

Chapter 7

Conclusion

Before discussing the results of this thesis as a whole, and what they might mean for groups making orthography decisions about spacing, I will compare the results from the tests of Hmong readers with results from the pilot tests, and with the analysis of the spacing practices of Hmong writers.

7.1 Comparing Akha and Hmong Daw results

Since the Lahu Si data was relatively sparse and inconclusive, I will focus my comparison on Akha and Hmong Daw.

7.1.1 Overall effect of spacing style

The Akha readers were clearly faster at reading the syllable-spaced stories than the word-spaced stories. In contrast, neither the stories test with Hmong readers in the US nor the sentence test with readers in Thailand showed a statistically significant difference between spacing styles overall, although the sentence test did show a significant advantage to syllable spacing for sentences with words readers had not yet seen in the test.

As discussed in Section 3.4.6, the Akha results are intriguing, but difficult to interpret. Particularly challenging to determine is the effect of the complex orthography situation in Akha, with one long-used tradition having syllable spacing and final tone punctuation marks, and a new, less familiar orthography using final tone letters and word spacing. Also complicating the analysis is the fact that the word-spaced text for Akha had no hyphens or apostrophes, whereas the word-spaced Common Akha Orthography uses both hyphens and apostrophes to mark certain syllable and morpheme boundaries.

Normally, we would expect a predominantly disyllabic language like Akha to show a greater preference for word spacing than a mainly monosyllabic language like Hmong Daw, since readers must mentally join more syllables together to understand the meaning of the text. Is the fact that we see the opposite effect relevant, or is it

simply a product of the different orthography situations for the two languages? Or, is the fact that Hmong Daw shows less of a benefit for syllable spacing mainly *because* of its monosyllabicity, so that there are fewer words in which spacing style even matters?

As it turns out, both the Akha test stories and the Hmong Daw stories have about 400 syllables each in total. Yet there are only 38 polysyllabic Hmong Daw words, but 141 polysyllabic Akha words. Moreover, the Akha test stories contain 52 three-syllable words and 12 four-syllable words. As a result, there are more than five times more syllables that are part of polysyllabic words in Akha than Hmong Daw. If Akha readers are mainly accustomed to processing text by syllable, then we would expect Akha readers to be more disrupted by the lack of intersyllable spacing than Hmong readers would be. Clearly, there are many complex factors that influence this comparison. But it seems at least plausible that the larger benefit to syllable spacing for Akha is partly *because* of, rather than in spite of, its relative polysyllabicity compared to Hmong Daw, if we assume that Akha readers are still mainly processing text syllable by syllable. The fact that readers who almost always read in the old, syllable-spaced orthography rather than the new word-spaced orthography showed an especially large advantage for syllable spacing corroborates this idea, as does the fact that older Akha readers who have read syllable-spaced text for longer also showed a greater advantage for syllable spacing.

Finally, although there was some hint in the Akha data that less proficient readers benefited more from syllable spacing in percentage terms than more proficient readers, this effect did not reach statistical significance. In contrast, there was no sign that reading proficiency had any relationship at all to the effect of spacing style in Hmong Daw in the various tests performed.

7.2 Comparing Hmong spacing in reading and in writing

Chapter 4 analyzed the factors that seem to influence Hmong writers as they decide whether to write polysyllabic words as a unit or with spaces between syllables. The word list test in Chapter 6 provided an alternative viewpoint, giving factors that are significant in predicting the relative ease of reading spaced or unspaced words of different types. From Kuperman & Bertram 2013, we would expect a great deal of overlap between these two sets of factors. To what extent do the variables that influence writing and those that influence reading match, and to what extent do they differ?

The following two variables were found to be significant in at least one writing test and one reading test:

- **Elaborate expressions** favor syllable spacing relative to other words⁵ (both the soc.culture.hmong corpus and the United Bible Societies New Testament for writing; only tested in US for reading)
- **Noun-noun forms.** Compounds with two nouns favor word spacing (in the UBS NT for writing; in the Thailand word list data for reading)

The markers of word status that interact significantly with spacing are different for reading and writing, but are still related:

- **Monomorphemic words** favor word spacing for writing (SCH corpus)
- **Fully reduplicated forms** favor syllable spacing for writing (SCH corpus)
- **Number-classifier forms**, which are tone sandhi compounds, favor syllable spacing for writing (SCH corpus)
- **Phonologically separated words** favor word spacing for reading (in the combined word list data)

Although monomorphemic words did not show a statistically significant difference in reading speed from polymorphemic words in the reading analysis, they did tend to favor word spacing when considered along with other phonologically separated word types (affixed words, semantically opaque compounds, coordinate compounds, and lexicalized terms often found written unspaced). They also showed a smaller mean advantage to syllable spacing than any other word type except affixed words. The fact that other morphologically or semantically connected words did not show an effect in writing may be because writers are more aware of when they create meaningless or misleading syllables than when they separate compound words or bound morphemes. The general direction for both reading and writing, though, favors word spacing for monomorphemic words more than the average word.

