PAYAP UNIVERSITY
TRUTH • SERVICE

COMPARING THE READABILITY OF SYLLABLE SPACING AND WORD SPACING IN HMONG DAW

SETH VITRANO-WILSON

Presented in Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS
IN
LINGUISTICS

Payap University
March 2015

Title: Comparing the readability of syllable spacing and word spacing in Hmong Daw

Researcher: Seth Vitrano-Wilson

Degree: Master of Arts in Linguistics

Advisor: Elissa A. Ikeda, Ph.D.

Approval Date: 19 March 2015

The members of the thesis examination committee:

1. Isara Choosri ____ Committee Chair
(Isara Choosri. Ph.D.)

2. Elissa Ikeda ____ Committee Member
(Elissa A. Ikeda, Ph.D.)

3. R. Wyn Owen ____ Committee Member
(Robert Wyn Owen. Ph.D.)

ACKNOWLEDGEMENTS

I am very grateful to my thesis advisers, Dr. Elissa Ikeda and Dr. Wyn Owen, for all their tremendous help in shaping and refining this thesis, spending many hours listening, advising, reading, and revising. Thank you to Aj. Isara Choosri for coming to Chiang Mai to chair my defense committee, and for his very helpful suggestions and feedback. Thank you to Mr. Train Wu at the United Hmong Association of Massachusetts for providing the Hmong Daw stories used in Chapter 5, as well as for helping to find Hmong readers in New England. Dr. Lantzia C. Thao at the Hmong District of the Christian Missionary Alliance in Colorado helped to refine the text and verify its suitability for Hmong readers of different levels. Aj. Tshaj Sawm, Aj. Chokchai, and Khun Nruas in Chiang Mai helped to modify the text for the sentence test and find Hmong readers. Dr. David Mortensen graciously provided his processed version of the 15-million word online corpus from the soc.culture.hmong Usenet group. Thank you to Aj. Prasert, Aj. Daniel, Aj. Somphet, Aj. Prasit, Aj. Tharawat, Khun Xovzoo, Khun Ib, Anabel Rulison, Katherine Dooley, and Alicia Ho for helping connect me to Hmong readers in Chiang Mai.

Thank you to Sansanee Saedu for all her help in analyzing the Akha texts and teaching me the Akha orthography used. Thank you to Lori Vernon for connecting me to the Akha Bible School and arranging my visit. Thank you to Nancy Kukaewkasem and everyone at the Akha Bible School for their help and hospitality. Thank you to Ah Suhn and Aj. Ah Tsa for connecting me to Akha readers in Chiang Mai.

Thank you to Amber and Upai Jasa for analyzing the morphology of the Lahu Si texts, and for connecting me to Lahu Si readers. Khun Wirachai helped refine the tests and find readers for me, while Aj. Prasaeng graciously connected me to other Lahu Si leaders.

Thank you to Aj. Art for your guidance in searching for a topic and feedback on the proposal, and to Aj. Pam and Khun Yui for bringing me through the program and handling so many details. Thank you to Aj. Terry Gibbs for his help in formatting this document and preventing hours of consternation. Thank you to Marcus Rice for his help with statistics and proofreading. Aj. Sigrid Lew and Stephanie Wong helped shape the research and guide me to resources in the literature. Thank you to Aj. Mark Holt for guiding me as I explored the topic. Thank you to Liz Foerster for connecting me to multiple people in the Hmong and Lahu communities in Chiang Mai. Thank you to my friendly wife Lauren, who, among innumerable other things in life, provided invaluable help in recruiting readers, and to my father, Dr. Brent Wilson, for help in proofreading. A tremendous thank you to all the readers in both the US and Thailand who helped make this study possible. Most of all, thank you to the Maker of words, syllables, and everything else. Ua tsaug ntau ntau.

