

APPENDIX A

LAHU SI STORIES

The four passages I used for Lahu Si were Mark 9:22-24 about the boy with seizures, Mark 1:30-31 about Jesus healing Peter's mother-in-law, Mark 1:32-38 about Jesus healing and praying alone, and Luke 12:13-15 about Jesus teaching against greed.

For the text of all Lahu Si stories presented here, the following key is used:

- (space) word break
- syllable break
- * syllable and morpheme boundary within a compound word
- % syllable and morpheme boundary within a compound word, replaced with a hyphen in word-spaced text and a space otherwise
- & syllable boundary for non-compound words, replaced with a hyphen in compound- and word-spaced text and a space otherwise

All character numbers exclude spaces and hyphens. Comprehension questions are listed after the text. Comprehension questions were asked verbally, not written.

Story 1: Mark 9:22b-24, practice text

77 words, 94 syllables, 331 characters

“Kheh•ma•khuhd, nawg ted ceng cear te gha ver, ngag•hehd haq har•kax lehq ga laoq lehd kud che yaog.”

Yer•suq koz che, “Te gha ver lehd nawg kud cheawg la. Yonr che chaw teq•pa awg•pun awg•ceng awg•gha•lia phehaq gha che yaog lehd, yawd haq kud vid che yaog.”

Awg•yand u•ve khuhn yad*par nehax ve awr•pa teq•pear lehax kug lehq, “Ngag yonr che yaog. Ngag yonr mad gag che awg•ceng haq a•cehax ga laoq lehd kud vid che yaog.”

Question 1: พระเยซูบอกว่าการทุกอย่างเป็นไปได้สำหรับใคร (Jesus said that all things are possible for whom?)

Question 2: ตอนสุดท้าย ผู้ชายคนนี้ขอให้พระเยซูช่วยอย่างไร (At the end, this man asks Jesus to help him in what way?)

Story 2: Mark 1:30-31, test text

51 words, 68 syllables, 217 characters

Sid•monr awg•mid•ma ve awr•ye kaz*nag lehq yuhq chehd che yaog. Yawd ved awg•lawn Yer•suq haq i•hehd har naq kud vid che yaog. Yer•suq lag lehq, yad*mid u•ve laq•kuhr haq yug lehq tuag vid tug te che yaog. Yawd ved kaz*nag che na vehr lehq, yawd i•hehd haq te•car*te•taw vid che yaog.

Question 1: ใครป่วย (ใครเป็นไข้) (Who was sick? (Who had a fever?))

Question 2: แม่ยายของซีโมนทำอะไรหลังจากที่พระเยซูช่วยรักษาเขา (What did Simon's mother-in-law do after Jesus healed her?)

Story 3: Mark 1:32-38, test text

168 words, 220 syllables, 739 characters

Muhd•ni kehg vehr lehq muhd•su•naz vehr teaz huh, nag•che*guh•che chaw awg•gha•lia he•lehq, ned geh che chaw teq•pa haq Yer•suq chehd lo sir vea lar che yaog. Med*yad awg•gha•lia ghaz•mir khuh awg•pad huh phoz chehd chehd che yaog. A•nar laex ceng te nag chehd che chaw kha•peh haq yawd ghu na vid lehq, ned teq•pa haq liz ghaq kog vid che yaog. Ned teq•pa yawd haq sir cheaq•te•lehq Yer•suq i•hehd haq tawd maq yaw cuh vid.

Muhd•sawx ted naq, muhd•ma•thid sear huh, Yer•suq tuag lehq heh•pir khuhn kae lehq, u•huh bon*lawg chehd che yaog. Sid•monr he•lehq yawd geh chehd che awg•chawd teq•pa.

Yer•suq haq ghaq ca che yaog. I•hehd Yer•suq haq mawg che te yand, “Chaw awg•gha•lia nawg haq ca chehd che yaog lehd yawd haq kud vid che yaog.”

Yer•suq i•hehd haq kud vid che, “Awg•pad%awg•ki huh cawg che med*khaz teq•pa haq liz ngag bon gha mar vid tug awg•nuz&awr kae veg. Awg•lawn•ghod•ver, che kheh te tug ngag lag che yaog lehd, yawd kawq kud vid che yaog.”

Question 1: เรื่องบอกว่าพระเยซูไม่ยอมให้ผีชั่วพูดเพราะอะไร (According to the story, why did Jesus not allow the evil spirits to speak?)

Question 2: พระเยซูทำอะไรก่อนเช้ามีคในที่เงียบสงบ (What did Jesus do early in the morning in a quiet place?)

Story 4: Luke 12:13-15, test text

106 words, 138 syllables, 457 characters

Chaw awg•mor nux u•ve khuhn chehd che te ghad Yer•suq haq kud vid che, “Sar•lar&og, ngag haq ngar*pa ve awg•sinx pehg ax tug nawg ngag*vix•var haq a•cehax kud viad,” lehd kud vid che yaog.

Kheh•ma•khuhd Yer•suq kawq koz che, “Yad par&og, nig•mag kanr huh suhx tar lehq awg•sinx pehg ax tug a•sug ngag haq teha lar cheawg le?” lehd yawd haq kud vid che yaog. Yawd i•hehd teq•pa haq kawq kud vid che, “Sa tiq te, su ved med•nged*jeg•kor haq sex lae che ni•ma tad cawg vid. Awg•lawn•ghod•ver, med•nged*jeg•kor cawg mad cheaq•te•lehq, chaw yad a•saq gha teq che ma hez hawg,” lehd kud vid che yaog.

Question 1: ชาวบ้านขอให้พระเยซูทำอะไร (What did the villagers ask Jesus to do?)

Question 2: พระเยซูเตือนเราว่าต้องระวังอะไร (What did Jesus warn us to watch out for?)

APPENDIX B

AKHA STORIES

The three passages I used for Akha were Mark 9:22-24 about the boy with seizures, Mark 1:32-38 about Jesus healing and praying alone, and Luke 12:13-15 about Jesus teaching against greed.

For the text of all Akha stories presented here, the following key is used:

- (space) word break
- syllable break
- * syllable break, hyphen in Bible text removed
- % syllable break, apostrophe in Bible text removed

All character numbers exclude spaces, hyphens, and apostrophes. Comprehension questions are listed after the text. Comprehension questions were asked verbally, not written.

Story 1: Mark 9:22b-24, practice text

53 words, 102 syllables, 334 characters

Nawr tiq•jeq•nav janr nya nar*aq, nga%nyaf da•zaq*anr ghawq•shaq•gaq*awr, .
caw•ba naeq•laf*awq laer shar ma.” 23 Yeq•sur aq•yawf*anr heu•lof*aer aer
naeq*euq mae. “Nawr jan nya nar*aq, jan*eu tsawr•haq*euq ghar•duq*aq,
yawq•jeq•nav•luf pyeuf ler nya taef*a.” 24 Khoer•ngae, aq•li khoer ghaq*euq aq•da
tiq•pov•taef*aer myav•bir ju•ju*aer heu•lof*aer gur dov•laf*euq mae. “Sanq•paq*or,
ngar jan•ma, ngar jan*eu myaq lar ni caw•ba naeq•laf*awq.”

Question 1: พระเยซูบอกว่าทุกสิ่งทุกอย่างเป็นไปได้สำหรับใคร (Jesus said that all things are possible for whom?)

Question 2: ตอนสุดท้าย ผู้ชายคนนั้นขอให้พระเยซูช่วยอย่างไร (At the end, this man asks Jesus to help him in what way?)

Note that this story is missing a beginning quote mark on the first sentence; it was, however, only a practice text and not timed.

Story 2: Mark 1:32-38, test text

110 words, 247 syllables, 819 characters

Nanr•ma ga%anr*awr, uq•cif cif•le*iq ngae, nar*eu gawr*eu tsawr•haq dawq•tawv•luf hawr*eu naef ur*eu tsawr•haq*anr aq•cawq Yeq•sur jawr•gar sheuq•laf*euq mae. Meuq*anr jawr*eu tsawr•haq dawq•tawv•luf khoer•nymr*euq ghov•khaeq*anr uiq•lar•taq*euq mae. Nar•gawr ghawq•mya•jeq nar•luf*eu tsawr•haq ghawq•mya%ni*anr Yeq•sur lav•taq•naeq*euq uq•taf, naef ghawq•mya•mawr*anr khawf tae•dov*euq mae. Naef khoer•deuq aq•yawf*anr siq•nya*eu mir•nae, Yeq•sur naef khoer•deuq*anr dawq maq•bin•gaeq*aq.

Uq•shawq yaw•naf*aer mq maq•bya•lar mi•nav*aer, Yeq•sur tur*awr dov*ir*awr nae, cof•ngaer•ni ghae*eu bawr•tsanq*anr ir*awr, khoer•gar guiq•lanq•shar*euq mae. Sir•moq hawr*eu aq•yawf naeq*aer jawr*eu deuq Yeq•sur*anr pov*ir*euq mae. Pov•mawr ngae, Yeq•sur*anr heu•lof*aer aer•naeq*euq mae. “Tsawr•haq dawq•tawv•luf nawq*anr pov•luf mae.” Yeq•sur aq•yawf•maf*anr heu•lof*aer aer•naeq*euq mae. “Dawq•pae jav*eu meuq•deuq pu•deuq*anr khawf ngar dawq•maeq ya%maeq•naeq•ni, aq•dui•maf yawq•gar•yawq ir kar*eur. Aq•jeq•mir•nae aer•nar*aq, ngar heu•lof*aer mr*eu ghar•duq oe*euq ma.”

Question 1: เรื่องบอกว่าพระเยซูไม่ยอมให้ผีชั่วพูดเพราะอะไร (According to the story, why did Jesus not allow the evil spirits to speak?)

Question 2: พระเยซูทำอะไรตอนเช้ามีดในที่เงียบสงบ (What did Jesus do early in the morning in a quiet place?)

Story 3: Luke 12:13-15, test text

70 words, 151 syllables, 506 characters

Khoer•ngae, “Sanq•paq*or, ngaq*euq maer•nmr ngaq*anr jeq•bir bir•naeq•laf ni aq•yawf*anr tiq•pov aer•naeq*iq sher*eq.” laer, tsawr•haq ar•maf maf•ma*euq laf•khoer tiq•ghaq Yeq•sur*anr shar*eu mae. “Cawq*or, naw•nyaf•nmr*anr dar•naeq*eu ghaq hawr*eu jeq•bir bir•naeq*eu ghaq geuf ni, aq•sur•ghaq nae ngaq*anr khaer•taq*euq la” laer, Yeq•sur aer•naeq*euq mae. Khoer•ngae, “Naw•maf

aq•cawq*euq myawq tiq•jeq*anr*iq maer•juir maq•juir ni, yawq•har yawq•ha
yaw•muiq*aer yanr*awq. Aq•jeq•mir•nae aer•nar*aq, pyur•shuir myawq•ganq
ghawq•mya%ni jav*eu mir•nae, tsawr•haq yaw•daef daef•nya*eu maq•ngeur*a” laer,
Yeq•sur tsawr•haq ar•maf maf•ma*anr aer•naeq*euq mae.

Question 1: ชาวบ้านขอให้พระเยซูทำอะไร (What did the villager ask Jesus to do?)

Question 2: พระเยซูเตือนเราว่าต้องระวังอะไร (What did Jesus warn us to watch out for?)

PAYAP UNIVERSITY

APPENDIX C

SPACING FREQUENCIES IN HMONG TEXTS

The table below contains the data used for the analysis of spacing practices by Hmong writers. The first two columns give the number of instances each word was written joined and separated in the soc.culture.hmong corpus. The third and fourth columns give the same information for the United Bible Societies New Testament. The final column gives the number of instances each word was found in the Catholic Bible translation. All syllables are consistently separated with a space in the Catholic Bible.

Spelling variants due to tone sandhi or to differences between Hmong Daw and Hmong Njua spellings in the soc.culture.hmong corpus are included in the overall joined and spaced numbers for the first two columns. The complete data set with the number of variants for each word is available upon request from the author.

Word	Joined in SCH	Spaced in SCH	Joined in NT	Spaced in NT	In Catholic Bible (spaced)
<i>aub ncaug</i>	4	60	0	0	2
<i>caj dab</i>	112	707	14	0	17
<i>caj npab</i>	9	75	1	0	2
<i>cheb cheb</i>	0	1	0	0	0
<i>cua daj cua dub</i>	0	29	0	11	5
<i>cua nab</i>	2	16	0	0	1
<i>dab tsi</i>	6,321	10,058	257	0	263
<i>dab tuag</i>	17	130	0	0	1
<i>dav hlau</i>	168	1,011	0	0	0
<i>dhia dhia</i>	3	78	0	0	0
<i>di ncauj</i>	5	26	0	0	37
<i>diav rawg</i>	0	1	0	0	0
<i>hauj lwm</i>	4,126	5,586	463	0	216
<i>hauv ncoo</i>	12	106	0	0	1
<i>hauv pliaj</i>	23	145	10	0	19
<i>hiav txwv</i>	82	215	114	0	82

Word	Joined in SCH	Spaced in SCH	Joined in NT	Spaced in NT	In Catholic Bible (spaced)
<i>huv si</i>	49	230	0	0	69
<i>ib leeg</i>	185	4,656	0	137	72
<i>ib qho</i>	58	7,481	0	75	80
<i>ib los</i>	37	2,160	0	38	4
<i>ib tug</i>	250	34,290	0	952	783
<i>kab tsib</i>	52	34	0	0	0
<i>kaj ntug</i>	391	353	0	24	44
<i>Kaj Siab</i>	439	539	0	91	28
<i>kawg nkaus</i>	70	2,010	0	130	4
<i>kawm ntawv</i>	168	4,389	0	5	0
<i>kev cai</i>	1,152	4,057	739	0	47
<i>kev mob kev tuag</i>	0	5	0	0	0
<i>kheej kheej</i>	0	59	0	0	0
<i>khov kho</i>	40	530	0	12	27
<i>khwv iab khwv daw</i>	0	1	0	0	0
<i>ko taw</i>	144	742	83	0	42
<i>loj loj</i>	13	1,009	0	15	24
<i>me ntsis</i>	776	3,875	43	0	26
<i>me nyuam</i>	5,511	6,691	327	0	230
<i>nab qa</i>	1	9	2	0	3
<i>nees nkaum</i>	4	30	0	18	21
<i>niam txiv</i>	636	3,979	3	7	4
<i>nkag siab</i>	1,217	1,321	0	0	0
<i>nkhaus niv nkhaus nom</i>	0	0	0	0	0
<i>nom tswv</i>	466	2,959	124	0	34
<i>nplooj ntoos</i>	1	96	0	1	4
<i>ntau ntau</i>	25	2,106	0	3	11
<i>ntev ntev</i>	9	442	0	5	3
<i>ntiaj teb</i>	1,415	4,854	382	0	371
<i>ntsej muag</i>	260	2,851	132	0	203
<i>ntxoov ntxoo</i>	2	25	0	0	14
<i>nus muag</i>	10	65	0	0	0
<i>ob tug</i>	135	6,252	0	163	144
<i>pas dej</i>	89	420	10	0	9