Similarly, the results that fully reduplicated forms and number-classifier forms favor syllable spacing in writing, while phonologically unified words favor syllable spacing in reading, overlap to a large degree. Both reduplicated words and the tone sandhi number-classifier compounds in this test are phonologically unified. However, number-classifier constructions do not show up as especially favoring syllable

⁵ That is, they favor syllable spacing more than the average word on the word list. The same meaning applies to all variables in this section—words that “favor syllable spacing” or “favor word spacing” do so only in comparison to the average word. For both the reading and the writing analyses, most words favored syllable spacing more than word spacing overall.

spacing in reading more than other phonologically joined words, the way they do in writing. It may be because we simply do not have enough data from the word list test—we only have three number-classifier words in the word list, and only two of those were tested in both the US and Thailand. We do already have a variable in the model for phonologically unified words, which covers all number-classifier constructions because they are all examples of tone sandhi. Perhaps the effect on reading existed, but was not strong enough to contribute independently to the model when compared to other tone sandhi words or reduplicated words. Indeed, if we choose to separate number-classifier words from other tone sandhi compounds and analyze the model results for both groups, we find that number-classifier forms do show a slightly higher mean advantage to syllable spacing than other tone sandhi words, with a mean advantage of 13% (95% confidence interval from -0.2% to 23%), compared with 9% (95% confidence interval from 0.3% to 17%) for other tone sandhi words. This difference is not large nor statistically significant, but it does suggest the possibility that a closer look would reveal the same effect for reading as for writing.

Both the reading and writing analyses show effects of the internal syllables in polysyllabic words. However, their motivation seems quite different. For reading, both significant variables relate to the clarity of the syllable boundary orthographically:

- **Tone letter on first syllable.** Lacking a tone letter on the first syllable favors syllable spacing (significant in both the Thailand word list data and the combined word list data for reading).
- **Number of letters in second syllable onset.** Having more initial consonants in the second syllable favors syllable spacing for reading (significant in the Thailand word list data).

As described in Section 6.9.2, these variables most likely relate to the way that final tone letters on the first syllable are helpful for marking syllable boundaries. Having many initial consonants in the second syllable makes it harder to see these final tone letters, thus obscuring the syllable boundary and making word-spaced text more difficult to read. A longer second syllable onset also makes words longer, favoring syllable spacing according to Bertram and Hyönä (2003)'s “visual acuity principle”.

For writing, however, the effects relate primarily to the length and frequency of the first syllable:

- **Ratio of first syllable frequency in target word over total frequency.** Words in which the first syllable nearly always occurs in that particular word, and not by itself or in another word, are more likely to be written unspaced (in both the SCH corpus and the UBS NT).
- **Number of letters in the first syllable.** Words with longer first syllables favor syllable spacing for writing (SCH corpus).

The differences between the factors affecting spacing style in reading and writing suggest that readers are primarily concerned with finding syllable boundaries, whereas writers are primarily thinking about the meaning and length of the first constituent. As for why first syllable effects are stronger for writing than second syllable effects, both here and in Kuperman & Bertram 2013, it may simply be due to the first syllable having mental primacy as writers make decisions about a word, since they come to it first. This would also help explain why words with syllables that appear most often within that word rather than in other words favor word spacing for writing. If a writer sees the first syllable and strongly associates it with the whole word, she would be more likely to conceive of the two syllables as a single unit and write them as such.

Finally, one factor was found to significantly influence spacing in writing, but not in reading, and does not seem to relate to other factors:

- **Number of Hmong Daw/Hmong Njua differences.** Words with many spelling differences between Hmong Daw and Hmong Njua are more likely to be written unspaced (in both the SCH corpus and the UBS NT), but show no difference for reading.

It is not clear why greater dialectal differences for a word should affect writing choices but not reading. Perhaps dialectal differences are highly salient to writers for sociolinguistic reasons, but spacing does not actually make a large difference in the ease of reading these words.

7.3 Comparing syllable spacing and word spacing overall

The main focus of testing in this thesis has been to compare syllable-spaced text to word-spaced text in Hmong Daw. What have we learned that might be helpful to people making decisions about their writing system?

7.3.1 Theoretical advantages

Before we discuss the results of the tests performed, let us consider some of the relevant literature reviewed in relation to syllable and word spacing.

7.3.1.1 Advantages of syllable spacing

The psycholinguistic grain size theory developed by Ziegler and Goswami (2005) argues that readers begin with awareness of larger phonological units and then proceed to smaller units as they learn to read an alphabetic script. In this view, words are the most easily accessible phonological level, followed by syllables, and then phonemes. By this argument, it would seem that since words are more easily accessible to nonreaders or beginning readers, word spacing would be easier for such readers than syllable spacing.

However, word-spaced text by necessity involves longer units than syllable-spaced text. Ziegler and Goswami (2005) call this the “granularity” of a writing system—a system dividing text by syllables but not words will have far fewer units needed to represent the whole language, whereas word-based systems will require many more. If phonemes are not yet accessible to readers, they may find the shorter, more limited number of possible syllables easier to master than the large variety of written words, even if they may initially be more aware of the word level than the syllable level. Of course, systems that efficiently mark both syllable and word boundaries would be ideal. But if only one level is chosen, syllable-based systems may be easiest for beginning readers to master, as they represent a compromise between too-numerous words and too-finely-grained phonemes. As readers learn to master the phonemic principle, they can more easily do away with the syllable level, using their phonemic awareness to put letters together into words that more closely map the meaningful units needed to interpret text.

There seems to be some disagreement about what level of awareness pre-readers and early readers have of the word and the syllable. While Ziegler and Goswami say that word awareness comes first, others such as Ferreiro (1999) and Chaney (1989) say that syllable awareness comes earlier. Part of the disagreement may relate to exactly what is meant by “word awareness.” A preschooler is undoubtedly more aware of the unit “water” than the syllables “wa” + “ter”. But an exhaustive understanding of word boundaries requires awareness of compound words, and of articles, prepositions, and other functor words that are often joined to content words in early writing.