Seth Vitrano-Wilson

Title:	Comparing the readability of syllable spacing and word spacing in Hmong Daw
Author:	Seth Vitrano-Wilson
Degree:	Master of Arts in Linguistics
Advisor:	Elissa A. Ikeda, Ph.D.
Approval Date:	19 March 2015
Institution:	Payap University, Chiang Mai, Thailand
Number of Pages:	160
Keywords:	Hmong Daw, reading, orthography testing, interword spacing, syllable, word boundaries

ABSTRACT

Language groups making decisions about how to write their language must consider not only how to represent different sounds, but also how to use spaces. Globally, languages that use the Latin script typically use spaces between words, as in European languages. Most linguists writing guidelines for orthography development assume that word spacing is optimal, and focus instead on exactly where word boundaries should be located.

In Mainland Southeast Asia, the use of spaces is quite diverse, ranging from spaces between every syllable for Vietnamese to no spaces at all for Chinese. As a result, groups often consider a number of options for spacing. While much research has compared reading speeds for word-spaced and unspaced text, the effects of word spacing and syllable spacing on reading speed have rarely been compared.

Three experiments were performed to compare people reading word-spaced texts and syllable-spaced texts in Hmong Daw, using the Latin script orthography known as RPA. RPA text is commonly found in both word-spaced and syllable-spaced formats. Readers in both the United States and Thailand were tested.

The first experiment, with Hmong readers in the US, examined the effect of spacing style on reading speed for a set of stories. In the second experiment, Hmong readers in Thailand read the same stories as the first experiment, but the stories were presented and timed one sentence at a time. In the third experiment, readers in both the US and Thailand read isolated polysyllabic words with and without intersyllable spaces.

The results showed no overall difference in reading speed between the two spacing styles when people read naturally connected stories. However, syllable spacing was found to be faster than word spacing for the test of isolated words, and for sentences with polysyllabic words that readers had not yet seen within the test. Meanwhile, pilot testing in Akha showed a potential advantage to syllable spacing over word spacing.

In addition, an analysis of the spacing practices of Hmong writers was performed, to see which word-related factors predict whether writers are more likely to write polysyllabic words as a unit or with intersyllable spaces. The test of the reading of isolated polysyllabic words in Hmong Daw found various orthographic, morphological, and syntactic factors that make words easier to read as a unit or with spaces between each syllable. Together, these results provide the basis for suggestions of which types of words tend to benefit most from joining or separation with spaces. The results suggest that a purely linguistic definition of a word does not necessarily correlate with the optimal spacing style for ease of reading in Hmong Daw.

The lack of an advantage for word spacing in both Hmong Daw and Akha, and the advantage for syllable spacing in certain contexts, is contrary to the common assumption among linguists involved in orthography development that word spacing is the optimal spacing choice for all languages. This study suggests that syllable spacing is a valid option for certain languages in the right sociolinguistic situations.

ชื่อเรื่อง:	การเปรียบเทียบภาษามังเด๊อวที่เขียนโดยมีช่องว่างระหว่างพยางค์และช่องว่างระหว่างคำที่ส่งผลต่อการอ่าน
ผู้จัดทำ:	เสท วิทราโน วิลสัน
ปริญญา:	ศิลปศาสตรมหาบัณฑิต สาขาวิชาภาษาศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก:	ดร. เอลิซ่า เอ. อิเคดา
วันที่อนุมัติผลงาน:	19 มีนาคม 2558
สถาบันการศึกษา:	มหาวิทยาลัยพายัพ จังหวัดเชียงใหม่ ประเทศไทย
จำนวนหน้า:	160
คำสำคัญ:	Hmong Daw, reading, orthography testing, interword spacing, syllable

บทคัดย่อ

การตัดสินใจเลือกวิธีการเขียนในกลุ่มภาษาใดภาษาหนึ่งไม่เพียงแต่จะต้องพิจารณาเรื่องเสียงแต่ยังต้องพิจารณาเรื่องการเว้นวรรคด้วย ในทางแบบลักษณ์ภาษาพบว่า ภาษาทั่วโลกที่ใช้อักษรละติน เช่น ภาษายุโรป นิยมเว้นวรรคระหว่างคำ นักภาษาศาสตร์ด้านพัฒนาการอักขรวิธีสันนิษฐานว่าการเว้นวรรคคำคือรูปแบบการเขียนที่ดีที่สุด ด้วยเหตุนี้จึงทำให้ประเด็นเรื่องการกำหนดขอบเขตของคำว่าคำควรมีขอบเขตที่ไหนอย่างไรเป็นประเด็นที่ได้รับความสนใจในงานวิจัยด้านอักขรวิธีเรื่อยมา