Word	Joined in SCH	Spaced in SCH	Joined in NT	Spaced in NT	In Catholic Bible (spaced)
<i>phooj ywg</i>	7,106	5,920	71	0	56
<i>plab hlaub</i>	7	55	0	0	2
<i>pluag tshais</i>	1	6	0	0	0
<i>pob caus</i>	19	100	0	0	0
<i>pob txha</i>	110	601	0	1	44
<i>pob zeb</i>	777	1,009	88	0	77
<i>pog koob yawg koob</i>	0	17	0	0	0
<i>poj niam</i>	10,608	11,500	377	0	338
<i>qaub ncaug</i>	10	70	9	0	1
<i>qhov ntswg</i>	17	144	0	0	10
<i>qhov rooj</i>	85	1,021	0	3	102
<i>qhov rai</i>	7	94	3	0	1
<i>qhov tsua</i>	21	274	0	2	26
<i>rhiab rhiab</i>	0	50	0	0	0
<i>ris tsho</i>	40	441	0	56	31
<i>rua lo</i>	0	20	0	0	0
<i>sawv ntxov</i>	37	360	0	13	53
<i>sib ceg</i>	127	1,459	0	26	7
<i>sov so</i>	8	47	0	0	1
<i>tab sis</i>	15,704	14,136	0	0	428
<i>tag kis</i>	282	622	14	0	63
<i>taj laj</i>	212	195	0	0	0
<i>tam sim</i>	1,844	3,551	115	0	23
<i>taub dag</i>	5	36	0	0	0
<i>teb chaws</i>	9,054	13,013	280	0	553
<i>tiv thaiv</i>	84	1,016	0	0	25
<i>tsaug zog</i>	58	886	0	31	32
<i>tseg tub tseg ki</i>	0	2	0	0	0
<i>tsev neeg</i>	577	4,162	16	0	121
<i>tshaib plab</i>	18	388	0	17	30
<i>tsov rog</i>	75	727	0	0	1
<i>tswv yim</i>	3,373	5,322	118	0	173
<i>tub nkeeg</i>	45	293	5	3	10
<i>ua dog ua dig</i>	0	370	0	10	1

Word	Joined in SCH	Spaced in SCH	Joined in NT	Spaced in NT	In Catholic Bible (spaced)
<i>ua qoob ua loo</i>	0	17	0	0	0
<i>ua si</i>	400	1,585	0	2	20
<i>viav vias</i>	2	12	0	0	1
<i>viv ncaus</i>	52	165	2	0	3
<i>xeeb ntxwv</i>	103	523	61	0	80
<i>xws li</i>	109	2,042	2	0	260
<i>yooj yim</i>	847	1,818	13	0	2
<i>zaj sawv</i>	0	10	0	1	4
<i>zaub mov</i>	59	636	0	32	3
<i>zib mu</i>	3	11	0	0	11
<i>zom zaws</i>	64	458	0	28	14
<i>zoo nkauj</i>	96	2,428	0	18	77

APPENDIX D

SOURCES FOR POLYSYLLABIC HMONG WORDS

Below is a table of the polysyllabic words in Hmong Daw used for testing in the US and Thailand, along with references for determining their word status. In addition to the sources listed for each word, all reduplicated forms are considered words on the basis of Ratliff 2009 and Ratliff 2010's classification of reduplication as a morphological feature. Elaborate expressions are considered compound words on the basis of Ratliff 2009's analysis of these expressions as compounds, as well as Mortensen 2003's description of elaborate expressions as a form of coordinate compounds.

Abbreviations for word types are as follows. Some words have more than one reason for being considered words.

op	semantically opaque compound
b	bound morpheme
mono	monomorphemic
sesqui	sesquimorphemic (one syllable is unanalyzable, perhaps a fossilized morpheme)
redup	reduplication
ee	4-syllable elaborate expression
ts	tone sandhi compound
orth	normally written unspaced
cc	coordinated compound, "semantic reduplication"
phon	phonological unity (apart from tone sandhi)
V-obj	a verbal compound with a fixed, obligatory object incorporated into the verb
lex	lexicalized

Word	US list	Thai list	Stories	Word type	Source
<i>aub ncaug</i>	x			ts	Ratliff 2010:164
<i>caj dab</i>		x		b	Ratliff 2009
<i>caj npab</i>	x	x		b	based on Ratliff 2009's description of <i>caj</i> - as body part noun class prefix
<i>cheb cheb</i>	x	x		redup	Owensby 1986:239
<i>cua daj cua dub</i>	x			ee	Ratliff 2009, Johns & Strecker 1987:109
<i>cua nab</i>			x	sesqui	Ratliff 2009
<i>dab tsi</i>	x	x		mono	Ratliff 2009
<i>dab tuag</i>	x	x		op	Ratliff 2009
<i>dav hlau</i>	x	x		op	Ratliff 2009
<i>dhia dhia</i>	x			redup	Owensby 1986:238
<i>di ncauj</i>	x			cc	Ratliff 2009
<i>diav rawg</i>	x			op	my own analysis ("spoon-chopsticks" = fork)
<i>hauj lwm</i>	x			mono	Ratliff 2009
<i>hauv ncoo</i>		x		phon	Ratliff 2009
<i>hauv pliaj</i>	x			cc	Ratliff 2009
<i>hiav txwv</i>	x			b	Ratliff 2009
<i>huv si</i>	x			sesqui	Ratliff 2009, Heimbach 1979:56
<i>ib leeg</i>	x			ts	Ratliff 2009
<i>ib qho</i>	x			ts	Ratliff 2009
<i>ib los</i>			x	ts	Ratliff 2010:37
<i>ib tug</i>			x	ts	Ratliff 2010:30
<i>kab tsib</i>	x	x		mono	Ratliff 2009
<i>kaj ntug</i>	x	x		ts	Ratliff 2009
<i>Kaj Siab</i>			x	name	see below
my own analysis based on it being a (single) name, and the fact that names are typically unspaced in Hmong texts which use some form of word spacing					
<i>kawg nkaus</i>			x	lex?	see below

Word	US list	Thai list	Stories	Word type	Source
The word <i>kawg nkaus</i> is a single entry in Xiong 2005, and is used as a unit in superlative constructions. <i>Kawg</i> can mean “extremity” or “end”, and <i>nkaus</i> is a restricted post-verbal intensifier according to Heimbach 1979:153. While there are no clear indicators of its word status, it appears to be lexicalized as a superlative with a fixed form.					
<i>kawm ntawv</i>	x	x		V-obj	Ratliff 2009
<i>kev cai</i>		x		cc	Ratliff 2009
<i>kev mob kev tuag</i>	x			ee	Heimbach 1979:81
<i>kheej kheej</i>	x			redup	Ratliff 2009
<i>khov kho</i>	x	x		redup	Ratliff 2010:86
<i>khwv iab khwv daw</i>	x			ee	Johns & Strecker 1987:106
<i>ko taw</i>	x	x		b	Ratliff 2009
<i>loj loj</i>			x	redup	based on Ratliff 2009, 2010
<i>me ntsis</i>		x		cc	Ratliff 2009
<i>me nyuam</i>	x	x		cc	Ratliff 2009
<i>nab qa</i>	x	x		ts, op	Ratliff 2009
<i>nees nkaum</i>		x		phon	Ratliff 2009
<i>niam txiv</i>	x	x		cc	Ratliff 2009
<i>nkag siab</i>	x	x		op	Ratliff 2009
<i>nkhaus niv</i> <i>nkhaus nom</i>	x			ee	Johns & Strecker 1987:106
<i>nom tswv</i>	x	x		cc	Ratliff 2009
<i>nplooj ntoos</i>			x	ts	Ratliff 2010:179
<i>ntau ntau</i>	x	x	x	redup	based on Ratliff 2009, 2010
Note: <i>ntau ntau</i> was used in the sentence-by-sentence test in Thailand, but not for the stories test in the US.					
<i>ntev ntev</i>	x	x		redup	Ratliff 2009
<i>ntiaj teb</i>	x	x		cc	Ratliff 2009
<i>ntsej muag</i>			x	cc	Ratliff 2009
<i>ntxoov ntxoo</i>	x	x		redup	Ratliff 2009
<i>nus muag</i>	x	x		cc	Ratliff 2009
<i>ob tug</i>		x		ts	Heimbach 1979:326
<i>pas dej</i>			x	lex?	see below

Word	US list	Thai list	Stories	Word type	Source
Ratliff 2009 says <i>pas dej</i> is phrasal according to its rules (no tone change, semantically predictable), but then gives it as an example of how the rules are somewhat arbitrary, since <i>pas dej</i> is usually thought of as a single unit, and when you ask how to say “lake,” you always get <i>pas dej</i> and not just <i>pas</i> . Xiong 2005 has it as a single entry, along with just <i>pas</i> . So it seems to be a lexicalized compound, albeit a semantically predictable one that does not show any phonological unity or have a coordinate form.					
<i>phooj ywg</i>	x	x		mono	Ratliff 2009
<i>plab hlaub</i>	x	x		op	Ratliff 2009
<i>pluag tshais</i>			x	ts	Heimbach 1979:253
<i>pob caus</i>		x		b	Ratliff 2009
<i>pob txha</i>	x	x		b	Ratliff 2009
<i>pob zeb</i>			x	b	Ratliff 2009
<i>pog koob yawg koob</i>	x			ee	Ratliff 2009
<i>poj niam</i>	x	x		cc	Ratliff 2009
<i>qaub ncaug</i>		x		ts	Ratliff 2010:191
<i>qhov ntswg</i>	x	x		b	Ratliff 2009
<i>qhov rooj</i>	x	x		b	Ratliff 2009
<i>qhov rai</i>			x	b	Ratliff 2009
<i>qhov tsua</i>			x	b	based on Ratliff 2009’s description of <i>qhov-</i> as a noun class prefix
<i>rhiab rhiab</i>			x	redup	based on Ratliff 2009, 2010
<i>ris tsho</i>			x	cc	Ratliff 2009
<i>rua lo</i>			x	orth	Xiong 2005:370
<i>sawv ntxov</i>			x	op	Ratliff 2009
<i>sib ceg</i>	x	x		b, ts	Ratliff 2009
<i>sov so</i>	x	x		redup	Ratliff 2010:86
<i>tab sis</i>	x	x	x	orth	Xiong 2005:388
<i>tag kis</i>			x	mono	Ratliff 2009
<i>taj laj</i>		x		mono	Ratliff 2009
<i>tam sim</i>	x	x		mono	Ratliff 2009
<i>taub dag</i>			x	ts	Ratliff 2009
<i>teb chaws</i>	x	x		ts	Ratliff 2009

Word	US list	Thai list	Stories	Word type	Source
<i>tiv thaiv</i>		x		cc	Ratliff 2009
<i>tsoog zog</i>	x	x	x	op	Ratliff 2009
<i>tseg tub tseg ki</i>	x			ee	Heimbach 1979:82; Johns & Strecker 1987:110
<i>tsev neeg</i>	x	x		op	Ratliff 2009
<i>tshaib plab</i>	x	x		V-obj	Ratliff 2009
<i>tsov rog</i>	x	x		op	Ratliff 2009
<i>tswv yim</i>	x	x		mono	Ratliff 2009
<i>tub nkeeg</i>	x	x		ts	Ratliff 2009
<i>ua dog ua dig</i>	x			ee	Ratliff 2010:162
<i>ua qoob ua loo</i>	x			ee	Heimbach 1979:265; Johns & Strecker 1987:109
<i>ua si</i>			x	sesqui	Ratliff 2009
<i>viav vias</i>	x	x		redup	based on Ratliff 2009, 2010
<i>viv ncaus</i>	x	x		mono	Ratliff 2009
<i>xeeb ntxwv</i>	x	x		phon, b	Ratliff 2009
<i>xws li</i>			x	cc	my own analysis (see Mortensen 2003)
<i>yooj yim</i>	x	x		mono	Ratliff 2009
<i>zaj sawv</i>	x	x		op	Ratliff 2009
<i>zaub mov</i>	x	x		cc	Ratliff 2009
<i>zib mu</i>			x	ts	Ratliff 2009
<i>zom zaws</i>			x	mono	Heimbach 1979:477
<i>zoo nkauj</i>			x	cc	Ratliff 2009

APPENDIX E

HMONG STORIES

I used four stories in Hmong Daw, all from a set of stories in the Houghton Mifflin Literary Readers: Beginning to Read, from levels A and C. These stories are all translated from English by a team working with the Merced City School District in Merced, California. The stories were all edited for length, plus one minor edit for word choice in Hmong.

Syllable boundaries in polysyllabic words are marked with the bullet symbol, •. Character counts exclude spaces but include punctuation marks.

Story 1: Based on Pollock 1989b, control text

*Only presented in word-spaced format
90 words, 97 syllables, 349 characters*

Koj siv tau dej ua ntau yam — xws•li haus, ua•si, thiab tso dej ntwis taug koj lub plhu. Koj rau koj nkawm khau yas mus kwm dej los tau, xuas pas nplawm dej, los muab pob•zeb txawb dej. Dej zoo coj los pab txhua yam kom hlob taus.

Nag yog dej los saum ntuj los. Nag txauj saum koj lub kaus kwv thiab ua kom koj cov ntiv tes rhiab•rhiab.

Dej zoo rau yus dhia ua•si, tso cev rau, thiab ntab saum cov nplaim. Dej ntub — thiab dej zoo•nkauj kawg•nkaus.

Question 1: Name two things that the writer says water is useful for.

Question 2: How does the writer describe rain?

Story 2: Based on Titherington 1989b, test text

Presented as syllable-spaced to group 1 and word-spaced to group 2

94 words, 112 syllables, 394 characters

Kaj•siab cog ib lub noob taub•dag, ces lub noob taub•dag tawg ua ib•tug kaus, ces tus kaus hlav ua ib tsob taub•dag, ces tsob taub•dag tawg ua ib lub paj taub•dag, ces lub paj taub•dag txi ua ib lub qe taub•dag. Ces lub qe taub•dag loj...thiab loj...thiab loj, txog thaum Kaj•siab mus muab nws de. Ces Kaj•siab kaus cov hlwb taub•dag tawm, tho lub taub•dag ua ib lub ntsej•muag, thiab coj nws mus txawb rau saum qhov•rai. Tab•sis...nws khaws rau lub noob taub•dag tseg tau cog rau lub caij ntuj tshiab.

Question 1: Name three things Kajsiaab did.

Question 2: Where did Kajsiaab put the carved pumpkin face after she made it?

Story 3: Based on Dabovich 1989b, control text

Presented only in syllable-spaced format

68 words, 75 syllables, 272 characters

Ntuj pib no. Nplooj•ntoo zeeg zom•zaws. Noog tsiv khiav mus lawm. Dais tsaug tsaug•zog. Nws nrhiav tau ib lub qhov•tsua. Ntuj los los daus. Tab•sis dais nyob sov so hauv nws lub qhov•tsua. Lub hnuv rov tawm tuaj dua. Noog ya rov qab los. Me kab me ntsaum rov qab tawm tuaj. Muv ya rov qab los. Dais nco qab txog zib•mu. Nws raws cov muv qab lawm.

Question 1: Where is the bear while it is snowing?

Question 2: Name two things that the story says come back in the spring.

Story 4: Based on Asch 1989b, test text

Presented as word-spaced to group 1 and syllable-spaced to group 2

95 words, 109 syllables, 377 characters

Tag•kis no sawv•ntxov thaum kuv sawv los kuv rua•lo ib•los loj•loj. Kuv ntxuav muag, hnav ris•tsho, thiab noj ib pluag•tshais loj•loj. Ces kuv hnav kuv lub tsho tiv no thiab rau kuv nkawm khau yas. Ces peb mus nuv ntses. Txog tom ib tog kev, kuv

de tau ib lub paj ces kuv muab rau kuv niam. Thaum peb mus txog nram pas•dej, kuv muab ib•tug cua•nab loj•loj chob rau kuv tus koob nuv ntse. Peb nuv ntse thiab nuv ntse ib hnuv, kuv nuv tau ib•tug ntse loj•loj...zoo ib yam li kuv niam.

Question 1: What did the child who tells the story give to his mother on the way to fishing?

Question 2: How long were the child and his family fishing for?

PAYAP UNIVERSITY

APPENDIX F

HMONG STORIES BY SENTENCES

Using the same four stories in Hmong Daw as for the US testing (see Appendix E), I broke each story up into sentences or long clauses. Some modifications were made to the stories in order to better fit the culture and weather of Thailand.