Word awareness, then, may begin before syllable awareness does, but it is completed later, if it is ever truly complete. Even linguists disagree at times on precisely what a “word” is, whereas there is much more consensus overall on syllable boundaries. The clarity of syllable boundaries will vary from language to language, but it seems clear that in general, writers can exhaustively determine syllable boundaries more easily than word boundaries. After all, the very topic of this thesis was spurred by the challenges many groups face in deciding where to put word breaks, and entire chapters of orthography-related works have been devoted to helping groups decide where to put word breaks (Kutsch Lojenga 2014). Syllable boundaries do not seem to have created nearly such a problem for the groups that have chosen to represent them (apart perhaps from the issue of sesquisyllabic words, which will be addressed below). Moreover, as Ferreiro (1999) and Bassetti (2009) argue, it can take upwards of two years of schooling before young readers learn how to apply the basic conventions of interword spacing of orthographies that use them. While we have no precise data on how long it would take readers to accurately mark syllable boundaries, it seems clear that syllable spacing would be more quickly mastered in writing than word spacing.

Syllable spacing also has one other advantage in a very particular context, which we saw in practice for the Hmong data in Chapter 6. With fully reduplicated words, writing as a unit has no theoretical benefit. It obscures the first constituent, the main unit needed for processing the meaning of the reduplicated word, but it does not reduce ambiguity at all. So regardless of what general decision language communities make about syllable versus word spacing, separating fully reduplicated words would be beneficial across languages.

7.3.1.2 Advantages of word spacing

Clearly, word spacing does have advantages in certain areas. What theoretical considerations help us understand when we might expect word spacing to be particularly beneficial?

One of the main advantages of word spacing over syllable spacing is that all segmented units are meaningful. For syllable-spaced text, any morpheme longer than one syllable will be split up, so that the individual syllables will either be meaningless or misleading (an example of the latter would be the English word “carpet”, which of course has nothing to do with cars or pets). Such separation might be easier to learn how to write—although in this case, one would expect that

the incomplete word awareness that Ziegler and Goswami (2005) argue comes before syllable awareness would help writers know that these syllables, at least, go together. But since research indicates that monomorphemic words are more likely to be read as a unit than polymorphemic words (Cui, Drieghe et al. 2013; Ji et al. 2011; Juhasz 2006; Duñabeitia et al. 2008; Muncer et al. 2014), one would expect that inserting a space within morphemes and creating non-meaningful or misleadingly meaningful units would slow readers down as they attempt to process the meaning of a full text.

Another potential area of benefit from word spacing is in cases where ambiguity exists between phrases and compound words, such as the English “green house” versus “greenhouse.” Word spacing separates the main independent units of syntax while joining syntactically connected compounds, thus eliminating the ambiguity in such cases. We saw this for English compounds in Juhasz et al. 2005, in which normally unspaced adjective-noun compounds showed a larger disruption from separation than noun-noun compounds, since the adjective-noun compounds are syntactically ambiguous in English but the noun-noun compounds are not. So we might expect to see a benefit from word spacing in such cases due to the elimination of ambiguity.

7.3.1.3 Reading level-dependent advantage

There are clear theoretical reasons why we would expect beginning readers to benefit from syllable spacing, and advanced readers to benefit from word spacing. Research has found that younger children have smaller perceptual spans as they read (Rayner 1986, Häikiö et al. 2009). Although these studies did not test reading proficiency directly, but rather age, it stands to reason that beginning readers, regardless of age, would have shorter perceptual spans. Since they are less adept at rapidly recognizing letters, beginning readers can store fewer letters in their short-term memory before they must process what they have seen, so they are necessarily processing smaller chunks of text than more proficient readers. For such readers, separating these smaller chunks would help them by matching their perceptual span with the units marked in the orthography. Alternatively, for proficient readers with longer perceptual spans, we would expect word spacing, with its closer match to meaning, to be helpful. Age apart from reading proficiency might play some minor role, but since the eye movements in reading of older elementary students is quite similar to adult eye movements, it seems that reading proficiency and not age is the far more important factor.

We should be careful, however, not to automatically extend these findings to people who are proficient readers in a national language, but not yet proficient readers in their mother tongue. Such readers are already well trained in the concept of sound-symbol connection, and have already developed the deeper phonological awareness that comes with reading. However, they have not yet developed the rapid pattern recognition particular to the orthography of their mother tongue that enables proficient reading. We might therefore expect such readers to have an intermediate perceptual span in their mother tongue, since they would be faster at letter recognition than those who are not yet literate in any language, but slower than proficient mother tongue readers in word and morpheme recognition, and less adept at the use of subtle cues in the text to predict upcoming content.

In general, therefore, we would expect completely new readers to benefit from a greater use of space, readers who are proficient in another language but not yet in their mother tongue to benefit from some intermediate spacing style, and the most proficient mother tongue readers to benefit from larger chunks of text, up to the typical limit of 7 to 9 letters (for English readers, at least) that define proficient readers' saccade lengths. However, this does not automatically translate to "syllable spacing is best for beginning readers, and word spacing is best for proficient readers." Each language and each orthography will interact with spacing style in different ways. The optimal psycholinguistic grain size for a given level of reading proficiency will depend on such factors as the frequency of polysyllabic words and morphemes, the clarity or ambiguity of syllable boundaries in word-spaced text, the frequency and type of polymorphemic words, the frequency of consonant clusters and diphthongs, the frequency of digraphs, trigraphs, and even tetragraphs (of which Hmong Daw has three), etcetera. We can predict from theory that beginning readers would benefit from more spacing than proficient readers, but we cannot say with any certainty what level of spacing would benefit beginning or advanced readers of a particular language reading a particular orthography.