ในภูมิภาคเอเชียตะวันออกเฉียงใต้ การเว้นวรรคคำมีได้ตั้งแต่การเว้นวรรคระหว่างพยางค์ในภาษาเวียดนาม ไปจนถึงการไม่เว้นวรรคเลยในภาษาจีน ด้วยเหตุนี้ในการเลือกวิธีการเขียนของภาษาในแถบเอเชียตะวันออกเฉียงใต้จึงต้องพิจารณากลวิธีการเว้นวรรคคำจำนวนมาก อย่างไรก็ตามแม้จะมีงานวิจัยเปรียบเทียบความเร็วในการอ่านของตัวบทที่มีการเว้นวรรค และไม่เว้นวรรคจำนวนมาก แต่แทบจะไม่มีงานวิจัยชิ้นใดเลยที่ศึกษาเปรียบเทียบความเร็วในการอ่านตัวบทที่มีการเว้นวรรคระหว่างคำ และการเว้นวรรคระหว่างพยางค์

งานวิจัยชิ้นนี้ ผู้วิจัยทำการทดลองสามการทดลองเพื่อเปรียบเทียบความเร็วในการอ่านข้อความที่มีการเว้นวรรคระหว่างคำกับข้อความที่เว้นวรรคระหว่างพยางค์ในภาษามังเด๊อว (Hmong Daw) โดยใช้ตัวบทที่เขียนด้วยตัวอักษรละติน RPA ซึ่งเป็นอักษรที่แพร่หลายในทั้งกลุ่มภาษาที่มีการเว้นวรรคระหว่างคำ และเว้นวรรคระหว่างพยางค์ กลุ่มตัวอย่างในการทดลองนี้ได้แก่ ผู้อ่านจากสหรัฐอเมริกา และประเทศไทย

การทดลองที่หนึ่งผู้วิจัยใช้กลุ่มตัวอย่างชาวอเมริกันศึกษาว่ารูปแบบการเว้นวรรคคำมีผลต่อความเร็วในการอ่านหรือไม่ จากการอ่านเรื่องจำนวนหนึ่ง ในการทดลองที่สองผู้วิจัยใช้กลุ่มตัวอย่างชาวไทยศึกษาว่าการเว้นวรรคคำมีผลต่อความเร็วในการอ่านหรือไม่ โดยให้ผู้ร่วมทดลองอ่านเรื่องชุดเดียวกันกับการทดลองที่หนึ่ง โดยผู้อ่านจะต้องอ่านทีละประโยคและถูกจับเวลาในการอ่าน

การทดลองที่สามผู้วิจัยให้ผู้อ่านทั้งชาวอเมริกันและชาวไทยอ่านคำหลายพยางค์ที่ถูกแยกออกเป็นคำ ๆ โดยมีทั้งคำที่มีและไม่มีการเว้นวรรคระหว่างพยางค์

ผลการทดลองพบว่าความเร็วในการอ่านตัวบทที่มีการเว้นวรรคทั้งสองแบบในกลุ่มตัวอย่างทั้งสองชาติไม่แตกต่างกัน เมื่ออ่านเรื่องที่มีการร้อยเรียงเป็นธรรมชาติ อย่างไรก็ตาม ในการทดลองให้ผู้ทดลองอ่านคำหลายพยางค์ที่แยกออกเป็นคำๆ และการทดลองให้ผู้ร่วมทดลองอ่านประโยคที่มีคำหลายพยางค์ในข้อความที่ไม่เคยเห็นมาก่อน ผู้วิจัยพบว่ากลุ่มตัวอย่างสามารถอ่านข้อความที่มีการเว้นวรรคระหว่างพยางค์ได้เร็วกว่าข้อความที่มีการเว้นวรรคระหว่างคำ ผลการทดลองนี้สอดคล้องกับผลการทดลองเบื้องต้นในภาษาอาข่า (Akha) ที่พบว่ามีการเว้นวรรคระหว่างพยางค์มีผลต่อการอ่านได้เร็วกว่าการเว้นวรรคระหว่างคำ