Syllable boundaries in multisyllabic words are marked with either a bullet symbol (•) or an asterisk (*). Bullets indicate a word that was joined for group 1, and asterisks indicate words joined for group 2. Polysyllabic words that are the target of comprehension questions are marked in bold.

Practice sentences (not measured)

- p1. Kuv hais tau lus Hmoob.
- p2. Nws nyiam noj zaub.
- p3. Nees muaj plaub txhais ceg.
- p4. Tub nyob hauv vaj.

Story 1: Based on Pollock 1989b

- 1. Koj siv dej ua tau ntau yam —
- 2. xws•li haus, ua•si, thiab tso dej ntw•s saum koj lub plhu.
- 3. Koj rau koj nkawm khau mus kw•m dej los tau,
- 4. xuas pas nplawm• dej, los muab pob•zeb txawb• dej.

Question 1: ช่วยบอกประโยชน์ของน้ำสองอย่างตามคนเขียนครับ (Please say two things that water is useful for according to the writer.)

- 5. Dej zoo koj los pab txhua yam kom hlob taus.
- 6. Nag yog dej los saum ntuj los.
- 7. Nag txauj saum koj lub kaus kwv thiab ua kom koj cov ntiv tes rhiab•rhiab.

Question 2: เวลาฝนตกโดนนิ้วมือคุณรู้สึกอย่างไรครับ (When rain falls on your finger, how do you feel?)

8. 1. Dej zoo rau yus dhia ua•si, tso cev rau, thiab ntab saum nplaim dej.
9. 2. Dej ntub — thiab dej zoo*nkauj kawg*nkaus.

Story 2: Based on Titherington 1989b

10. Kaj*siab cog ib lub noob taub*dag,

Question 3: กำเขี่ยทำอะไรครับ (What did Kajsiaab do?)

11. ces lub noob taub tuaj ua ib•tug kaus,
12. ces tus kaus hlav ua ib tsoob taub*dag,
13. ces tsoob taub*dag tawg ua ib lub paj taub,
14. ces lub paj taub txi ua ib lub qe taub.
15. Ces lub qe taub loj...thiab loj...thiab loj,
16. txog thaum Kaj*siab mus muab de.
17. Ces Kaj*siab kaus cov hlwb taub tawm,
18. tho lub taub ua ib lub ntsej•muag,
19. thiab coj nws mus txawb rau saum qhov•rai.

Question 4: กำเขี่ยเอาหน้าฟักทองไว้ที่ไหนครับ (Where did Kajsiaab put the pumpkin face?)

20. Tab*sis...nws khaws rau lub noob taub tseg tau cog rau lub cajj ntuj tshiab.

Story 3: Based on Asch 1989b

21. Tag*kis no sawv•ntxov thaum kuv sawv los kuv rua•lo ib•los loj•loj.

Question 5: เด็กตื่นนอนทำอะไรครับ (What does the child do when s/he wakes up?)

22. Kuv ntxuav muag, hnav ris*tsho, thiab noj ib pluag*tshais ntau*ntau.

Question 6: แล้วเด็กก็ทำอะไรครับ (Then what did the child do?)

23. Ces kuv hnav kuv lub tsho tiv no thiab rau kuv nkawm khau.
24. Ces peb mus nuv ntsej.
25. Txog tom ib tog kev, kuv de tau ib lub paj ces kuv muab rau kuv niam.
26. Thaum peb mus txog nram pas*dej,
27. kuv muab ib•tug cua•nab loj•loj chob rau kuv tus koob nuv ntsej.

28. Peb nuv ntses thiab nuv ib hnub,
29. kuv nuv tau ib•tug ntses loj•loj...
30. zoo ib yam li kuv niam tus.

Story 4: Based on Dabovich 1989b

31. Ntuj pib no.
32. Nplooj*ntoo zeeg zom*zaws.
33. Noog ya khiav mus lawm.
34. Dais tsaug tsaug•zog.
35. Nws nrhiav tau ib lub qhov*tsua.

Question 7: หมีเจออะไรครับ (What did the bear find?)

36. Ntuj los los nag tshauv.
37. Tab•sis dais nyob sov so hauv nws lub qhov*tsua.
38. Lub hnub rov tawm tuaj lawm.
39. Noog ya rov qab los.
40. Me kab me ntsaum rov qab tawm tuaj.
41. Muv ya rov qab los.
42. Dais nco qab txog zib•mu.

Question 8: หมีคิดถึงอะไรครับ (What did the bear think about?)

43. Nws raws cov muv qab lawm.

APPENDIX G

HMONG WORD LISTS

Since several changes were made on the word test between testing readers in the US and Thailand, I list them separately.

Abbreviations for word types are as follows. Some words have more than one reason for being considered words.

- op semantically opaque compound
- b bound morpheme
- mono monomorphemic
- sesq sesquimorphemic (second syllable is unanalyzable, perhaps a fossilized morpheme)
- redup reduplication
- ee 4-syllable elaborate expression
- ts tone sandhi compound
- orth normally written unspaced
- cc coordinated compound, “semantic reduplication”
- phon phonological unity (apart from tone sandhi)
- V-obj a verbal compound with a fixed, obligatory object incorporated into the verb

Word list for US readers

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
p1	<i>plab hlaub</i>	stomach-lower.leg	calf of the leg	op	s
p2	<i>aub ncaug</i>	water-mouth	saliva	ts	w
p3	<i>khwv iab khwv daw</i>	toil-bitter-toil-salty	arduous toil	ee	w
p4	<i>nom tswv</i>	leader-master	officials, government	cc	s
p5	<i>kab tsib</i>	-	sugar cane	mono	s

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
p6	<i>tab sis</i>	always-even.though	but	orth	w
1	<i>huv si</i>	all-?	all; clean	sesq	s
2	<i>ib leeg</i>	one-person	alone, single person	ts	w
3	<i>pog koob</i> <i>yawg koob</i>	grandmother-great-grandfather-great	ancestors	ee	s
4	<i>caj npab</i>	Noun.class.body.part-arm	arm	b	w
5	<i>me nyuam</i>	small-little	baby, child	cc	s
6	<i>nab qa</i>	snake-frog	lizard	ts, op	s
7	<i>tseg tub</i> <i>tseg ki</i>	bereft-son-bereft-daughter	bereft of children	ee	w
8	<i>tsev neeg</i>	house-people	family	op	w
9	<i>pob txha</i>	Noun.class.ball-bone	bone	b	s
10	<i>teb chaws</i>	land-place	country	ts	w
11	<i>nkhaus niv</i> <i>nkhaus nom</i>	bent-INTENSIFIER-bent-INTENSIFIER	curvy, crooked	ee	s
12	<i>kaj ntug</i>	bright-sky	dawn, daylight	ts	w
13	<i>qhov rooj</i>	Noun.class.hole-door	door, gate	b	s
14	<i>yooj yim</i>	-	easy	mono	s
15	<i>zaub mov</i>	vegetable-rice	food	cc	s
16	<i>ko taw</i>	Noun.class.handle-foot	foot	b	w
17	<i>ntev ntev</i>	long.long	long	redup	s
18	<i>hauv pliaj</i>	head-forehead	forehead	cc	w
19	<i>diav rawg</i>	spoon-chopsticks	fork	op	s
20	<i>phooj ywg</i>	-	friend	mono	s
21	<i>xeeb ntxwv</i>	descendant-NOMINALIZER	grandchild, descendant	b, phon	s
22	<i>tswv yim</i>	-	idea, wisdom	mono	w
23	<i>tub nkeeg</i>	person-tired	lazy, lethargic	ts	s
24	<i>tam sim</i>	-	immediately	mono	w
25	<i>di ncauj</i>	lip-mouth	lip	cc	s
26	<i>dav hlau</i>	hawk-iron	airplane	op	s

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
27	<i>qhov ntswg</i>	Noun.class.hole-nose	nose	b	w
28	<i>niam txiv</i>	mother-father	parents, husband/wife	cc	w
29	<i>ib qho</i>	one-thing	piece, part	ts	s
30	<i>ntau ntau</i>	much-much	very much/many	redup	w
31	<i>zaj sawv</i>	dragon-rise	rainbow	op	s
32	<i>kheej kheej</i>	round.round	round	redup	w
33	<i>hiav txwv</i>	sea-NOMINALIZER	sea	b	w
34	<i>ntxoov ntxoo</i>	-	shade, shadow; cloudy	redup	w
35	<i>nus muag</i>	brother-sister	sibling	cc	s
36	<i>kev mob kev tuag</i>	way-sick-way-death	sickness	ee	w
37	<i>viv ncaus</i>	-	sisters; female cousins	mono	s
38	<i>cua daj cua dub</i>	wind-yellow-wind- black	storm	ee	w
39	<i>khov kho</i>	-	strong	redup	s
40	<i>viav vias</i>	swing-swing	swing (n)	redup	s
41	<i>cheb cheb</i>	sweep-sweep	sweeping	redup	s
42	<i>tshaib plab</i>	hungry-stomach	to be hungry	V-obj	w
43	<i>ua dog ua dig</i>	do-?-do-? (dog-dig = "badly, haphazardly")	to do badly, haphazardly	ee	w
44	<i>dhia dhia</i>	jump-jump	to keep jumping/running	redup	w
45	<i>sib ceg</i>	RECIPROCAL-scold	to quarrel	b, ts	s
46	<i>ua qoob ua loo</i>	do-crop-do-crop?	to raise crops	ee	w
47	<i>ntiaj teb</i>	surface-earth	world, earth	cc	w
48	<i>tsaug zog</i>	weak-strength	to sleep	op	w
49	<i>kawm ntawv</i>	study-paper	to study	V-obj	s
50	<i>nkag siab</i>	enter-liver	to understand	op	s
51	<i>sov so</i>	-	warm(er)	redup	s

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
52	<i>dab tuag</i>	ghost-dead	ugly, sloppy; demon	op	w
53	<i>dab tsi</i>	-	what?	mono	w
54	<i>poj niam</i>	female-mother	woman	cc	w
55	<i>tsov rog</i>	tiger-war	war, battle	op	s
56	<i>hauj lwm</i>	-	work (n)	mono	w

Word list for Thailand readers

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
1	<i>ntau ntau</i>	much-much	very much/many	redup	s
2	<i>tab sis</i>	always-even.though	but	orth	s
3	<i>tsaug zog</i>	weak-strength	to sleep	op	w
4	<i>sov so</i>	-	warm(er)	redup	s
5	<i>nees nkaum</i>	two-twenty	twenty	phon	w
6	<i>me ntsis</i>	little-few	few, a little bit	orth	w
7	<i>hauv ncoo</i>	head-pillow	pillow	phon	s
8	<i>viav vias</i>	swing-swing	swing (n)	redup	w
9	<i>ib leeg</i>	one-person	alone, single person	ts	w
10	<i>huv si</i>	all-?	all; clean	sesq	s
11	<i>caj npab</i>	noun.class.body.part -arm	arm	b	s
12	<i>tiv thaiv</i>	oppose-protect	to protect, defend	cc	w
13	<i>nab qa</i>	snake-frog	lizard	ts, op	w
14	<i>key cai</i>	path-custom	custom	op	s
15	<i>tsev neeg</i>	house-people	family	op	w
16	<i>pob txha</i>	noun.class.ball-bone	bone	b	s
17	<i>teb chaws</i>	land-place	country	ts	w
18	<i>plab hlaub</i>	stomach-lower.leg	calf of the leg	op	s
19	<i>kaj ntug</i>	bright-sky	dawn, daylight	ts	w
20	<i>qhov rooj</i>	noun.class.hole-door	door, gate	b	s
21	<i>yooj yim</i>	-	easy	mono	s
22	<i>zaub mov</i>	vegetable-rice	food	cc	s

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
23	<i>ko taw</i>	noun.class.handle-foot	foot	b	w
24	<i>ntev ntev</i>	long.long	long	redup	s
25	<i>hauv pliaj</i>	head-forehead	forehead	cc, phon	w
26	<i>phooj ywg</i>	-	friend	mono	s
27	<i>xeeb ntxwv</i>	descendant-NOMINALIZER	grandchild, descendant	b, phon	s
28	<i>tswv yim</i>	-	idea, wisdom	mono	w
29	<i>tub nkeeg</i>	person-tired	lazy, lethargic	ts	s
30	<i>tam sim</i>	-	immediately	mono	w
31	<i>di ncauj</i>	lip-mouth	lip	cc	s
32	<i>dav hlau</i>	hawk-iron	airplane	op	s
33	<i>qhov ntswg</i>	noun.class.hole-nose	nose	b	w
34	<i>niam txiv</i>	mother-father	parents, husband/wife	cc	w
35	<i>ib qho</i>	one-thing	piece, part	ts	s
36	<i>zaj sawv</i>	dragon-rise	rainbow	op	s
37	<i>dab tsi</i>	-	what?	mono	w
38	<i>hiav txwv</i>	sea-NOMINALIZER	sea	b	w
39	<i>ntxoov ntxoo</i>	-	shade, shadow; cloudy	redup	w
40	<i>caj dab</i>	noun.class.body.part -neck	neck	b	w
41	<i>viv ncaus</i>	-	sisters; female cousins	mono	s
42	<i>taj laj</i>	-	market	mono	w
43	<i>khov kho</i>	-	strong	redup	s
44	<i>kab tsib</i>	-	sugar cane	mono	s
45	<i>cheb cheb</i>	sweep-sweep	sweeping	redup	w
46	<i>tshaib plab</i>	hungry-stomach	to be hungry	V-obj	w
47	<i>nom tswv</i>	leader-master	officials, government	cc	s

#	Hmong Daw	Gloss by morpheme	Word gloss	Word type	Group 1 spacing
48	<i>dhia dhia</i>	jump-jump	to keep jumping/running	redup	w
49	<i>sib ceg</i>	RECIPROCAL-scold	to quarrel	b, ts	s
50	<i>me nyuam</i>	small-little	baby, child	cc	s
51	<i>ntiaj teb</i>	surface-earth	world, earth	cc	w
52	<i>kawm ntawv</i>	study-paper	to study	V-obj	s
53	<i>dab tuag</i>	ghost-dead	ugly, sloppy; demon	op	w
54	<i>nkag siab</i>	enter-liver	to understand	op	s
55	<i>qaub ncaug</i>	sour-water	saliva	ts, op	s
56	<i>poj niam</i>	female-mother	woman	cc	w
57	<i>tsov rog</i>	tiger-war	war, battle	op	w
58	<i>nus muag</i>	brother-sister	sibling	cc	s
59	<i>hauj lwm</i>	-	work (n)	mono	w
60	<i>ob tug</i>	two-CLASSIFIER.animate	two people/animals	ts	s
61	<i>pob caus</i>	noun.class.ball-knot	knot	b	w

APPENDIX H

MODELS, RESULTS, AND SPSS SYNTAX

Independent variables considered

Reader-related variables

The following demographic or test-related variables were considered in the models for all tests of Akha and Hmong reading speed:

- Age (in years)
- Gender (coded as 0.5 or -0.5)
- Akha/Hmong Daw reading level (self-reported, 0 to 10)
- Unspaced national language reading level (Thai, Lao, Burmese, or Chinese, depending on the reader; self-reported, 0 to 10)
- Years of schooling
- Years of schooling in Akha/Hmong
- Group assignment (coded as 0.5 or -0.5)

The Akha test considered these additional specific variables:

- Reading frequency in the CAO orthography (times per week)
- Log ratio of reading frequency (in times per week) in the Baptist orthography over the CAO orthography

All Hmong tests also considered these variables:

- Dialect of reader (coded as Hmong Njua = 1, Hmong Daw = 0)
- Hmong Daw speaking level (self-reported, 0 to 10)
- English (or French) reading level (self-reported, 0 to 10)
- Reading frequency in Hmong Daw (times per week)
- Reading frequency in Thai or Lao (times per week)
- Reading frequency in English or French (times per week)
- Spacing style readers said they read most often (word = 1, syllable = -1, both/equal = 0)

- Religion (coded as a nominal categorical variable, P = Protestant, C = Catholic, O = Other)

Both the Hmong stories and the Hmong word list test considered the ratio of the reader's lifetime spent in the US. Hmong stories test also considered the readers' time in seconds on story 3, a test story (story 1 was also a test story, but as the first story, there were more problems and variation with the data).