7.3.1.4 Implications of Chinese hanzi studies

As we mentioned in Section 2.1.1, studies on readers of Chinese hanzi, which uses morphosyllabic characters without spaces between words, are inconclusive on whether reading in Chinese is fundamentally word-based or syllable-based. There are, however, clear word-based effects in reading Chinese (Li et al. 2014, Li et al. 2009, Bai et al. 2008, Li & Ma 2012). Since words are a significant unit for Chinese readers, even when unmarked in the orthography, we might expect word spacing to

help Chinese readers, and potentially readers of other languages as well (Rayner & Pollatsek 2013, Li et al. 2014). Indeed, the fact that studies comparing spaced and unspaced text in Chinese hanzi find no detriment to the unfamiliar word-spaced format, whereas they do find a detriment to adding spaces within words or between every syllable (Bai et al. 2008, Shen et al. 2010), suggests that word spacing might benefit Chinese readers if they became accustomed to the now-unfamiliar format.

However, it is not entirely appropriate to generalize from the potential (though not entirely proven) benefits of word spacing for Chinese hanzi to a benefit for orthographies using other scripts. Hanzi marks syllable boundaries with 100% clarity, so adding interword spacing results in the marking of both syllable and word boundaries. For most orthographies using the Latin script or other alphabetic systems, however, using interword spaces typically means syllable boundaries are not clearly marked. If anything, the potential benefit of interword spacing to Chinese hanzi underlines the idea that orthographies that manage to efficiently mark both syllable and word boundaries would serve readers better than orthographies that only mark one clearly. This does not tell us directly about which of the two should be chosen if only one can be marked clearly in a given orthography.

7.3.2 Tests of word and syllable spacing in Latin script

Besides this thesis, the other studies comparing the reading of syllable-spaced text to word-spaced text have used Chinese pinyin.

7.3.2.1 Chinese pinyin

Bassetti (2009) compared adults reading Chinese pinyin presented in both syllable-spaced and word-spaced format. She found no significant difference between the two, despite the fact that word spacing is both the official standard and the most common format for pinyin. Bassetti and Masterson (2012) found that while adults showed no difference between syllable-spaced and word-spaced pinyin, children read syllable-spaced text faster, by an average of 16%.

This supports the idea that beginning readers benefit from greater segmentation than advanced readers. It also suggests that syllable spacing may have more of an advantage than is commonly assumed, if even adult readers of pinyin show no advantage for word spacing, despite its official status and greater prevalence than syllable-spaced pinyin. However, it is also possible that the hanzi system, which is

orthographically syllable-based, may be influencing Chinese readers and causing them to process by syllable more than would be the case for readers in another language.

7.3.2.2 Akha

In Chapter 3 on pilot testing, we saw an average advantage of 16% in favor of syllable spacing for adult and adolescent readers of Akha in Thailand. However, unlike in Chinese, where pinyin has been officially standardized for decades and is officially written with word spacing, the Akha readers in this study have only recently switched from a syllable-spaced orthography to a word-spaced one. Although the test was performed in the new orthography, called the Common Akha Orthography, most readers said they were more proficient in the old system (known as the Baptist orthography). The word-spaced format presented in the test also differed from the new word-spaced orthography in that it lacked any hyphens or apostrophes to segment words internally. It is therefore difficult to say whether the results for Akha, a predominantly disyllabic language with many affixes and compounds, truly show a benefit to syllable spacing that could be applied to other languages, or whether it is mainly a result of particularities in the orthography situation of Akha. We can at least say, however, that the results from Akha are consistent with the idea that there is no inherent universal benefit to word spacing over syllable spacing for reading.

7.3.2.3 Hmong Daw

The stories test with Hmong readers in the US showed no difference between syllable spacing and word spacing. The sentence test given to readers in Thailand, based on the same stories but presented and measured one sentence at a time, showed no significant difference overall. However, sentences with polysyllabic words that were appearing for the first time in the test did show a significant advantage for syllable spacing. An advantage for syllable spacing was also seen in the word list test, wherein readers of Hmong Daw read disyllabic words in isolation. The overall results from Hmong suggest that there is either no difference between word spacing and syllable spacing, or that syllable spacing may be slightly more advantageous, though not enough in natural text to be of any practical importance.

Are issues of unfamiliarity with the word-spaced format used in this test masking an advantage for word spacing? Readers of Hmong Daw are typically exposed to both

syllable-spaced and word-spaced formats. However, precisely which forms are considered “words” to be joined varies from writer to writer, and from community to community. The word-spaced text presented to the Hmong readers in this thesis was based on a “linguistic” definition of a word. But most Hmong materials that have word spacing (including the materials and style that the Protestant Christian readers in my sample would be most familiar with, such as the UBS Bible and other evangelical literature) tend to join fewer words together than my “linguistic” definition would join. Since one of the main objectives of this study is to compare syllable-spaced text to the “word-spaced” text that would result from the work of outside linguists consulting on orthography decisions, the spacing style in the test materials properly reflects that goal. This does mean, though, that the presentation of certain words in a joined format may have been unfamiliar to some readers, even those who said they were used to “word-spaced” text.

Thankfully, we have multiple ways of addressing this problem. First of all, when I included the spacing style people said they read more often as a factor in the model, it did not significantly influence the effect of spacing style for any tests, nor did it change the final results.

Secondly, the analysis of the spacing practices of Hmong writers in Chapter 4 provided us with a way of incorporating the fact that readers may be more used to seeing some words joined than others. Neither the frequency of joining in the 15-million word SCH corpus, nor the frequency of joining in the UBS New Testament, interacted significantly with spacing style in the word list test, and including these variables in the model did not change the advantage for syllable spacing in the word list at all.