นอกจากนี้ ผู้วิจัยยังวิเคราะห์การเว้นวรรคของผู้เขียนชาวม้งเพื่อดูว่าปัจจัยที่เชื่อมโยงกับคำปัจจัยใดสามารถใช้บอกได้ว่าผู้เขียนจะเขียนคำที่มีหลายพยางค์ให้เป็นหน่วยเดียว หรือเว้นวรรคระหว่างพยางค์มากกว่า การทดลองอ่านคำหลายพยางค์ที่ถูกแยกออกจากกันในภาษาม้งเด๊อวแสดงให้เห็นว่ามีปัจจัยทางด้านอักขรวิธี สัทหน่วยคำ และวากยสัมพันธ์หลากหลายปัจจัยที่ส่งผลให้อ่านคำแบบที่เป็นหนึ่งหน่วย หรือที่มีการเว้นวรรคระหว่างพยางค์ง่ายขึ้น ผลการทดลองดังกล่าวทำให้คาดได้ว่าคำประเภทใดที่จะที่ได้ประโยชน์จากการรวม หรือเว้นวรรคด้วยช่องไฟ นอกจากนี้ผลการวิจัยยังชี้ให้เห็นว่าคำจัดกัดความของคำแต่เพียงอย่างเดียวอาจไม่สอดคล้องกับรูปแบบการเว้นวรรคที่ดีที่สุดในการอ่านภาษาม้งเด๊อวด้วย

การที่การเว้นวรรคคำในภาษาม้งเด๊อวและภาษาอาข่าไม่ได้ช่วยให้การอ่านเร็วขึ้น แต่เป็นการเว้นวรรคระหว่างพยางค์ในบางบริบทที่มีผลต่อความเร็วในการอ่านขัดแย้งกับสมมติฐานเบื้องต้นของนักภาษาศาสตร์ด้านพัฒนาการอักขรวิธีที่สันนิษฐานว่าการเว้นวรรคคำเป็นรูปแบบการเขียนที่ดีที่สุดในทุกภาษา งานวิจัยชิ้นนี้ชี้ให้เห็นว่าการเว้นวรรคระหว่างพยางค์เป็นทางเลือกที่ใช้ได้กับภาษาอีกจำนวนหนึ่งในสถานการณ์ทางภาษาศาสตร์สังคมที่เหมาะสม

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

AICc — The corrected Akaike Information Criterion, also known as Hurvich and Tsai's criterion, a measure of the relative fit of a statistical model to a set of data. A model with a lower (better) AICc number fits the data more accurately than models with higher AICc numbers, and is more likely to reflect real statistical relationships.

ANOVA — "Analysis of variance," a statistical method that divides the variance in a set of data into separate components coming from different sources

CAO — The Common Akha Orthography, an orthography for Akha developed in 2008 and 2009 that uses final tone letters and word spacing

CQD — The Chuanqiandian orthography, common among Hmong varieties in China, for which syllable spacing is standard

L1 — First language, mother tongue

MSEA — Mainland Southeast Asia, the region consisting roughly of Myanmar, Thailand, Cambodia, Laos, Vietnam, and southern China

MT — Mother tongue

RPA — Romanized Popular Alphabet, the most commonly used Latin script orthography for Hmong Daw in the US and Thailand

SCH — soc.culture.hmong, an online Hmong group from which a 15-million word corpus of Hmong text was derived. This corpus was used to analyze Hmong spacing practices, as well as to create input variables into the models for the reading tests in Hmong Daw.

SD — Standard deviation

SPSS — Statistical Package for the Social Sciences, a statistical software package created by IBM

UBS NT — The United Bible Societies New Testament, a translation in Hmong Daw used especially by Protestant Christians (United Bible Societies 2000)

