

H.4. Factors that can obscure actual cognacy. Analogical change can alter or dissolve previously existing regular correspondences. Also, sound symbolism of the "conventional" type, or in Hock's (1991) terms, "systematic onomatopeia," is characterized by the analogical extension of specific sounds to reshape existing words or produce new words with specific meanings. At this stage, however, the goal should be to focus on regular sound correspondences. Should such correspondences be demonstrated and supported by additional evidence, then it might be possible to find evidence for analogical change. Another factor could be cases of cognate sounds that have undergone little change in the respective language families. Cases of forms composed of such sounds could give the impression of loans rather than cognates. And archaic loans could very well abide by regular correspondences present in potential cognates. There is a chance that some of the forms proposed to be cognates could be cases of archaic loans.