Word-related variables

The Hmong spacing practices analysis and the word list test both considered the following word-related variables (variables related to morphological type, syntactic type, or the first or second syllable were not considered for the elaborate expressions analysis):

- Number of letters
- Number of letters in the first syllable
- Having a final tone letter on the first syllable (yes = 1, no = 0)
- Having a bound morpheme (yes = 1, no = 0)
- Being a fully reduplicated word (yes = 1, no = 0)
- Number of morphemes (monomorphemic = 1, sesquimorphemic = 1.5, dimorphemic = 2)
- Number of phonological words (tone sandhi words, reduplicated words, and words with spreading nasalization = 1, other = 2)
- Semantically opaque (monomorphemic words and semantically opaque compounds = 1, all others = 0)
- Number-classifier form (yes = 1, no = 0)
- Noun-noun form (yes = 1, no = 0)
- Having a verbal constituent (yes = 1, no = 0)
- Frequency of first syllable in the SCH corpus (number of instances)
- Ratio of instances in the SCH corpus of the first syllable of the word in that target word over total instances (including in other words or on its own)
- Log of instances of word in the SCH corpus
- Number of instances of word in the Catholic Bible
- Number of spelling differences between Hmong Daw and Hmong Njua for target word (counted as number of grapheme substitutions necessary)

The analysis of the UBS NT also considered the number of instances of the word in the UBS NT text. The word list test considered the following variables as well:

- Number of initial consonants in the second syllable
- Ratio of spaced instances to total instances in the SCH corpus (including tone sandhi and dialect variants)
- Ratio of Hmong Njua variant instances to total instances of word
- Ratio of tone sandhi variant instances to total instances of word
- Order presented in the word list (integer from 1 to 62 for US list, 1 to 61 for Thailand list)

Sentence-related variables

All the analyses of the Hmong sentence test data considered the following sentence-related variables:

- Number of syllables
- Letters per syllable
- Syllables per word
- Number of polysyllabic words
- Number of polysyllabic words that readers had not yet seen in test
- Number of polysyllabic words that readers had not yet seen in that particular story
- Ratio of words in the sentence that are found on a list of high frequency words in Hmong Daw (Lewis et al. 2012)
- Number of lines on the screen a sentence took (most were one, some two)
- Having a line break in the middle of a word (yes = 1, no = 0; note that only one sentence had one of these)
- Order of sentence in the test (integer from 1 to 43)
- Set assignment (whether a sentence was word-spaced or syllable-spaced for group 1; coded as a nominal string variable with “1” for set 1, “2” for set 2, and “m” for monosyllabic-only sentences)

All the analyses of sentences that excluded sentences with monosyllabic words only used two additional variables:

- Mean reading speed (in Box-Cox transformed seconds per syllable) on sentences with monosyllabic words only
- Standard deviation of reading speed (Box-Cox seconds/syllable) on sentences with monosyllabic words only

The analysis of sentences with only one new polysyllabic word also considered two other variables:

- Log of instances of target word in the SCH corpus
- Ratio of spaced instances over total instances of target word in the SCH corpus

Meanwhile, the analysis of repeated polysyllabic words (which included both the first and later instances of the words) used the following variable:

- Having a polysyllabic word readers had already seen in test (yes = 1, no = 0)

Spacing, interaction effects, and random effects

All the models for the Hmong sentence test and word list test included spacing style as an independent variable in the model (coded as w = word-spaced, s = syllable-spaced; in the case of Lahu Si, there was also c = “compound-spaced” or free morpheme-spaced), regardless of whether it improved the AICc measure or not. In addition, for these multilevel models, interaction effects between spacing and all other independent variables in the model were considered. Random effects of spacing style and all other variables in the model were also considered, but few were retained unless they made a large contribution to the AICc measure. No models retained a random slope for spacing style. Random effects for reader (coded “Person”) and item (coded “Sentence” or “Word”) were included in multilevel models when they improved the AICc fit. A higher-level random effect for story (“Story”) was included in one case as well.

In general, all the independent variables used in the models for reading speed were standardized (that is, their means were transformed to 0 and their standard deviations to 1) to make coefficients easier to compare, and to simplify interpretation of the intercept. In certain cases, a non-standardized variable was also used, either because its mean was already zero, or in order to give parameters in more practically understandable units.

Variable selection methods

Linear regressions

All the linear regressions in this study used a backward selection method, with the Adjusted R^2 value as the criterion for selection. If removing a variable resulted in an improvement of Adjusted R^2 , or a less than .01 improvement, the variable was

removed. Variables were removed more liberally than suggested by Adjusted R^2 because it is known to underpenalize for model complexity (Fortmann-Roe 2012).

Besides removing variables when their retention did not sufficiently improve Adjusted R^2 , variables were also removed if significant multicollinearity was found. In some cases, multicollinearity was retained if it was not extremely strong and did not affect the spacing style variable or its interactions.

Multilevel models

The multilevel models in this study used a forward selection method, with the AICc value as the criterion for selection. Variables that improved the AICc measure by 1 or more were added. Since AICc numbers were available in the SPSS output, they were used instead of significance or Adjusted R^2 , because variable selection methods based on information theoretic criteria are less prone to the problems of multiple hypothesis testing and stepwise regression (Faraway 2002, Whittingham et al. 2006, Mundry & Nunn 2009, Fortmann-Roe 2012). The use of multilevel models also helps make the model more robust to the problems of variable selection (Gelman et al. 2012). Random effects were generally retained only if they both improved the AICc measure and had a significance level of less than 0.2.

After the forward selection process was completed, a check for multicollinearity sometimes resulted in additional variables being removed.

1. Lahu Si stories multilevel model

This analysis (in Section 3.3) used a multilevel model with crossed random effects for reader and story. Since the random effect for story did not improve the AICc fit, it was removed. Seconds per syllable was the dependent variable, and spacing was the only independent variable considered and used.

The following tables show the SPSS output from comparing word spacing to the other two styles:

```
MIXED SecPerSyl BY Spacing
  /FIXED=Spacing
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(Person) COVTYPE(VC).
```


Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4	87.748	.001
Spacing	2	8	1.833	.221

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.860102	.099546	5.828	8.640	.000	.614770	1.105435
[Spacing = C]	.044118	.073466	8	.601	.565	-.125296	.213531
[Spacing = S]	-.093626	.073466	8	-1.274	.238	-.263039	.075787
[Spacing = W]	0 ^b	0

b. This parameter is set to zero because it is redundant.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.013493	.006747	2.000	.046	.005064	.035951
Intercept [subject Variance = Person]	.036054	.028762	1.254	.210	.007549	.172189

Next is the SPSS output comparing compound spacing to the other two styles:

Fixed Effects

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.904220	.099546	5.828	9.083	.000	.658888	1.149552
[Spacing = S]	-.137744	.073466	8	-1.875	.098	-.307157	.031669
[Spacing = W]	-.044118	.073466	8	-.601	.565	-.213531	.125296
[Spacing = C]	0 ^b	0

b. This parameter is set to zero because it is redundant.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.013493	.006747	2.000	.046	.005064	.035951
Intercept [subject Variance = Person]	.036054	.028762	1.254	.210	.007549	.172189

2. Akha stories regression

The analysis of Akha stories (Section 3.4) used a fairly simple multiple linear regression. The dependent variable was the natural log of the ratio of word-spaced reading speed over the syllable-spaced reading speed (both in seconds per syllable). In other words, a positive number means syllable spacing is faster, and a negative number means word spacing is faster. The independent variables in the model were:

- Age
- Akha reading level
- Years of schooling
- Log ratio of reading frequency in the Baptist orthography over the CAO orthography

Here is the main analysis:

```
REGRESSION
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/DEPENDENT lnWSratio
/METHOD=ENTER ZAge ZAkhardlev ZYrsschool zLnratioorth.
```

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.778	.606	.448	.09893

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.150	4	.038	3.838	.038 ^b
	Residual	.098	10	.010		
	Total	.248	14			

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	.177	.029		6.048	.000
Zscore(Age)	.124	.040	.742	3.127	.011
Zscore(Akhardlev)	-.106	.050	-.476	-2.125	.060
Zscore(Yrsschool)	-.040	.033	-.308	-1.209	.255
Zscore(Lnratioorth)	.098	.034	.736	2.851	.017

Coefficients (continued)

Model	95.0% Confidence Interval for B		Collinearity Statistics	
	Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	.112	.243		
Zscore(Age)	.036	.213	.701	1.426
Zscore(Akhardlev)	-.218	.005	.787	1.271
Zscore(Yrsschool)	-.114	.034	.606	1.650
Zscore(Lnratioorth)	.021	.175	.591	1.691

The analysis was also run including the grouping variable, which did not improve the model:

REGRESSION
 /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
 /DEPENDENT lnWSratio
 /METHOD=ENTER ZAge ZAkhardlev ZYrsschool zLnratioorth Group.

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.783	.613	.397	.10334

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.152	5	.030	2.847	.082 ^b
	Residual	.096	9	.011		
	Total	.248	14			

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	.178	.031		5.801	.000
Zscore(Age)	.126	.042	.751	3.018	.015
Zscore(Akhardlev)	-.104	.052	-.466	-1.983	.079
Zscore(Yrsschool)	-.036	.036	-.277	-.999	.344
Zscore(Lnratioorth)	.093	.038	.702	2.478	.035
group	-.012	.030	-.093	-.406	.694

Coefficients (continued)

Model		95.0% Confidence Interval for B		Collinearity Statistics	
		Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	.108	.247		
	Zscore(Age)	.032	.221	.694	1.440
	Zscore(Akhardlev)	-.223	.015	.779	1.284
	Zscore(Yrsschool)	-.118	.046	.559	1.788
	Zscore(Lnratioorth)	.008	.179	.537	1.862
	group	-.079	.055	.822	1.217

Finally, another analysis was run with a non-standardized log ratio of reading frequency in the two orthographies:

```
REGRESSION
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/DEPENDENT lnWSratio
/METHOD=ENTER ZAge ZAkhardlev ZYrsschool Lnratioorth.
```

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.778	.606	.448	.09893

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.150	4	.038	3.838	.038 ^b
	Residual	.098	10	.010		
	Total	.248	14			

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	.067	.046		1.454	.177
Zscore(Age)	.124	.040	.742	3.127	.011
Zscore(Akhardlev)	-.106	.050	-.476	-2.125	.060
Zscore(Yrsschool)	-.040	.033	-.308	-1.209	.255
Lnratioorth	.095	.033	.736	2.851	.017

Coefficients (continued)

Model		95.0% Confidence Interval for B		Collinearity Statistics		
		Lower Bound	Upper Bound	Tolerance	VIF	
1	(Constant)		-.036	.169		
	Zscore(Age)		.036	.213	.701	1.426
	Zscore(Akhardlev)		-.218	.005	.787	1.271
	Zscore(Yrsschool)		-.114	.034	.606	1.650
	Lnratioorth		.021	.169	.591	1.691

3. Among spacing practices

3.1 SCH corpus

The SCH corpus analysis used a linear regression, with the log ratio of unspaced over spaced instances of the word as the dependent variable. For words that have zero unspaced instances, the negative infinite value of the log ratio was replaced by the following function:

$$f(n) = \ln \left(\mu_{spaced}^{\frac{-1}{n+1}} - 1 \right)$$

where n is the number of total instances (that is, spaced instances) for the word, and μ_{spaced} is the mean ratio of spaced to total instances for all words in the data set, equal to 0.838. This function represents an estimate of the log ratio of unspaced over spaced instances that each word would have in a larger corpus, based on how often words in general are spaced or unspaced in this data set and how many spaced instances occur for a given word.

The independent variables kept in the model were:

- Number of morphemes

- Letters in the first syllable
- Number-classifier form
- Fully reduplicated form
- Ratio of first syllable frequency in target word over total frequency
- Number of Hmong Daw/Hmong Njua grapheme differences

Here is the SPSS syntax and output for the model:

```
REGRESSION
/STATISTICS COEFF R ANOVA COLLIN TOL CI
/DEPENDENT schlnjs
/METHOD=ENTER zmorphemes zlet1stsyl znumclf zfullredup z1stsylinword
zHLspellldiff.
```

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.817	.667	.642	.860466

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	117.377	6	19.563	26.422	.000
	Residual	58.492	79	.740		
	Total	175.868	85			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.193	.093		-23.619	.000
	Zscore(morphemes)	-.306	.104	-.215	-2.954	.004
	Zscore: let 1st syl	-.377	.109	-.264	-3.446	.001
	Zscore(numclf)	-.618	.106	-.434	-5.849	.000
	Zscore(fullredup)	-.581	.104	-.380	-5.562	.000
	Zscore: 1st syl ratio in.word/tot	.421	.105	.294	3.991	.000
	Zscore: HD-HL # grapheme differences	.242	.099	.169	2.451	.016

Coefficients (continued)

Model	95.0% Confidence Interval for B		Collinearity Statistics	
	Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	-2.378	-2.008		
Zscore(morphemes)	-.512	-.100	.796	1.257
Zscore: let 1st syl	-.595	-.159	.717	1.395
Zscore(numclf)	-.828	-.407	.764	1.308
Zscore(fullredup)	-.789	-.373	.900	1.111
Zscore: 1st syl ratio in.word/tot	.211	.631	.776	1.289
Zscore: HD-HL # grapheme differences	.046	.439	.885	1.130

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Zscore(morphemes)	Zscore: let 1st syl
1	1	1.548	1.000	.00	.12	.07
	2	1.467	1.027	.00	.10	.13
	3	1.100	1.186	.00	.00	.11
	4	.999	1.245	.97	.01	.00
	5	.836	1.361	.02	.15	.00
	6	.656	1.536	.00	.35	.08
	7	.394	1.982	.00	.28	.61

Collinearity Diagnostics^a

Model	Dimension	Variance Proportions			
		Zscore(numclf)	Zscore(fullredup)	Zscore: 1st syl ratio in.word/tot	Zscore: HD-HL # grapheme differences
1	1	.14	.00	.16	.01
	2	.07	.17	.06	.00
	3	.01	.03	.00	.63
	4	.00	.00	.00	.00
	5	.14	.67	.02	.02
	6	.23	.00	.46	.07
	7	.40	.12	.30	.26

3.2 UBS New Testament

The UBS NT analysis (Section 4.4.3) used a logistic regression, with the spacing style of each test word found in the NT as the dependent variable (coded as 1 = unspaced, 0 = spaced).