We clearly cannot use this data for the stories test, since we only have one time measurement per story, and each story has many different polysyllabic words. However, when we look at the sentence test and analyze only the sentences with one new polysyllabic word, we find that neither spacing frequency in the SCH corpus nor in the UBS New Testament interact with spacing style. Likewise, if we choose to include these variables in the model anyway, there is no change in the advantage of syllable spacing for these sentences. In other words, we see no evidence at all in any of the tests that issues of the relative unfamiliarity of word spacing for certain readers or certain words are masking an advantage for word spacing.

We can be confident, then, that there is either no difference to reading speed overall in Hmong Daw between word and syllable spacing, or there is a slight advantage for syllable spacing. In neither case is it likely that the difference between syllable spacing and word spacing would be large enough to have any practical importance. In other words, the Hmong Daw results overall, as with the Akha results, indicate that there is no meaningful advantage for word spacing over syllable spacing for reading.

These results fit with those of Bassetti and Masterson (2012), who found no difference between word-spaced and syllable-spaced pinyin for adult readers. However, they found that unlike adults, children did show an advantage for syllable spaced pinyin. Other studies have also found that less proficient readers benefit from greater use of spaces and other segmentation devices than more proficient readers (Häikiö et al. 2011, Kasisopa 2011). In contrast, the slower and faster readers who participated in this study did not differ in their reaction to the two spacing formats presented.

One possible explanation is that the readers of Hmong Daw in this study were not different enough in reading proficiency to show a different reaction to spacing style. The research above compared adults or older students to young elementary school students. Perhaps beginning readers prefer more space, but once readers reach a certain intermediate level of proficiency, their reaction to spacing remains constant. Most of the readers in this study were fairly proficient and frequent readers of Hmong Daw, most having college or graduate degrees, and most reporting that they read in Hmong Daw every day. Moreover, some of the less proficient readers were screened out due to low comprehension scores, while other less proficient readers may have screened themselves out by not volunteering for the test. While there certainly were slower readers tested, they may have still been at a high enough level that they would not show a difference in their reaction to spacing style as seen in previous research on children.

7.3.2.4 Spacing question responses

The final piece of evidence that helps us broadly compare word spacing and syllable spacing from these tests comes from the question asked to all readers about which spacing style is easier to read. For the Hmong readers, 40 of 52 respondents said that syllable spacing was easier to read than word spacing, and only nine said word spacing was easier to read. Out of seven Lahu Si respondents, all seven said syllable

spacing was easier to read (although with Lahu Si, syllable spacing is the standard, as it is with Lahu Na as well). Out of 21 Akha readers, 12 said syllable spacing was easier to read, and 9 said word spacing was easier.

It is interesting that among Hmong readers there is a clear sense that syllable spacing is easier to read, yet the Hmong data does not present a clear statistically significant advantage to reading speed for syllable spacing. The opposite is true for Akha readers—nearly equal numbers said syllable spacing and word spacing were easier, but the Akha reading times show a clear advantage for syllable spacing. This raises the possibility that sociolinguistic attitudes may be playing a large role in shaping beliefs about ease of reading, and that readers may be less consciously aware of actual difficulties in reading the different styles.

7.3.3 The effect of spacing style on different kinds of words

The word list test in Hmong Daw and the Hmong spacing practices analysis are the only test results from this thesis that are fine-grained enough to address questions of what types of words are benefited by different spacing styles. Based on the results for Hmong Daw described in Section 6.8, the general factors given in Table 10 below may be influencing how spacing affects reading in Hmong Daw. We assume that variables shown to affect writing in Hmong may also affect reading if there is some relationship to other known variables affecting reading in Hmong Daw. These findings fit well overall with our theoretical expectations about the comparative advantages and disadvantages of syllable spacing and word spacing. The results here may in some cases be applicable to reading in other languages, especially ones that are morphologically or orthographically similar to Hmong Daw.

Table 10 General factors influencing the effect of spacing style in Hmong

Favor syllable spacing more than average	Favor word spacing more than average
longer words	shorter words
words with orthographically unclear syllable boundaries	words with orthographically clear syllable boundaries
phonologically united words (reduplicated words, tone sandhi compounds, words with spreading nasalization)	phonologically separated words (monomorphemic words, affixed words, semantically opaque compounds, and coordinate compounds)

Note that in Table 10 above, variables that “favor syllable/word spacing more than average” relate to a shift from the average word. This does not mean necessarily that word types on the left column are read more easily with syllable spacing and word types on the right more easily with word spacing (this would only be true if the average were a net zero difference between the spacing styles).

7.3.4 Sesquisyllabic structures

One problematic area for syllable boundaries that is common in Mainland Southeast Asia is the existence of sesquisyllabic structures. Sesquisyllabic words and morphemes are ones in which one syllable, usually the first, is reduced phonetically in some fashion, for instance lacking (or having significantly reduced) vowel contrast or tone. In this sense, they are not always considered “full” syllables by native speakers, and may present a problem for marking syllable boundaries. Sometimes the minor syllable is a morpheme, such as a prefix, or perhaps the first constituent of an old compound word that has become phonetically reduced (in which case it might have already lost its morpheme status). Often it forms a single morpheme with the adjacent major syllable. Minor syllables are usually not properly pronounceable on their own, but are phonologically bound to the adjacent major syllable.

Since none of the languages tested for this thesis have sesquisyllabic structures, there are no direct results that deal with the question of whether minor syllables are best written on their own, or joined with their major syllable partners. However, there is some indirect evidence from this thesis that may be helpful. If most minor syllables in a language are not morphemes, then we would expect from the results of the word list test that separating them off would not provide any benefit to reading and may slow readers down. Also, if they cannot be pronounced on their own, this might hinder the phonological processing of early readers as they try to sound words out. On the other hand, longer words tend to favor syllable spacing, so if a language commonly has minor syllables attached to two or more major syllables in a word, words may begin to extend beyond the perceptual span of some readers.