The independent variables kept in the model were:

- Number of morphemes
- Log of word frequency in the SCH corpus
- Letters in the first syllable
- Number-classifier form
- Ratio of first syllable frequency in target word over total frequency
- Number of Hmong Daw/Hmong Njua spelling differences

Here is the SPSS syntax and output for the model:

```
LOGISTIC REGRESSION VARIABLES UBSjtotratio
/METHOD=ENTER zHDHL#graphemedifferences z@1stsylratioin.wordtot
zNNforms
/PRINT=GOODFIT CI(95) corr.
```

Logistic Regression

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nageikerke R Square
1	46.711	.432	.577

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.079	8	.929

Classification Table

Observed		Predicted		
		UBSjtotratio		Percentage Correct
		.0	1.0	
Step 1	UBSjtotratio .0	22	6	78.6
	1.0	6	23	79.3
Overall Percentage				78.9

a. The cut value is .500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ZHDHL#graphemedifferences	1.259	.515	5.986	1	.014	3.522	1.285	9.656
Z@1stsylratioin.wordtot	1.862	.541	11.868	1	.001	6.437	2.232	18.570
ZNNforms	1.220	.420	8.455	1	.004	3.388	1.488	7.712
Constant	.291	.381	.584	1	.445	1.338		

Correlation Matrix

		Constant	ZHDHL#grapheme differences	Z@1stsylratioin.wordtot	ZNNforms
Step	Constant	1.000	.190	.225	.039
1	ZHDHL#graphemedifferences	.190	1.000	.311	.340
	Z@1stsylratioin.wordtot	.225	.311	1.000	.471
	ZNNforms	.039	.340	.471	1.000

4. US Hmong stories regression

4.1 Reading speed

The Hmong stories test analysis (Section 5.2.7) used a linear regression, with the log ratio of word-spaced reading speed over syllable-spaced reading speed in seconds/syllable as the dependent variable. The only independent variable kept in the model was group assignment.

Here is the SPSS syntax and output for the model:

```
REGRESSION
/STATISTICS COEFF CI(95) R ANOVA COLLIN TOL
/DEPENDENT lnWS
/METHOD=ENTER Group.
```

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.518	.268	.232	.17974

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.237	1	.237	7.340	.014 ^b
	Residual	.646	20	.032		
	Total	.883	21			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.026	.038		.673	.509
	Group	.104	.038	.518	2.709	.014

Coefficients (continued)

Model		95.0% Confidence Interval for B		Collinearity Statistics	
		Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)				
	Group	-.054	.106	1.000	1.000

4.2 Comprehension

The comprehension analysis for the stories test data (Section 5.2.8) used a multilevel ordinal logistic model, with the comprehension score as the dependent variable (coded as 0 = incorrect/no answer; 0.5 = partially correct on sentences with multi-part answers; and 1 = correct). Since there were two questions per story, the actual comprehension values varied from 1 to 2.

Spacing was the only independent variable included in the model. The random effects for reader and for story did not contribute to the model and were removed.

Here is the SPSS syntax and output for the model:

Descriptives

Spacing = s

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	22	1.0	2.0	1.917	.2399
Valid N (listwise)	22				

Spacing = w

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	22	1.0	2.0	1.955	.2132
Valid N (listwise)	22				

GENLINMIXED

```

/ DATA_STRUCTURE SUBJECTS=Person*Story
/ FIELDS TARGET=Comp TRIALS=NONE OFFSET=NONE
/ TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
/ FIXED_EFFECTS=Spacing USE_INTERCEPT=TRUE
/ BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING
INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
/ EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
    
```

Generalized Linear Mixed Models

Fixed Coefficients

Target: Comp
Reference Category: 2.0

Model Term	Coefficient ▼	Std. Error	t	Sig.
1	-.3774	0.956	-3.906	.060
Threshold for Comp*				
1.5	-.3341	0.901	-3.709	.066
1.66666666666667	-.3020	0.862	-3.503	.073
Spacing*s	-.1147	1.007	-1.139	.373
Spacing*w	0 ^a			

Probability distribution: Multinomial
Link function: Cumulative logit

^aThis coefficient is set to zero because it is redundant.

Fixed Coefficients

Target:Comp
Reference Category:2.0

Model Term	95% Confidence Interval	
	Lower	Upper
1	-7.931	0.383
Threshold for Comp=		
1.5	-7.218	0.535
1.666666666666667	-6.728	0.689
Spacing=s	-5.481	3.187
Spacing=w		

Probability distribution:Multinomial
Link function:Cumulative logit

^aThis coefficient is set to zero because it is redundant.

5. Hmong sentence test, Thailand

5.1 All sentences

The analysis of the sentence test for readers in Thailand (Section 5.3.6.1) used a multilevel model, with the Box-Cox transformed reading speed in seconds per syllable as the dependent variable.

The independent variables kept in the model were:

- Spacing
- Hmong Daw reading level
- Hmong Daw reading frequency
- Thai/Lao reading frequency
- Gender
- Average number of letters per syllable
- Number of syllables

- Ratio of words on the high frequency word list for Hmong Daw (Lewis et al. 2012)
- Order of sentence in the test
- Spacing * Number of syllables
- Spacing * Hmong Daw reading level
- Spacing * Letters per syllable

Here is the SPSS syntax and output for the model:

```
MIXED BoxSecSyl BY Spacing WITH zHDrdlevel Gender zLetSyl
zratiohighfreqwords ZSentOrder zSyl HDrdxwk TLrdxwk
/CRITERIA=MXITER(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zHDrdlevel Gender zLetSyl zratiohighfreqwords zSyl
ZSentOrder Spacing*zSyl Spacing*zHDrdlevel Spacing*zLetSyl
/PRINT=SOLUTION TESTCOV CORB
/RANDOM=INTERCEPT | SUBJECT(Person) COVTYPE(VC)
/RANDOM=HDrdxwk TLrdxwk | SUBJECT(Sentence) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	611.561
Akaike's Information Criterion (AIC)	619.561
Hurvich and Tsai's Criterion (AICC)	619.608
Bozdogan's Criterion (CAIC)	642.537
Schwarz's Bayesian Criterion (BIC)	638.537

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	30.074	16.106	.000
Spacing	2	281.187	1.991	.138
zHDrdlevel	1	18.176	12.690	.002
Gender	1	18.008	8.046	.011
zLetSyl	1	152.814	9.190	.003
zratiohighfreqwords	1	153.750	16.500	.000
zSyl	1	157.108	2.302	.131
ZSentOrder	1	157.167	9.422	.003
Spacing * zSyl	2	306.422	4.584	.011
Spacing * zHDrdlevel	2	797.127	4.177	.016
Spacing * zLetSyl	2	292.520	3.399	.035

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.323073	.080691	44.203	-4.004	.000	-.485675	-.160471
[Spacing=m]	.121775	.076875	164.386	1.584	.115	-.030015	.273565
[Spacing=s]	-.025942	.031559	809.926	-.822	.411	-.087888	.036004
[Spacing=w]	0 ^b	0
zHDrlevel	-.235586	.064221	20.875	-3.668	.001	-.369190	-.101982
Gender	.179315	.063217	18.008	2.836	.011	.046505	.312124
zLetSyl	.145528	.042871	188.696	3.395	.001	.060961	.230095
zratiohighfreqwords	-.156282	.038474	153.750	-4.062	.000	-.232288	-.080277
zSyl	-.001962	.042233	206.003	-.046	.963	-.085226	.081302
ZSentOrder	-.089944	.029302	157.167	-3.070	.003	-.147822	-.032067
[Spacing=m] * zSyl	-.161361	.059520	174.268	-2.711	.007	-.278833	-.043888
[Spacing=s] * zSyl	.021548	.030422	809.354	.708	.479	-.038168	.081263
[Spacing=w] * zSyl	0 ^b	0
[Spacing=m] * zHDrlevel	.055297	.025343	809.426	2.182	.029	.005550	.105043
[Spacing=s] * zHDrlevel	-.011496	.028035	767.972	-.410	.682	-.066531	.043539
[Spacing=w] * zHDrlevel	0 ^b	0
[Spacing=m] * zLetSyl	-.075416	.063827	167.716	-1.182	.239	-.201425	.050592
[Spacing=s] * zLetSyl	-.076475	.029987	809.290	-2.550	.011	-.135337	-.017614
[Spacing=w] * zLetSyl	0 ^b	0

b. This parameter is set to zero because it is redundant.

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=m]	[Spacing=s]	[Spacing=w]	zHDrlevel
Intercept	1	-.434	-.203	. ^b	-.028
[Spacing=m]	-.434	1	.227	. ^b	-.001
[Spacing=s]	-.203	.227	1	. ^b	.003
[Spacing=w]	. ^b	. ^b	. ^b	1	. ^b
zHDrlevel	-.028	-.001	.003	. ^b	1
Gender	.181	-.001	-.001	. ^b	-.084
zLetSyl	-.092	-.020	.078	. ^b	.022
zratiohighfreqwords	.239	-.562	-.045	. ^b	.015
zSyl	-.248	.346	.153	. ^b	-.015
ZSentOrder	.067	-.141	.023	. ^b	-.002
[Spacing=m] * zSyl	.172	-.077	-.099	. ^b	.008
[Spacing=s] * zSyl	.061	-.063	-.338	. ^b	.010

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=m]	[Spacing=s]	[Spacing=w]	zHDrlevel
[Spacing=w] * zSyl
[Spacing=m] * zHDrlevel	.011	-.021	-.008	.	-.242
[Spacing=s] * zHDrlevel	-.011	.037	-.022	.	-.215
[Spacing=w] * zHDrlevel
[Spacing=m] * zLetSyl	.074	.185	-.056	.	-.015
[Spacing=s] * zLetSyl	.063	-.051	-.354	.	-.013
[Spacing=w] * zLetSyl

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	Gender	zLetSyl	zratiohighfreqwords	zSyl	ZSentOrder
Intercept	.181	-.092	.239	-.248	.067
[Spacing=m]	-.001	-.020	-.562	.346	-.141
[Spacing=s]	-.001	.078	-.045	.153	.023
[Spacing=w]
zHDrlevel	-.084	.022	.015	-.015	-.002
Gender	1	.004	.000	-.001	.001
zLetSyl	.004	1	.360	.143	.021
zratiohighfreqwords	.000	.360	1	-.298	.114
zSyl	-.001	.143	-.298	1	.201
ZSentOrder	.001	.021	.114	.201	1
[Spacing=m] * zSyl	.001	-.128	.151	-.638	.076
[Spacing=s] * zSyl	.004	-.118	-.018	-.323	-.031
[Spacing=w] * zSyl
[Spacing=m] * zHDrlevel	.002	-.057	-.041	.037	.003
[Spacing=s] * zHDrlevel	.003	-.082	-.066	.055	.011
[Spacing=w] * zHDrlevel
[Spacing=m] * zLetSyl	-.003	-.654	-.193	-.119	-.045
[Spacing=s] * zLetSyl	-.010	-.382	-.004	-.116	-.021
[Spacing=w] * zLetSyl

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=m] * zSyl	[Spacing=s] * zSyl	[Spacing=w] * zSyl	[Spacing=m] * zHDrlevel
Intercept	.172	.061	.	.011
[Spacing=m]	-.077	-.063	.	-.021
[Spacing=s]	-.099	-.338	.	-.008
[Spacing=w]

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=m] * zSyl	[Spacing=s] * zSyl	[Spacing=w] * zSyl	[Spacing=m] * zHDrlevel
zHDrlevel	.008	.010	. ^b	-.242
Gender	.001	.004	. ^b	.002
zLetSyl	-.128	-.118	. ^b	-.057
zratiohighfreqwords	.151	-.018	. ^b	-.041
zSyl	-.638	-.323	. ^b	.037
ZSentOrder	.076	-.031	. ^b	.003
[Spacing=m] * zSyl	1	.223	. ^b	-.024
[Spacing=s] * zSyl	.223	1	. ^b	-.023
[Spacing=w] * zSyl	. ^b	. ^b	. ^b	. ^b
[Spacing=m] * zHDrlevel	-.024	-.023	. ^b	1
[Spacing=s] * zHDrlevel	-.031	-.009	. ^b	.542
[Spacing=w] * zHDrlevel	. ^b	. ^b	. ^b	. ^b
[Spacing=m] * zLetSyl	.153	.079	. ^b	.033
[Spacing=s] * zLetSyl	.078	.287	. ^b	.030
[Spacing=w] * zLetSyl	. ^b	. ^b	. ^b	. ^b

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=s] * zHDrlevel	[Spacing=w] * zHDrlevel	[Spacing=m] * zLetSyl	[Spacing=s] * zLetSyl
Intercept	-.011	. ^b	.074	.063
[Spacing=m]	.037	. ^b	.185	-.051
[Spacing=s]	-.022	. ^b	-.056	-.354
[Spacing=w]	. ^b	. ^b	. ^b	. ^b
zHDrlevel	-.215	. ^b	-.015	-.013
Gender	.003	. ^b	-.003	-.010
zLetSyl	-.082	. ^b	-.654	-.382
zratiohighfreqwords	-.066	. ^b	-.193	-.004
zSyl	.055	. ^b	-.119	-.116
ZSentOrder	.011	. ^b	-.045	-.021
[Spacing=m] * zSyl	-.031	. ^b	.153	.078
[Spacing=s] * zSyl	-.009	. ^b	.079	.287
[Spacing=w] * zSyl	. ^b	. ^b	. ^b	. ^b
[Spacing=m] * zHDrlevel	.542	. ^b	.033	.030
[Spacing=s] * zHDrlevel	1	. ^b	.051	.016
[Spacing=w] * zHDrlevel	. ^b	. ^b	. ^b	. ^b
[Spacing=m] * zLetSyl	.051	. ^b	1	.258
[Spacing=s] * zLetSyl	.016	. ^b	.258	1
[Spacing=w] * zLetSyl	. ^b	. ^b	. ^b	. ^b

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.087617	.004504	19.452	.000	.079219	.096905
Intercept [subject = Variance Person]	.076318	.026175	2.916	.004	.038966	.149475
HDrdxwk [subject = Variance Sentence]	.000480	.000210	2.291	.022	.000204	.001130
TLrdxwk [subject = Variance Sentence]	.001053	.000280	3.763	.000	.000625	.001772

Correlation Matrix for Estimates of Covariance Parameters

Parameter	Residual	Intercept [subject = Person]	HDrdxwk [subject = Sentence]	TLrdxwk [subject = Sentence]
		Variance	Variance	Variance
Residual	1	-.006	-.092	-.013
Intercept [subject = Variance Person]	-.006	1	.004	.003
HDrdxwk [subject = Variance Sentence]	-.092	.004	1	-.111
TLrdxwk [subject = Variance Sentence]	-.013	.003	-.111	1

5.2 Sentences with polysyllabic words only

For this cross-section of the sentence test data (Section 5.3.6.2), the independent variables kept in the model were:

- Spacing
- Average number of letters per syllable
- Number of syllables
- Set assignment of sentence
- Reader's mean Box-Cox reading speed on monosyllabic sentences
- Spacing * Letters per syllable

This model also retained a random effect for story. Here is the SPSS syntax and output for the model:

```
MIXED BoxSecSyl BY Spacing WITH Set zMonoMean zLetSyl Syl
/CRITERIA=MXITER(200) MXSTEP(200)
```

```

/METHOD=ML
/FIXED=Spacing zLetSyl Spacing*zLetSyl zMonoMean Set
/PRINT TESTCOV SOLUTION corb
/RANDOM=Syl | SUBJECT(Person) COVTYPE(VC)
/RANDOM=INTERCEPT | SUBJECT(Story) COVTYPE(VC)
/RANDOM=INTERCEPT | SUBJECT(Story*Sentence) COVTYPE(VC).

```

Mixed Model Analysis

Information Criteria

-2 Log Likelihood	234.172
Akaike's Information Criterion (AIC)	254.172
Hurvich and Tsai's Criterion (AICC)	254.646
Bozdogan's Criterion (CAIC)	305.805
Schwarz's Bayesian Criterion (BIC)	295.805

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4.238	5.562	.074
Spacing	1	431.571	1.717	.191
zLetSyl	1	21.425	12.641	.002
Spacing * zLetSyl	1	449.130	7.539	.006
zMonoMean	1	46.051	250.274	.000
Set	1	20.587	6.599	.018

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.229453	.105249	4.368	-2.180	.089	-.512217	.053310
[Spacing=s]	-.033821	.025808	431.571	-1.310	.191	-.084546	.016905
[Spacing=w]	0 ^b	0
zLetSyl	.215785	.052007	24.596	4.149	.000	.108586	.322985
[Spacing=s] * zLetSyl	-.074282	.027053	449.130	-2.746	.006	-.127447	-.021116
[Spacing=w] * zLetSyl	0 ^b	0
zMonoMean	.355690	.022484	46.051	15.820	.000	.310435	.400946
Set	-.247027	.096163	20.587	-2.569	.018	-.447254	-.046801

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zLetSyl	[Spacing=s] * zLetSyl
Intercept	1	-.122	.	-.033	.003
[Spacing=s]	-.122	1	.	-.001	.006
[Spacing=w]	.	.	1	.	.
zLetSyl	-.033	-.001	.	1	.
[Spacing=s] * zLetSyl	.003	.006	.	.	1

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zLetSyl	[Spacing=s] * zLetSyl
zLetSyl	-.033	-.001	. ^b	1	-.258
[Spacing=s] * zLetSyl	.003	.006	. ^b	-.258	1
[Spacing=w] * zLetSyl	. ^b	. ^b	. ^b	. ^b	. ^b
zMonoMean	-.005	.009	. ^b	-.002	.007
Set	-.010	.007	. ^b	-.360	-.002

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=w] * zLetSyl	zMonoMean	Set
Intercept	. ^b	-.005	-.010
[Spacing=s]	. ^b	.009	.007
[Spacing=w]	. ^b	. ^b	. ^b
zLetSyl	. ^b	-.002	-.360
[Spacing=s] * zLetSyl	. ^b	.007	-.002
[Spacing=w] * zLetSyl	. ^b	. ^b	. ^b
zMonoMean	. ^b	1	.000
Set	. ^b	.000	1

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.078610	.005357	14.674	.000	.068782	.089843
Syl [subject = Person] Variance	.000109	4.477801E-5	2.427	.015	4.847254E-5	.000244
Intercept [subject = Story] Variance	.033605	.030028	1.119	.263	.005832	.193643
Intercept [subject = Story * Sentence] Variance	.041050	.014297	2.871	.004	.020742	.081242

Correlation Matrix for Estimates of Covariance Parameters

Parameter	Residual	Syl [subject = Person]	Intercept [subject = Story]	Intercept [subject = Story * Sentence]
		Variance	Variance	Variance
Residual	1	-.052	-.002	-.019
Syl [subject = Person] Variance	-.052	1	.013	.004
Intercept [subject = Story] Variance	-.002	.013	1	-.098
Intercept [subject = Story * Sentence] Variance	-.019	.004	-.098	1

5.3 Sentences with new polysyllabic words only

For this cross-section of the sentence test data (Section 5.3.6.3), the independent variables kept in the model were:

- Spacing
- Average number of letters per syllable
- Average number of syllables per word
- Ratio of words on the high frequency word list for Hmong Daw (Lewis et al. 2012)
- Reader’s mean Box-Cox reading speed on monosyllabic sentences
- Gender
- Spacing * Letters per syllable

Here is the SPSS syntax and output for the model:

```
MIXED BoxSecSyl BY Spacing WITH zMonoMean Gender zLetSyl zSylWord
zratiohighfreqwords
/CRITERIA=MXITER(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zMonoMean zLetSyl zSylWord zratiohighfreqwords
Spacing*zLetSyl
/PRINT=SOLUTION TESTCOV corb
/RANDOM=INTERCEPT | SUBJECT(Person) COVTYPE(VC)
/RANDOM=INTERCEPT | SUBJECT(Sentence) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	184.790
Akaike’s Information Criterion (AIC)	190.790
Hurvich and Tsai’s Criterion (AICC)	190.864
Bozdogan’s Criterion (CAIC)	205.150
Schwarz’s Bayesian Criterion (BIC)	202.150

The information criteria are displayed in smaller-is-better form.

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	16.905	9.384	.007
Spacing	1	298.941	4.458	.036
zMonoMean	1	19.249	185.295	.000
zLetSyl	1	12.949	1.224	.289

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
ZSylWord	1	12.999	12.134	.004
zratiohighfreqwords	1	12.896	11.833	.004
Spacing * zLetSyl	1	308.884	4.366	.037

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.129604	.055501	20.005	-2.335	.030	-.245376	-.013832
[Spacing=s]	-.066823	.031650	298.941	-2.111	.036	-.129108	-.004538
[Spacing=w]	0 ^b	0
zMonoMean	.348055	.025569	19.249	13.612	.000	.294585	.401525
zLetSyl	.089753	.054486	15.246	1.647	.120	-.026217	.205724
ZSylWord	-.231612	.066489	12.999	-3.483	.004	-.375255	-.087969
zratiohighfreqwords	-.240665	.069964	12.896	-3.440	.004	-.391936	-.089393
[Spacing=s] * zLetSyl	-.063801	.030533	308.884	-2.090	.037	-.123881	-.003721
[Spacing=w] * zLetSyl	0 ^b	0

b. This parameter is set to zero because it is redundant.

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zMonoMean	zLetSyl
Intercept	1	-.284	.	-.005	-.212
[Spacing=s]	-.284	1	.	.003	.061
[Spacing=w]	.	.	1	.	.
zMonoMean	-.005	.003	.	1	-.005
zLetSyl	-.212	.061	.	-.005	1
ZSylWord	.044	.001	.	-.004	-.150
zratiohighfreqwords	-.064	.004	.	-.002	.310
[Spacing=s] * zLetSyl	.064	-.205	.	.012	-.280
[Spacing=w] * zLetSyl	.	.	1	.	.

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	ZSylWord	zratiohighfreqwords	[Spacing=s]*zLetSyl	[Spacing=w]*zLetSyl
Intercept	.044	-.064	.064	.
[Spacing=s]	.001	.004	-.205	.
[Spacing=w]	.	.	.	1
zMonoMean	-.004	-.002	.012	.
zLetSyl	-.150	.310	-.280	.

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	ZSylWord	zratiohighfreqwords	[Spacing=s]*zLetSyl	[Spacing=w]*zLetSyl
ZSylWord	1	.584	-.001	b
zratiohighfreqwords	.584	1	-.006	b
[Spacing=s] * zLetSyl	-.001	-.006	1	b
[Spacing=w] *	b	b	b	1
zLetSyl				b

b. The correlation is system missing because it is associated with a redundant parameter.

Covariance Parameters

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
						Residual	.079359
Intercept [subject = Person]	Variance	.008872	.004539	1.955	.051	.003255	.024185
Intercept [subject = Sentence]	Variance	.034711	.015249	2.276	.023	.014674	.082112

Correlation Matrix for Estimates of Covariance Parameters

Parameter		Residual	Intercept [subject = Person]	Intercept [subject = Sentence]
			Variance	Variance
Residual		1	-.093	-.025
Intercept [subject = Person]	Variance	-.093	1	.013
Intercept [subject = Sentence]	Variance	-.025	.013	1

5.4 Sentences with a single new polysyllabic word only

For this cross-section of the sentence test data (Section 5.3.6.4), the independent variables kept in the model were:

- Spacing
- Average number of letters per syllable
- Average number of syllables per word
- Ratio of words on the high frequency word list for Hmong Daw (Lewis et al. 2012)
- Reader's mean Box-Cox reading speed on monosyllabic sentences
- Spacing * Letters per syllable

Here is the SPSS syntax and output for the model:

```
MIXED BoxSecSyl BY Spacing WITH zratiohighfreqwords zMonoMean zLetSyl
zSylWord
/CRITERIA=MXITER(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zLetSyl Spacing*zLetSyl zratiohighfreqwords zMonoMean
zSylWord
/PRINT TESTCOV SOLUTION CORB
/RANDOM=INTERCEPT | SUBJECT(Sentence) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	103.045
Akaike's Information Criterion (AIC)	107.045
Hurvich and Tsai's Criterion (AICC)	107.109
Bozdogan's Criterion (CAIC)	115.549
Schwarz's Bayesian Criterion (BIC)	113.549

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	5.867	6.920	.040
Spacing	1	185.214	5.516	.020
zLetSyl	1	5.864	3.901	.097
Spacing * zLetSyl	1	184.959	6.956	.009
zratiohighfreqwords	1	5.980	28.734	.002
zMonoMean	1	185.119	289.342	.000
ZSylWord	1	6.249	11.085	.015

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.060167	.045965	9.146	-1.309	.222	-.163895	.043562
[Spacing=s]	-.096014	.040881	185.214	-2.349	.020	-.176665	-.015362
[Spacing=w]	0 ^b	0
zLetSyl	.169747	.059450	8.461	2.855	.020	.033943	.305551
[Spacing=s] * zLetSyl	-.125264	.047495	184.959	-2.637	.009	-.218966	-.031561
[Spacing=w] * zLetSyl	0 ^b	0
zratiohighfreqwords	-.340939	.063603	5.980	-5.360	.002	-.496697	-.185181
zMonoMean	.345785	.020328	185.119	17.010	.000	.305680	.385890
ZSylWord	-.213475	.064117	6.249	-3.329	.015	-.368861	-.058090

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zLetSyl	[Spacing=s] * zLetSyl
Intercept	1	-.447	. ^b	-.176	.065
[Spacing=s]	-.447	1	. ^b	.065	-.126
[Spacing=w]	. ^b	. ^b	1	. ^b	. ^b
zLetSyl	-.176	.065	. ^b	1	-.410
[Spacing=s] * zLetSyl	.065	-.126	. ^b	-.410	1
[Spacing=w] * zLetSyl	. ^b	. ^b	. ^b	. ^b	. ^b
zratiohighfreqwords	.015	.002	. ^b	.098	-.007
zMonoMean	-.003	.007	. ^b	.003	.017
ZSylWord	.083	-.009	. ^b	-.225	-.002

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=w] * zLetSyl	zratiohighfreqwords	zMonoMean	ZSylWord
Intercept	. ^b	.015	-.003	.083
[Spacing=s]	. ^b	.002	.007	-.009
[Spacing=w]	. ^b	. ^b	. ^b	. ^b
zLetSyl	. ^b	.098	.003	-.225
[Spacing=s] * zLetSyl	. ^b	-.007	.017	-.002
[Spacing=w] * zLetSyl	. ^b	. ^b	. ^b	. ^b
zratiohighfreqwords	. ^b	1	-.013	.711
zMonoMean	. ^b	-.013	1	-.022
ZSylWord	. ^b	.711	-.022	1

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.081150	.008439	9.616	.000	.066187	.099496
Intercept [subject = Sentence] Variance	.012212	.009492	1.287	.198	.002662	.056029

5.5 Sentences with repeated polysyllabic words only

For this cross-section of the sentence test data (5.3.6.5), the independent variables kept in the model were:

- Spacing
- Average number of letters per syllable
- Ratio of words on the high frequency word list for Hmong Daw (Lewis et al. 2012)

- Repeated instance of a word
- Reader's mean Box-Cox reading speed on monosyllabic sentences
- Spacing * Repeated instance of a word

Here is the SPSS syntax and output for the model when the repeated variable (Rep) is standardized, giving the overall significance for spacing in these sentences:

```
MIXED BoxSecSyl BY Spacing WITH zMonoMean zLetSyl zratiohighfreqwords
zRep
/CRITERIA=MXITER(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zMonoMean zLetSyl zratiohighfreqwords zRep Spacing*zRep
/PRINT TESTCOV SOLUTION CORB
/RANDOM=INTERCEPT | SUBJECT(Person) COVTYPE(VC)
/RANDOM=INTERCEPT | SUBJECT(Sentence) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	105.636
Akaike's Information Criterion (AIC)	111.636
Hurvich and Tsai's Criterion (AICC)	111.761
Bozdogan's Criterion (CAIC)	124.470
Schwarz's Bayesian Criterion (BIC)	121.470

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	8.370	2.515	.150
Spacing	1	186.260	.059	.808
zMonoMean	1	18.880	134.579	.000
zLetSyl	1	6.029	18.086	.005
zratiohighfreqwords	1	6.099	24.335	.003
ZRep	1	6.079	24.350	.003
Spacing * ZRep	1	172.869	4.230	.041

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.105569	.072978	9.922	-1.447	.179	-.268349	.057210
[Spacing=s]	-.010577	.043402	186.260	-.244	.808	-.096201	.075046
[Spacing=w]	0 ^b	0
zMonoMean	.387852	.033433	18.880	11.601	.000	.317846	.457859

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
zLetSyl	.359118	.084443	6.029	4.253	.005	.152736	.565501
zratiohighfreqwords	-.389157	.078888	6.099	-4.933	.003	-.581433	-.196881
ZRep	-.271224	.050639	8.174	-5.356	.001	-.387566	-.154882
[Spacing=s] * ZRep	.078467	.038153	172.869	2.057	.041	.003161	.153773
[Spacing=w] * ZRep	0 ^b	0

b. This parameter is set to zero because it is redundant.

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zMonoMean	zLetSyl
Intercept	1	-.288	.	-.017	.555
[Spacing=s]	-.288	1	.	.039	.008
[Spacing=w]	.	.	1	.	.
zMonoMean	-.017	.039	.	1	.002
zLetSyl	.555	.008	.	.002	1
zratiohighfreqwords	-.469	-.002	.	-.001	-.261
ZRep	-.083	-.016	.	.001	.055
[Spacing=s] * ZRep	-.014	.040	.	-.010	-.003
[Spacing=w] * ZRep	.	.	1	.	.

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	zratiohighfreqwords	ZRep	[Spacing=s] * ZRep	[Spacing=w] * ZRep
Intercept	-.469	-.083	-.014	.
[Spacing=s]	-.002	-.016	.040	.
[Spacing=w]	.	.	.	1
zMonoMean	-.001	.001	-.010	.
zLetSyl	-.261	.055	-.003	.
zratiohighfreqwords	1	.342	-.004	.
ZRep	.342	1	-.372	.
[Spacing=s] * ZRep	-.004	-.372	1	.
[Spacing=w] * ZRep	.	.	.	1

b. The correlation is system missing because it is associated with a redundant parameter.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.072636	.007859	9.242	.000	.058756	.089795
Intercept [subject = Person] Variance	.016292	.007870	2.070	.038	.006321	.041991
Intercept [subject = Sentence] Variance	.014947	.010666	1.401	.161	.003691	.060528

Correlation Matrix for Estimates of Covariance Parameters

Parameter	Residual	Intercept [subject = Person] Variance	Intercept [subject = Sentence] Variance
		Variance	Variance
Residual	1	-.109	-.036
Intercept [subject = Person] Variance	-.109	1	.007
Intercept [subject = Sentence] Variance	-.036	.007	1

Here is the SPSS syntax and output for the model when the repeated variable (Rep) is not standardized, giving the predicted parameters for spacing on the first instance of repeated words:

```
MIXED BoxSecSyl BY Spacing WITH zMonoMean zLetSyl zratiohighfreqwords
Rep
/CRITERIA=MXITER(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zMonoMean zLetSyl zratiohighfreqwords Rep Spacing*Rep
/PRINT TESTCOV SOLUTION CORB
/RANDOM=INTERCEPT | SUBJECT(Person) COVTYPE(VC)
/RANDOM=INTERCEPT | SUBJECT(Sentence) COVTYPE(VC).
```

Mixed Model Analysis

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	7.242	3.417	.106
Spacing	1	183.894	2.937	.088
zMonoMean	1	18.880	134.579	.000
zLetSyl	1	6.029	18.086	.005

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
zratiohighfreqwords	1	6.099	24.335	.003
Rep	1	6.079	24.350	.003
Spacing * Rep	1	172.869	4.230	.041

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.229908	.100023	8.798	2.299	.048	.002845	.456971
[Spacing=s]	-.107634	.062809	183.894	-1.714	.088	-.231553	.016286
[Spacing=w]	0 ^b	0
zMonoMean	.387852	.033433	18.880	11.601	.000	.317846	.457859
zLetSyl	.359118	.084443	6.029	4.253	.005	.152736	.565501
zratiohighfreqwords	-.389157	.078888	6.099	-4.933	.003	-.581433	-.196881
Rep	-.553674	.103374	8.174	-5.356	.001	-.791173	-.316174
[Spacing=s] * Rep	.160182	.077885	172.869	2.057	.041	.006454	.313911
[Spacing=w] * Rep	0 ^b	0

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval
					Lower Bound
Residual	.072636	.007859	9.242	.000	.058756
Intercept [subject = Person] Variance	.016292	.007870	2.070	.038	.006321
Intercept [subject = Sentence] Variance	.014947	.010666	1.401	.161	.003691

Estimates of Covariance Parameters

Parameter	95% Confidence Interval
	Upper Bound
Residual	.089795
Intercept [subject = Person] Variance	.041991
Intercept [subject = Sentence] Variance	.060528

Here is the SPSS syntax and output for the model when we use the opposite of Rep (called "New"), giving the predicted parameters for spacing on repeated instances of repeated words:

```

MIXED BoxSecSyl BY Spacing WITH zMonoMean zLetSyl zratiohighfreqwords
New
/CRITERIA=MXITER(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zMonoMean zLetSyl zratiohighfreqwords New Spacing*New
/PRINT TESTCOV SOLUTION CORB
/RANDOM=INTERCEPT | SUBJECT(Person) COVTYPE(VC)
/RANDOM=INTERCEPT | SUBJECT(Sentence) COVTYPE(VC).