In terms of writing, we would expect that most speakers of a language with sesquisyllabic structures should be easily able to recognize the difference between a major syllable and a minor syllable, especially if minor syllables cannot be pronounced on their own and rarely if ever have independent meaning. So we would not expect unspaced sesquisyllabic structures to be much more difficult for

writers than spaced, and in fact might be easier. The results from this thesis, along with theoretical considerations, generally support the idea that joining sesquisyllables orthographically would make reading and writing easier overall, although this may depend on the language and the orthography.

7.4 Overall conclusions

In Section 2.5.2, we saw the results of experiments comparing traditionally unspaced texts in Chinese, Japanese, and Thai. The overall results suggested that adding interword spaces would have either no detrimental effect or possibly a beneficial effect in such cases, depending on how much readers improved as they became more familiar with word-spaced text.

If interword spaces have the potential to assist reading for orthographies that lack them, they also come at some cost to writing. When writing hanzi, no Chinese writer ever has to wonder about whether a certain compound construction qualifies as a word or not, or whether certain functors are affixes, independent words, or clitics. As we have already discussed, avoiding the question of word breaks makes learning to write easier and saves time in instruction. It also makes standardization of an orthography easier, since it eliminates variation in word breaks from writer to writer.

In contrast to the challenge of learning to write with interword spaces, learning to write with intersyllable spaces is relatively simple, although the magnitude of the difference depends on how clear both word and syllable boundaries are in a given language. The development of syllable awareness usually occurs before literacy, and can in most cases be easily mastered by young or early readers.

Due to the word effects that seem to be present even in writing systems without interword spaces, we expected to see some benefit to interword spacing in the languages we tested. Instead, we found either no difference between word-spaced text and syllable-spaced text, or an advantage for syllable-spaced text. The advantage for syllable spacing in Akha may easily be due to the complex orthography situation, while the advantage for syllable spacing in Hmong Daw may not apply to natural connected text. But clearly, no advantage for word spacing was seen for readers of either language.

What would we see if we ran this same experiment across a larger, more diverse sample of Hmong readers? What would we see if we increased the diversity of the texts? What would we see if we could somehow test a group of readers exactly

equivalent in all relevant covariates, who are all equally familiar with syllable-spaced and word-spaced text for all the polysyllabic words in the text? We can of course never know the answer to the final question. Overall, though, the results in this thesis do not support the idea that word spacing provides a greater benefit to reading than syllable spacing does for Hmong Daw or Akha.

If there is no apparent benefit to word spacing for reading, but there is a clear benefit to writing for syllable spacing, then the results from this test suggest that syllable spacing is easier to read and write overall for Hmong Daw than word spacing.

Whether these results can be extended to other languages is a more difficult question. We can certainly say that since the native speakers of a language should be the ones primarily involved in orthography decision making in any case, the results from this thesis bolster the case against outsiders pressuring groups not to use syllable spacing if the native speakers see clear benefits to it for their language. In terms of the empirical question of whether a given language with a given orthography would show a benefit for word spacing or not (or even a benefit for syllable spacing), we offer the following general guidelines based on what we have seen:

- Languages with a high degree of monosyllabic morphemes and monomorphemic words are especially likely to show no benefit to word spacing.
- Languages in areas such as Mainland Southeast Asia and China, where the dominant writing systems are syllable-oriented, and where pedagogy focuses on the syllable level, may be especially likely to show no benefit to word spacing.
- Languages with clear syllable boundaries but less clear word boundaries are especially likely to show no benefit to word spacing in writing.
- Orthographies with unclear syllable boundaries inherent in the system (apart from explicit marking with spaces) are more likely to show no benefit to word spacing.
- Orthographies that can find clear and efficient ways of marking both syllable and word boundaries are likely to be easier to read than orthographies that mark one or the other.

- Although we did not test this directly, other research suggests that orthographies that use word spacing should consider secondary segmentation devices such as hyphens or apostrophes to mark certain morpheme or syllable boundaries. These would be especially helpful for long words, for morphologically complex words, and for words whose syllable or morpheme boundaries are orthographically unclear. They may, however, come at some cost to the simplicity of writing, depending on how they are used.

7.5 Objections to syllable spacing

Let us now consider some objections to the use of syllable spacing.

Two anonymous reviewers expressed the opinion that while syllable spacing may benefit early readers, it holds more advanced readers back from their full potential and creates a plateau that they cannot rise above without switching to a word-based system. Although the syllable level may be helpful for sounding out text, it is not directly related to meaning, so syllable spacing encourages a phonological approach to reading that hinders actual comprehension. Though not explicitly stated as such, this idea might be connected to the image of “mindless chanting” that some critics see as dominant in Southeast Asian pedagogy, where recitation or pronunciation is viewed as the goal, not comprehension.

What do the results of this thesis say about this idea? For one, although we did see the hint of a greater advantage for syllable spacing among less proficient readers in Akha, we saw no such relationship in any of the Hmong tests. This is despite a large number of highly proficient Hmong readers participating. Many of the readers in the sample read Hmong every day, most have at least a Bachelor’s degree and many have advanced graduate degrees, some have had up to two years of schooling using Hmong as the language of instruction, and some had reading speeds roughly comparable to my own in English. If anything, the self-selection of readers who did not feel their Hmong reading was good enough to be tested by a foreigner from a university means that this sample of readers is biased toward the more advanced, more confident Hmong readers. There certainly were less proficient readers in the sample as well, of course, but comparing these to the more advanced readers showed no statistically significant relationship with spacing style for any of the several possible proxy variables for reading proficiency available in any Hmong test. A post hoc analysis for the word list test did not show any interaction between average reading speeds in the test and the effect of spacing style either. In addition,

many of the readers said they were more used to word-spaced text than syllable-spaced. If there truly were a plateau that Hmong readers reach for syllable spacing that is lower than the word-spaced plateau, we would have expected this group to have at least a few readers who had surpassed that plateau in their ability to read word-spaced text but been stifled as they read syllable-spaced text. Instead, we see no evidence for this. We also saw no relationship between spacing style and comprehension. For Hmong at least, then, it is difficult to reconcile the results in this thesis with the idea that syllable spacing would create a lower maximum reading speed, and that readers could only advance further by switching to a word-spaced format.