```

Mixed Model Analysis

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	7.936	15.284	.005
Spacing	1	189.553	.941	.333
zMonoMean	1	18.880	134.579	.000
zLetSyl	1	6.029	18.086	.005
zratiohighfreqwords	1	6.099	24.335	.003
New	1	6.079	24.350	.003
Spacing * New	1	172.869	4.230	.041

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.323766	.080586	9.958	-4.018	.002	-.503424	-.144107
[Spacing=s]	.052549	.054164	189.553	.970	.333	-.054293	.159390
[Spacing=w]	0 ^b	0
zMonoMean	.387852	.033433	18.880	11.601	.000	.317846	.457859
zLetSyl	.359118	.084443	6.029	4.253	.005	.152736	.565501
zratiohighfreqwords	-.389157	.078888	6.099	-4.933	.003	-.581433	-.196881
New	.553674	.103374	8.174	5.356	.001	.316174	.791173
[Spacing=s] * New	-.160182	.077885	172.869	-2.057	.041	-.313911	-.006454
[Spacing=w] * New	0 ^b	0

b. This parameter is set to zero because it is redundant.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.072636	.007859	9.242	.000	.058756	.089795
Intercept [subject = Person] Variance	.016292	.007870	2.070	.038	.006321	.041991
Intercept [subject = Sentence] Variance	.014947	.010666	1.401	.161	.003691	.060528

5.6 Sentence test comprehension model

The comprehension analysis for the sentence test data (Section 5.3.7) used a multilevel ordinal logistic model, with the comprehension score as the dependent variable (coded as 0 = incorrect/no answer; 0.2, 0.4, 0.6, 0.8 = partially correct on sentences with multi-part answers; and 1 = correct).

Spacing was the only independent variable included in the model. The random effect for reader did not contribute to the model and was removed.

Here is the SPSS syntax and output for the model:

Descriptives

Spacing = s

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	74	0	1	.90	.267
Valid N (listwise)	74				

Spacing = w

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	73	0	1	.93	.217
Valid N (listwise)	73				

```

GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=sentence
  /FIELDS TARGET=Comp TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
  /FIXED EFFECTS=Spacing USE_INTERCEPT=TRUE
    
```

```

/RANDOM USE_INTERCEPT=TRUE SUBJECTS=sentence
COVARIANCE_TYPE=VARIANCE_COMPONENTS
/BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING
INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
/EMMEANS OPTIONS SCALE=ORIGINAL PADJUST=LSD.

```

Generalized Linear Mixed Models

Fixed Coefficients

Target: Comp
Reference Category: 1

Model/Term	Coefficient	Std. Error	t	Sig.
0	-3.765	0.743	-5.069	.000
0.2	-3.623	0.729	-4.968	.000
Threshold for Comp=				
0.4	-3.260	0.701	-4.652	.000
0.6	-2.773	0.673	-4.121	.000
0.8	-2.301	0.653	-3.523	.001
Spacing=s	-0.252	0.502	-0.501	.617
Spacing=w	0 ^a			

Probability distribution: Multinomial
Link function: Cumulative logit

^a This coefficient is set to zero because it is redundant

Fixed Coefficients

Target: Comp
Reference Category: 1

Model/Term	95% Confidence Interval	
	Lower	Upper
0	-5.233	-2.297
0.2	-5.064	-2.181
Threshold for Comp=		
0.4	-4.646	-1.875
0.6	-4.103	-1.443
0.8	-3.593	-1.010
Spacing=s	-1.245	0.741
Spacing=w		

Probability distribution: Multinomial
Link function: Cumulative logit

^a This coefficient is set to zero because it is redundant

6. Hmong word test

6.1 US word list, disyllabic words

The word list test analysis for US readers (Section 6.8.1.1) used a multilevel model, with the Box-Cox transformed reading time in seconds as the dependent variable.

The independent variables kept in the model were:

- Spacing (not significant)
- Log of word frequency in the SCH corpus
- Years of school in Hmong
- English reading level
- Hmong Daw reading level
- Age

Here is the SPSS syntax and output for the model:

```
MIXED BoxTime BY Spacing WITH zln tot ZYr sschoolHm zEngrdlevel
zHDrdlevel zAge
/METHOD=REML
/Fixed=zln tot ZYr sschoolHm zEngrdlevel zHDrdlevel zAge Spacing
/PRINT=SOLUTION TESTCOV CORB
/RANDOM INTERCEPT | SUBJECT(Word) COVTYPE(VC)
/RANDOM INTERCEPT | SUBJECT(Person) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	276.067
Akaike's Information Criterion (AIC)	282.067
Hurvich and Tsai's Criterion (AICC)	282.093
Bozdogan's Criterion (CAIC)	299.566
Schwarz's Bayesian Criterion (BIC)	296.566

The information criteria are displayed in smaller-is-better form.

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	30.930	436.815	.000
zln tot	1	51.363	76.558	.000
ZYr sschoolHm	1	15.097	50.376	.000

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
zEngrdlevel	1	14.998	9.368	.008
zHDrdlevel	1	15.101	11.754	.004
zAge	1	15.054	9.030	.009
Spacing	1	865.888	1.795	.181

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.557711	.027528	38.052	20.259	.000	.501986	.613437
zIntot	-.151302	.017292	51.363	-8.750	.000	-.186011	-.116592
ZYrsschoolHm	-.158386	.022315	15.097	-7.098	.000	-.205924	-.110848
zEngrdlevel	-.083628	.027323	14.998	-3.061	.008	-.141865	-.025391
zHDrdlevel	-.085615	.024972	15.101	-3.428	.004	-.138810	-.032419
zAge	.091237	.030362	15.054	3.005	.009	.026542	.155931
[Spacing=s]	-.022839	.017049	865.888	-1.340	.181	-.056300	.010622
[Spacing=w]	0 ^b	0

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	zIntot	ZYrsschoolHm	zEngrdlevel	zHDrdlevel	zAge
Intercept	1	.019	-.002	-.010	.018	-.005
zIntot	.019	1	-.005	-.002	.007	-.002
ZYrsschoolHm	-.002	-.005	1	.268	-.134	.299
zEngrdlevel	-.010	-.002	.268	1	-.010	.562
zHDrdlevel	.018	.007	-.134	-.010	1	-.454
zAge	-.005	-.002	.299	.562	-.454	1
[Spacing=s]	-.314	-.002	-.007	-.006	.005	-.004
[Spacing=w]	.	^b	^b	^b	^b	^b

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=s]	[Spacing=w]
Intercept	-.314	^b
zIntot	-.002	^b
ZYrsschoolHm	-.007	^b
zEngrdlevel	-.006	^b
zHDrdlevel	.005	^b
zAge	-.004	^b
[Spacing=s]	1	^b
[Spacing=w]	^b	^b

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval
					Lower Bound
Residual	.067399	.003245	20.771	.000	.061330
Intercept [subject = Word] Variance	.012333	.003225	3.824	.000	.007387
Intercept [subject = Person] Variance	.007439	.003230	2.303	.021	.003176

Estimates of Covariance Parameters

Parameter	95% Confidence Interval
	Upper Bound
Residual	.074068
Intercept [subject = Word] Variance	.020590
Intercept [subject = Person] Variance	.017421

Correlation Matrix for Estimates of Covariance Parameters

Parameter		Intercept [subject = Word]	Intercept [subject = Person]
		Variance	Variance
Residual	Residual	-.057	-.022
Intercept [subject = Word] Variance		1	.009
Intercept [subject = Person] Variance		-.022	1

6.2 Elaborate expressions

The analysis of elaborate expressions in the US word list test (Section 6.8.1.2) used a multilevel model, with the Box-Cox transformed reading time in seconds as the dependent variable.

The independent variables kept in the model were:

- Spacing
- Log of word frequency in the SCH corpus
- Age
- Group
- Spacing*Group

Here is the SPSS syntax and output for the model with the *nkhaus niv nkhaus nom* data included:

```

MIXED BoxTime BY Spacing WITH zIntot zAge Group
/METHOD=REML
/FIXED=Spacing zIntot zAge Group Spacing*Group
/PRINT=SOLUTION TESTCOV CORB
/RANDOM INTERCEPT | SUBJECT(Person) COVTYPE(VC).

```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	-82.988
Akaike's Information Criterion (AIC)	-78.988
Hurvich and Tsai's Criterion (AICC)	-78.893
Bozdogan's Criterion (CAIC)	-71.269
Schwarz's Bayesian Criterion (BIC)	-73.269

The information criteria are displayed in smaller-is-better form.

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.389	1621.004	.000
Spacing	1	116.970	6.230	.014
zIntot	1	116.497	39.970	.000
zAge	1	18.426	11.261	.003
Group	1	18.497	1.602	.221
Spacing * Group	1	116.665	5.059	.026

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.839127	.024563	38.781	34.163	.000	.789435	.888818
[Spacing=s]	-.068565	.027471	116.970	-2.496	.014	-.122969	-.014161
[Spacing=w]	0 ^b	0
zIntot	-.088933	.014067	116.497	-6.322	.000	-.116793	-.061073
zAge	.065305	.019461	18.426	3.356	.003	.024487	.106124
Group	.115749	.050063	42.114	2.312	.026	.014726	.216772
[Spacing=s] * Group	-.130424	.057988	116.665	-2.249	.026	-.245269	-.015579
[Spacing=w] * Group	0 ^b	0

b. This parameter is set to zero because it is redundant.

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing =s]	[Spacing =w]	zIntot	zAge	Group	[Spacing=s] * Group	[Spacing=w] * Group
Intercept	1	-.582	. ^b	-.098	.066	.191	-.175	. ^b
[Spacing=s]	-.582	1	. ^b	.068	-.036	-.164	.041	. ^b
[Spacing=w]	. ^b	. ^b	. ^b	. ^b	. ^b	. ^b	. ^b	. ^b
zIntot	-.098	.068	. ^b	1	.032	-.216	.328	. ^b
zAge	.066	-.036	. ^b	.032	1	.058	-.003	. ^b
Group	.191	-.164	. ^b	-.216	.058	1	-.604	. ^b
[Spacing=s] * Group	-.175	.041	. ^b	.328	-.003	-.604	1	. ^b
[Spacing=w] * Group	. ^b	. ^b	. ^b	. ^b	. ^b	. ^b	. ^b	1

b. The correlation is system missing because it is associated with a redundant parameter.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.023130	.003077	7.517	.000	.017821	.030019
Intercept [subject = Person] Variance	.004139	.002600	1.592	.111	.001208	.014180

Correlation Matrix for Estimates of Covariance Parameters

Parameter	Residual	Intercept [subject = Person] Variance
		Variance
Residual	1	-.159
Intercept [subject = Person] Variance	-.159	1

Here is the model without *nkhaus niv nkhaus nom*:

```
MIXED BoxTime BY Spacing WITH zIntot zHDrdlevel zAge
/METHOD=REML
/FIXED=Spacing zIntot zHDrdlevel zAge
/PRINT=SOLUTION TESTCOV
/RANDOM zIntot | SUBJECT(Person) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	-96.398
Akaike's Information Criterion (AIC)	-92.398
Hurvich and Tsai's Criterion (AICC)	-92.293
Bozdogan's Criterion (CAIC)	-84.856
Schwarz's Bayesian Criterion (BIC)	-86.856

The information criteria are displayed in smaller-is-better form.

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	101.719	3658.445	.000
Spacing	1	104.734	2.674	.105
zIntot	1	15.630	9.945	.006
zHDrdlevel	1	106.211	7.183	.009
zAge	1	98.728	34.897	.000

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.793985	.018918	117.903	41.970	.000	.756522	.831448
[Spacing=s]	-.045201	.027643	104.734	-1.635	.105	-.100013	.009611
[Spacing=w]	0 ^b	0
zIntot	-.052143	.016535	15.630	-3.154	.006	-.087264	-.017023
zHDrdlevel	-.038738	.014454	106.211	-2.680	.009	-.067394	-.010082
zAge	.084505	.014305	98.728	5.907	.000	.056120	.112890

b. This parameter is set to zero because it is redundant.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.019758	.002839	6.959	.000	.014908	.026185
zIntot [subject = Person] Variance	.002155	.001997	1.079	.281	.000350	.013250

6.3 Thailand word list

The analysis of the word list test for readers in Thailand (Section 6.8.2) used a multilevel model, with the Box-Cox transformed reading time in seconds as the dependent variable.

The independent variables kept in the model were:

- Spacing
- Log of word frequency in the SCH corpus
- Number of initial consonant letters in the second syllable
- The presence of a final tone letter on the first syllable
- Noun-noun compounds
- Spacing * Noun-noun compounds
- Spacing * Second syllable initial consonants
- Spacing * First syllable tone letter

Here is the SPSS syntax and output for the model:

```
MIXED TBoxTime BY Spacing WITH z1ntot z2ndsylnitcons ztone1stsyl
ZNNforms
/METHOD=REML
/Fixed=Spacing z1ntot z2ndsylnitcons ZNNforms ztone1stsyl
Spacing*ZNNforms Spacing*z2ndsylnitcons Spacing*ztone1stsyl
/PRINT=SOLUTION TESTCOV corb
/RANDOM INTERCEPT | SUBJECT(Word) COVTYPE(VC)
/RANDOM INTERCEPT | SUBJECT(Person) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	-195.555
Akaike's Information Criterion (AIC)	-189.555
Hurvich and Tsai's Criterion (AICC)	-189.537
Bozdogan's Criterion (CAIC)	-170.865
Schwarz's Bayesian Criterion (BIC)	-173.865

The information criteria are displayed in smaller-is-better form.

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	30.584	470.576	.000
Spacing	1	1300.925	5.021	.025

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
zIntot	1	55.974	61.869	.000
z2ndsylinitcons	1	54.882	3.313	.074
ZNNforms	1	54.949	.252	.618
ztone1stsyl	1	56.024	.324	.571
Spacing * ZNNforms	1	1301.558	5.941	.015
Spacing * z2ndsylinitcons	1	1302.260	8.556	.004
Spacing * ztone1stsyl	1	1300.531	6.024	.014

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.833524	.038262	31.933	21.785	.000	.755580	.911467
[Spacing=s]	-.024856	.011093	1300.925	-2.241	.025	-.046618	-.003094
[Spacing=w]	0 ^b	0
zIntot	-.103422	.013149	55.974	-7.866	.000	-.129762	-.077082
z2ndsylinitcons	.041635	.014888	76.343	2.796	.007	.011984	.071285
ZNNforms	-.006989	.014354	76.674	-.487	.628	-.035573	.021595
ztone1stsyl	-.021471	.014477	80.535	-1.483	.142	-.050278	.007337
[Spacing=s] * ZNNforms	.027230	.011172	1301.558	2.437	.015	.005314	.049147
[Spacing=w] * ZNNforms	0 ^b	0
[Spacing=s] * z2ndsylinitcons	-.033391	.011416	1302.260	-2.925	.004	-.055786	-.010996
[Spacing=w] * z2ndsylinitcons	0 ^b	0
[Spacing=s] * ztone1stsyl	.027897	.011366	1300.531	2.454	.014	.005600	.050195
[Spacing=w] * ztone1stsyl	0 ^b	0

b. This parameter is set to zero because it is redundant.

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zIntot	z2ndsylinitcons
Intercept	1	-.146	.	.016	-.001
[Spacing=s]	-.146	1	.	.001	-.006
[Spacing=w]	.	.	1	.	.
zIntot	.016	.001	.	1	.172
z2ndsylinitcons	-.001	-.006	.	.172	1
ZNNforms	.000	-.003	.	.030	.075
ztone1stsyl	.003	.005	.	.063	.206

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zIntot	z2ndsylnitcons
[Spacing=s] * ZNNforms	-.001	.000	. ^b	.002	-.027
[Spacing=w] * ZNNforms	. ^b	. ^b	. ^b	. ^b	. ^b
[Spacing=s] * z2ndsylnitcons	-.002	.003	. ^b	-.007	-.391
[Spacing=w] * z2ndsylnitcons	. ^b	. ^b	. ^b	. ^b	. ^b
[Spacing=s] * ztone1ststyl	.002	-.001	. ^b	.005	-.073
[Spacing=w] * ztone1ststyl	. ^b	. ^b	. ^b	. ^b	. ^b

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	ZNNforms	ztone1ststyl	[Spacing=s] * ZNNforms	[Spacing=w] * ZNNforms
Intercept	.000	.003	-.001	. ^b
[Spacing=s]	-.003	.005	.000	. ^b
[Spacing=w]	. ^b	. ^b	. ^b	. ^b
zIntot	.030	.063	.002	. ^b
z2ndsylnitcons	.075	.206	-.027	. ^b
ZNNforms	1	-.006	-.393	. ^b
ztone1ststyl	-.006	1	.006	. ^b
[Spacing=s] * ZNNforms	-.393	.006	1	. ^b
[Spacing=w] * ZNNforms	. ^b	. ^b	. ^b	. ^b
[Spacing=s] * z2ndsylnitcons	-.028	-.077	.074	. ^b
[Spacing=w] * z2ndsylnitcons	. ^b	. ^b	. ^b	. ^b
[Spacing=s] * ztone1ststyl	.005	-.409	-.025	. ^b
[Spacing=w] * ztone1ststyl	. ^b	. ^b	. ^b	. ^b

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=s] * z2ndsylnitcons	[Spacing=w] * z2ndsylnitcons	[Spacing=s] * ztone1ststyl	[Spacing=w] * ztone1ststyl
Intercept	-.002	. ^b	.002	. ^b
[Spacing=s]	.003	. ^b	-.001	. ^b
[Spacing=w]	. ^b	. ^b	. ^b	. ^b
zIntot	-.007	. ^b	.005	. ^b
z2ndsylnitcons	-.391	. ^b	-.073	. ^b
ZNNforms	-.028	. ^b	.005	. ^b
ztone1ststyl	-.077	. ^b	-.409	. ^b
[Spacing=s] * ZNNforms	.074	. ^b	-.025	. ^b
[Spacing=w] * ZNNforms	. ^b	. ^b	. ^b	. ^b
[Spacing=s] * z2ndsylnitcons	1	. ^b	.205	. ^b

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=s] * z2ndsylnitcons	[Spacing=w] * z2ndsylnitcons	[Spacing=s] * ztone1stsyl	[Spacing=w] * ztone1stsyl
[Spacing=w] * z2ndsylnitcons	b	b	b	b
[Spacing=s] * ztone1stsyl	.205	b	1	b
[Spacing=w] * ztone1stsyl	b	b	b	b

b. The correlation is system missing because it is associated with a redundant parameter.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.042411	.001665	25.473	.000	.039270	.045803
Intercept [subject = Word] Variance	.008675	.002019	4.296	.000	.005497	.013690
Intercept [subject = Person] Variance	.032723	.009485	3.450	.001	.018540	.057755

Correlation Matrix for Estimates of Covariance Parameters

Parameter	Residual	Intercept [subject = Word]	Intercept [subject = Person]
		Variance	Variance
Residual	1	-.041	-.004
Intercept [subject = Word] Variance	-.041	1	.001
Intercept [subject = Person] Variance	-.004	.001	1

6.4 Word list, combined US-Thailand data

The analysis of the combined US-Thailand word list data (Section 6.8.3.1) used a multilevel model, with the Box-Cox transformed reading time in seconds as the dependent variable.

The independent variables kept in the model were:

- Spacing
- Country of testing
- Log of word frequency in the SCH corpus
- The presence of a final tone letter on the first syllable
- Years of school in Hmong

- Number of phonological words
- Spacing * Number of phonological words
- Spacing * First syllable tone letter

Here is the SPSS syntax and output for the model:

```
MIXED BoxTime BY Spacing Country WITH zln tot ZYr schoolHm ztone l st syl
zphon words
/criteria=mxiter(200) MXSTEP(200)
/METHOD=REML
/FIXED=Spacing zln tot Country ztone l st syl ZYr schoolHm zphon words
Spacing*ztone l st syl Spacing*zphon words
/PRINT=SOLUTION TESTCOV corb
/RANDOM INTERCEPT | SUBJECT(Word) COVTYPE(VC)
/RANDOM INTERCEPT | SUBJECT(Person) COVTYPE(VC).
```

Mixed Model Analysis

Information Criteria

-2 Restricted Log Likelihood	401.154
Akaike's Information Criterion (AIC)	407.154
Hurvich and Tsai's Criterion (AICC)	407.164
Bozdogan's Criterion (CAIC)	427.402
Schwarz's Bayesian Criterion (BIC)	424.402

The information criteria are displayed in smaller-is-better form.

Fixed Effects

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	61.213	439.309	.000
Spacing	1	2217.834	6.004	.014
zln tot	1	58.132	82.474	.000
Country	1	43.231	31.646	.000
ztone 1st syl	1	56.382	.057	.812
ZYr schoolHm	1	43.078	5.152	.028
zphon words	1	56.831	.088	.767
Spacing * ztone 1st syl	1	2219.436	8.744	.003
Spacing * zphon words	1	2217.729	8.356	.004

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.559800	.049687	52.810	11.267	.000	.460133	.659468
[Spacing=s]	-.024939	.010178	2217.834	-2.450	.014	-.044899	-.004980
[Spacing=w]	0 ^b	0
zIntot	-.140097	.015427	58.132	-9.081	.000	-.170975	-.109218
[Country=T]	.353123	.062772	43.231	5.625	.000	.226550	.479695
[Country=US]	0 ^b	0
ztone1stsyl	-.018693	.016166	70.449	-1.156	.251	-.050933	.013546
ZYrsschoolHm	-.069478	.030609	43.078	-2.270	.028	-.131204	-.007752
zphonwords	-.010005	.016706	69.030	-.599	.551	-.043332	.023323
[Spacing=s] * ztone1stsyl	.030093	.010177	2219.436	2.957	.003	.010136	.050051
[Spacing=w] * ztone1stsyl	0 ^b	0
[Spacing=s] * zphonwords	.029467	.010194	2217.729	2.891	.004	.009477	.049457
[Spacing=w] * zphonwords	0 ^b	0

b. This parameter is set to zero because it is redundant.

Correlation Matrix for Estimates of Fixed Effects

Parameter	Intercept	[Spacing=s]	[Spacing=w]	zIntot	[Country=T]
Intercept	1	-.104	.	.025	-.716
[Spacing=s]	-.104	1	.	.002	.001
[Spacing=w]	.	.	1	.	.
zIntot	.025	.002	.	1	-.004
[Country=T]	-.716	.001	.	-.004	1
[Country=US]
ztone1stsyl	.003	.002	.	.033	.000
ZYrsschoolHm	-.061	-.002	.	.000	.077
zphonwords	-.004	-.005	.	-.224	.000
[Spacing=s] * ztone1stsyl	-.001	.002	.	.003	.001
[Spacing=w] * ztone1stsyl
[Spacing=s] * zphonwords	-.002	-.005	.	-.001	.001
[Spacing=w] * zphonwords

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Country=US]	ztone1stsyl	ZYrsschoolHm	zphonwords
Intercept	.	.003	-.061	-.004
[Spacing=s]	.	.002	-.002	-.005
[Spacing=w]

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Country=US]	ztone1stsyl	ZYrsschoolHm	zphonwords
zIntot	b	.033	.000	-.224
[Country=T]	b	.000	.077	.000
[Country=US]	b	b	b	b
ztone1stsyl	b	1	.001	.021
ZYrsschoolHm	b	.001	1	.000
zphonwords	b	.021	.000	1
[Spacing=s] * ztone1stsyl	b	-.326	-.002	-.008
[Spacing=w] * ztone1stsyl	b	b	b	b
[Spacing=s] * zphonwords	b	-.007	-.002	-.305
[Spacing=w] * zphonwords	b	b	b	b

Correlation Matrix for Estimates of Fixed Effects (continued)

Parameter	[Spacing=s] *	[Spacing=w] *	[Spacing=s] *	[Spacing=w] *
	ztone1stsyl	ztone1stsyl	zphonwords	zphonwords
Intercept	-.001	b	-.002	b
[Spacing=s]	.002	b	-.005	b
[Spacing=w]	b	b	b	b
zIntot	.003	b	-.001	b
[Country=T]	.001	b	.001	b
[Country=US]	b	b	b	b
ztone1stsyl	-.326	b	-.007	b
ZYrsschoolHm	-.002	b	-.002	b
zphonwords	-.008	b	-.305	b
[Spacing=s] * ztone1stsyl	1	b	.024	b
[Spacing=w] * ztone1stsyl	b	b	b	b
[Spacing=s] * zphonwords	.024	b	1	b
[Spacing=w] * zphonwords	b	b	b	b

b. The correlation is system missing because it is associated with a redundant parameter.

Covariance Parameters

Estimates of Covariance Parameters

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Residual	.059902	.001800	33.283	.000	.056476	.063535
Intercept [subject = Word] Variance	.013293	.002828	4.700	.000	.008760	.020170
Intercept [subject = Person] Variance	.042964	.009524	4.511	.000	.027825	.066342

Correlation Matrix for Estimates of Covariance Parameters

Parameter	Residual	Intercept [subject = Word]	Intercept [subject = Person]
		Variance	Variance
Residual	1	-.025	-.004
Intercept [subject = Word] Variance	-.025	1	.002
Intercept [subject = Person] Variance	-.004	.002	1

6.5 Word list comprehension model

The comprehension analysis for the word list data (Section 6.8.4) used a multilevel ordinal logistic model, with the comprehension score (coded as 0 = incorrect/no answer, 0.5 = partially correct, and 1 = correct) as the dependent variable. Spacing was the only independent variable included in the model.

Here is the SPSS syntax and output for the model for the US data only:

Descriptives

Spacing = s

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	528	.0	1.0	.940	.2361
Valid N (listwise)	528				

Spacing = w

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	528	.0	1.0	.921	.2685
Valid N (listwise)	528				

```

GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=person*word
  /FIELDS TARGET=Comp TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
  /FIXED EFFECTS=Spacing USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=word
  COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=person
  COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING
  INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
  CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
    
```

Generalized Linear Mixed Models

Fixed Coefficients

Target: Comp
Reference Category: 1.0

Model/Term	Coefficient	Std. Error	t	Sig.	
Threshold for Comp=	0	-3.003	0.330	-9.087	.000
	0.5	-2.986	0.330	-8.997	.000
Spacing=s	0.312	0.263	1.186	.236	
Spacing=w	0 ^a				

Probability distribution: Multinomial
Link function: Cumulative logit

^aThis coefficient is set to zero because it is redundant.

Fixed Coefficients

Target: Comp
Reference Category: 1.0

Model/Term	95% Confidence Interval		
	Lower	Upper	
Threshold for Comp=	0	-3.652	-2.355
	0.5	-3.613	-2.319
Spacing=s	-0.205	0.828	
Spacing=w			

Probability distribution: Multinomial
Link function: Cumulative logit

^aThis coefficient is set to zero because it is redundant.

Here is the SPSS syntax and output for the model for the Thailand data only:

Descriptives

Spacing = s

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	791	.0	1.0	.908	.2882
Valid N (listwise)	791				

Spacing = w

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	794	.0	1.0	.902	.2954
Valid N (listwise)	794				

```

GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=person*word
  /FIELDS TARGET=Comp TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
  /FIXED EFFECTS=Spacing USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=word
  COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=person
  COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING
  INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
  CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
  
```

Generalized Linear Mixed Models

Fixed Coefficients

Target: Comp
Reference Category: 1.0

Model Term	Coefficient	Std. Error	t	Sig.
0	-2.837	0.305	-9.311	.000
Threshold for Comp=				
0.5	-2.800	0.304	-9.201	.000
Spacing=s	0.117	0.187	0.629	.529
Spacing=w	0 ^a			

Probability distribution: Multinomial
Link function: Cumulative logit

^a This coefficient is set to zero because it is redundant.

Fixed Coefficients

Target: Comp
Reference Category: 1.0

Model Term	95% Confidence Interval	
	Lower	Upper
Threshold for Comp=		
0	-3.434	-2.239
0.5	-3.396	-2.203
Spacing=s	-0.249	0.484
Spacing=w		

Probability distribution: Multinomial
Link function: Cumulative logit

*This coefficient is set to zero because it is redundant

Here is the SPSS syntax and output for the model for the combined data set:

Descriptives

Spacing = s

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	1400	0	1	.92	.271
Valid N (listwise)	1400				

Spacing = w

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Comp	1402	0	1	.91	.289
Valid N (listwise)	1402				

```

GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=person*word
  /FIELDS TARGET=Comp TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
  /FIXED EFFECTS=Spacing USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=word
  COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=person
    
```



```

COVARIANCE_TYPE=VARIANCE_COMPONENTS
/BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING
INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
/EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.

```

Generalized Linear Mixed Models

Fixed Coefficients

Target: Comp
Reference Category: 1

Model Term	Coefficient	Std. Error	t	Sig.	
Threshold for Comp=	0	-2.874	0.248	-11.607	.000
	0.5	-2.833	0.247	-11.459	.000
Spacing=s		0.199	0.148	1.349	.178
Spacing=w		0 ^a			

Probability distribution: Multinomial
Link function: Cumulative logit

^a This coefficient is set to zero because it is redundant.

Fixed Coefficients

Target: Comp
Reference Category: 1

Model Term	95% Confidence Interval		
	Lower	Upper	
Threshold for Comp=	0	-3.359	-2.388
	0.5	-3.318	-2.348
Spacing=s		-0.090	0.499
Spacing=w			

Probability distribution: Multinomial
Link function: Cumulative logit

^a This coefficient is set to zero because it is redundant.

RESUME

Name: Seth Vitrano-Wilson

Date of Birth: 13 January 1980

Place of Birth: USA

Institutions Attended:

1998-1999, Pomona College, USA

1999-2002, Bachelor of Arts in International Affairs, University of Colorado at Boulder, USA

2008-2009, Graduate Institute of Applied Linguistics, USA

2011-Present, Master of Arts in Linguistics, Payap University, Thailand

PAYAP UNIVERSITY