Here we can reiterate the concept of the “paraorthographic linkage hypothesis” presented by Gary Feng, which states that “readers constantly adapt and optimize their reading behaviors,” and that therefore, mature readers “are well-adapted to their own writing systems” (2008:414). Chinese and English readers use remarkably different systems, yet overall, they process information at similar rates (Sun & Feng 1999). This does not mean that orthographies do not matter, and that any system is as good as any other. But it does mean that readers who want to learn to read well, and who practice consistently, will overcome any challenges their orthography presents them and utilize the information it gives them to the greatest extent possible.

We should not be surprised to see, then, that there is no evidence in this data for a “plateau” effect for syllable-spaced text in Hmong. Just as Chinese readers can develop high levels of proficiency despite the lack of word boundary information in their orthography, and just as English readers can develop high proficiency despite the phonemic irregularity of their writing system and the lack of consistent morpheme boundary marking, so too can readers of syllable-spaced orthographies learn to read for meaning by joining syllables together into words in their minds when they encounter polysyllabic words.

7.6 The value and the challenge of compromise

In the previous section, I presented the options of word spacing and syllable spacing as if the choice were a dichotomous one between two well-defined possibilities. In reality, words, and to a much lesser extent syllables, can be defined in a number of ways, which may lead to different results for reading and writing (not to mention for acceptability to a language group). Moreover, it is not at all the case that

language groups have only these two options to choose from. Groups may decide to use spaces to mark some kinds of words but not others, or some morpheme or syllable boundaries but not others. And they may also choose to add hyphens, apostrophes, or some other segmentation device to supplement their use of spaces.

Although I did not directly compare natural text using various types of intermediate spacing styles between word and syllable, nor did I test any secondary segmentation devices, the results in this study and the literature we have reviewed can provide some indirect insight into the benefits and drawbacks of intermediate or alternative approaches to text segmentation.

In general, the Hmong word list results along with the Hmong writing results suggest that to fully optimize the use of spaces for reading would involve a complex application of multiple orthogonal factors, both discrete and continuous, at different linguistic levels. It is impossible to “teach” such a system for writers apart from memorization of each case, along with some helpful guidelines that describe the tendencies for spacing (e.g., “longer words tend to be divided more often than shorter words”). Moreover, certain factors such as phonological unity or semantic opacity would change over time as the language changes, so the system would have to either constantly adapt or become less optimal. However, the evidence from Hmong writing in Section 4.4 and from English writing (Kuperman & Bertram 2013) suggests that writers of word-spaced orthographies do indeed attempt to approximate many of these optimal features in their writing to various degrees, whether consciously or unconsciously.

Even if writers as a group are able to approximate certain features of “optimal spacing” to some degree, it is clear that such a system takes much more time to learn to write than a more regular system. The easiest system for writers would either be syllable spacing, if syllable boundaries in a language are clear, or perhaps some intermediate style with a single simple rule, based on some dichotomous factor that is intuitive to native speakers.

For instance, it would be fairly easy overall—arguably even easier than syllable spacing perhaps—for writers to learn a system that forbids meaningless syllables on their own. We could call such a system “morpheme-spaced,” with the exception that morphemes that are phonologically smaller than a full syllable would not be separated off. Indeed, there is some limited evidence from the word list data that such a system would be easier to read overall than syllable spacing. Since monomorphemic words tended to show less benefit to syllable spacing in isolated

words than other word types, they might show an advantage for word spacing in natural connected texts. And since basic word awareness comes before syllable awareness according to Ziegler & Goswami 2005, it might even be easier to learn to write than a purely syllable-spaced format as well. Such a system would eliminate the problem of misleadingly meaningful subunits of a morpheme that could confuse readers. It would also cause borrowed words or names to be joined as a unit, which would help readers to know that they should not look for meaningful units within these borrowed terms.

A similar possibility would be to only join monomorphemic words and affixed words, so that spaces would mark off free morphemes. Since affixed words also tended to show a relatively smaller advantage to syllable spacing than most words in the word list test, free morpheme spacing may prove beneficial to reading for some languages. Depending on the morphology of the language, it may also not be too difficult to learn for writers. Although this style was not well-liked in the Lahu Si pilot test (see Section 3.3), it may find a warmer reception among other language groups.

These are only two examples of intermediate styles between syllable spacing and word spacing, and other possibilities could be explored by language groups. It is important to keep in mind the tradeoff between what is optimal for reading and what is optimal for writing. With that said, any number of possibilities are learnable if they are most acceptable to a language group. One general possibility would be to have a basic default style (e.g., syllable spacing, morpheme spacing, etc.), but to allow for exceptions if there is a consensus in the language community that certain words should be joined or separated for whatever particular reasons. This would produce a system that is closer to optimality for reading, but at a lower cost to writing than a more irregular system purely optimized for reading.

As mentioned above, hyphens, apostrophes, and other secondary segmentation devices should be explored to see if they make reading easier at not too great a cost to the simplicity of the writing rules. Such segmentation can sometimes serve to make syllable or morpheme boundaries clearer in a way that is easily learnable for writers. For instance, one could imagine a language with final tone letters having a rule that words or morphemes with unmarked tones on a non-final syllable should have a hyphen or apostrophe at the syllable boundary, so that all syllable boundaries are clearly marked. The Hmong word list data suggests that final tone letters are helpful to readers in marking syllable boundaries, so such a rule might help making reading easier, although we do not have direct evidence of this.

Another possibly beneficial use of secondary segmentation devices would be to break up especially long words. For instance, a language might choose to use hyphens at the major boundary of three-constituent compound words, as Bertram et al. 2011 found to be helpful in Finnish and Dutch. However, such a rule would only be advisable if it is fairly intuitive to native speakers which boundary is the “major” boundary in such words, and if there are enough such words to justify adding another spelling rule to the system.

One important caveat to all the suggestions above is that it is quite probable that some of the advantages of syllable spacing or other sub-word level spacing styles seen in this thesis are a product of both the typical morphological and phonological profile of languages in Mainland Southeast Asia, as well as the pedagogical focus on the syllable. This in turn is partly due to the preponderance of syllable-based writing systems in the region, whether Chinese hanzi or Brahmi-derived alphasyllabaries. As Page 2014 suggests, the writing systems of the region help to reinforce the syllable as a unit in reading more than is typical for European word-based systems. Although there may be some results from this thesis that are true universally, one should be quite hesitant to extrapolate these results far beyond the region and language types from which they derive.

7.7 Further research

Having broadly explored issues of text segmentation, but having actually tested very few issues in a limited way, this thesis leaves a great deal open to further research. For instance, would the results of this thesis still hold if a greater variety of texts had been studied? Does the type of text interact with the effect of spacing style, so that certain types of texts would be read more easily with syllable spacing, and others with word spacing? The particular Hmong texts chosen for the story and sentence tests were written for children. Would material written for adults, with potentially greater grammatical complexity and lexical diversity, behave differently?

Similarly, would the results of this thesis hold with a more diverse group of readers? If Hmong readers from China, Vietnam, or elsewhere were included, would they show the same patterns as we have seen? Would readers from more diverse religious backgrounds show the same patterns as we see in this thesis?

In relation to how different types of words are affected by spacing, it would be quite valuable to know more about the interaction between factors at different levels, whether syntactic, morphological, phonological, semantic, or orthographic. The

methodology of this thesis was designed primarily with morphological criteria in mind, and was therefore not ideally suited to the exploration of other factors. What else can we learn about the types of words that tend to be read more easily as a whole or separated by syllables?

Since this thesis did not include languages with sesquisyllabic structures, further research on such languages would help to answer many questions about how spacing styles affect reading and writing in these languages. Do readers find it easier or harder to read when minor syllables are joined to their main syllables? Do writers intuitively know the difference between minor syllables and major syllables enough to easily distinguish them in writing?

Regarding hyphenation and other secondary segmentation devices, what can we learn about the types of words and morphemes that would benefit from secondary segmentation and those that would not? How would different intermediate styles that use spaces between words but secondary segmentation between some or all morpheme or syllable boundaries compare to each other? How would they compare to full word spacing or syllable spacing without secondary segmentation? What can we learn about the additional burden, if any, of learning to write using different styles of secondary segmentation?

Another very interesting area of further research would be to compare syllable spacing and word spacing in non-Latin scripts, especially in Brahmi-derived alphasyllabaries that typically mark syllable boundaries more clearly than Latin script alphabets. Would the greater clarity of syllable boundaries in such scripts mean that syllable spacing would be redundant, and word spacing would show a greater benefit? What factors would influence the effect of spacing style, and would these be the same as with the Latin script orthographies studied in this thesis?

7.8 Summary of answers to research questions

This thesis began with the following three research questions:

1. Is word spacing optimal for Hmong Daw?
2. What word-related factors predict the choice between syllable spacing and word spacing by Hmong writers?
3. What factors predict any differences in reading speed between syllable spacing and word spacing in Hmong Daw?

For question 1, the results of this thesis, summarized in Section 7.3.2.3, indicate that word spacing is not more beneficial than syllable spacing to readers of Hmong Daw. Meanwhile, the results from the spacing survey questions in Section 5.4, the research on reading development and phonological awareness reviewed in Section 2.6, and the general challenges faced by writers across many languages in determining word boundaries (Kutsch Lojenga 2014, Eaton & Schroeder 2012, Cahill & Karan 2008) all suggest that syllable spacing is likely easier to learn for writers of Hmong Daw.

For question 2, the analysis of Hmong spacing practices in Section 4.4 suggests that several factors significantly influence the choice of spacing for Hmong writers, among them the number of morphemes, the number of syllables, the length of the first syllable, the presence of full reduplication, and the ratio of the frequency of the first syllable in the target word over its frequency in other words or on its own. Section 7.2 and 7.3.3 summarize the analysis from Section 4.4.

For question 3, the results from the word list test in the US and Thailand together provide various orthographic, morphological, and syntactic factors that influence the effect of spacing on reading speed. Some of these factors more clearly influence the effect of spacing than others. No demographic variables were found to influence the effect of spacing style. Details are found in Section 6.8 and 6.9 (see also Table 10), and summaries are given in Section 7.2 and 7.3.3.

7.9 Final remarks

More important than letters or the spaces between them is the content that fills them. I hope that the content of this thesis is helpful and encouraging to those who are involved with the development of writing systems for their own language or for others. And I hope that in some small way, this thesis can make mother-tongue reading and writing a more blessed and beautiful experience for those who have yet to enjoy this unique pleasure in life.