US	United States
C	Consonant, in syllable structure descriptions
G	Glide, in syllable structure descriptions
N	Nucleus, in syllable structure descriptions
T	Tone, in syllable structure descriptions
V	Vowel, in syllable structure descriptions
(n)	Noun, in glosses
β_{stand}	The standardized β, or standardized regression coefficient, is a statistic that indicates effect size. It is standardized by making the variance of all effects equal to 1, so that direct comparisons between effects can be made. An effect with a larger absolute value for β_{stand} is a stronger effect. Positive and negative numbers indicate effects working in opposite directions.
p	The p-value is a statistic that indicates whether the null hypothesis (i.e., that there is no effect of the variable) should be rejected or not, based on a given significance level or α. For this thesis, an α of .05 was used, meaning that the null hypothesis was rejected when there was a less than 5% chance of it being true.
( )	In syllable structure descriptions, indicates that the element in the parentheses is optional
/ /	Phonemic examples
< >	Orthographic examples, focusing on the spelling itself
italics	Examples from non-English Latin script languages, given in orthographic form, when the spelling itself is not under discussion
.	Syllable boundary in phonemic examples. In glosses, a period indicates two ideas in English that are conveyed by one morpheme in the language being glossed.
-	Morpheme boundary in glosses and phonemic examples. Syllable boundary in Hmong examples, if necessary to disambiguate from word or morpheme boundaries.
(space)	Syllable boundary in Hmong, Vietnamese, and Lao examples, except when the text indicates otherwise.
?	In glosses, an uncertain or unknown meaning for a given morpheme

GLOSSARY

Alphasyllabary

A writing system in which consonants have an inherent vowel which can be modified by diacritics or some other vowel symbol. Unlike syllabaries, the symbol for each consonant and vowel sound is consistent across syllables. Unlike alphabets, an unmarked consonant symbol always includes a vowel, so that there is usually no way of representing independent units below the syllable level. Alphasyllabaries are also known as *abugidas*, and include Brahmi-based scripts in South and Southeast Asia.

Box-Cox power transformation

A transformation of the data that preserves rank, but can make the distribution of the data closer to a normal curve. All Box-Cox transformed data in this thesis was re-transformed back to the original units when presenting results.

Boxplot

A plot that gives information on the distribution of values within a data set. The central bar represents the median, and the main box represents the central 50% of values (that is, the second and third quartiles). Lines extend above and below the box to either 1.5 times the "interquartile value" (that is, the difference between the 25th and 75th percentile figures), or to the minimum or maximum values, whichever are smaller. For boxplots in this thesis, values between 1.5 and 3 times the interquartile value are shown with a circle, and values more than 3 times the interquartile values away from the central box are shown with an asterisk.

Confidence interval The range within which the true mean of a population is likely to lie. All confidence intervals given in this thesis are 95% confidence intervals, meaning that if the experiment were repeated many times, 95% of the observed confidence intervals will hold the true value for the parameter.

Coordinate compound A two-part compound word whose parts are semantically closely related, either as synonyms or antonyms. Hence they are sometimes called the product of "semantic reduplication".

Effect size The magnitude of the effect of an independent variable on a dependent variable, as measured by β_{stand}. The larger the effect size, the more the dependent variable changes when an independent variable changes a given number of standard deviations from its original value.

Elaborate expression A special type of four-syllable expression found in many languages of Mainland Southeast Asia. Typically, they include repeated elements of the form ABAC or ABCB. Sometimes the non-repeated elements together form a coordinate compound, or are related by partial reduplication. They typically show a fixed, lexicalized form, so that the order cannot be changed and elements cannot be replaced with other synonyms.

Fisher's exact test A statistical test of significance useful for comparing small numbers of categories to one another.

Fossilized morpheme A morpheme that has no meaning on its own, but derives its meaning from the meaning of the word as a whole, and without which the word would mean something else. For instance, the morpheme "mul" in "mulberry" has no meaning on its own, but it serves to distinguish a "mulberry" specifically from a "berry" in general. Typically, fossilized morphemes used to mean something in the past, or meant something in another language before it was borrowed, but has lost its independent meaning and distribution; hence the term "fossilized".

Fovea
The area at the center of vision where visual perception is most acute. Typically, the eyes gather most of the information while reading from the foveal area.

Grapheme
The smallest unit used in a writing system (i.e. a letter or symbol).

Hanzi
Chinese characters, the traditional writing system for Mandarin Chinese

Hurvich and Tsai's criterion
See AICc in abbreviations and symbols.

Ideophone
A word that represents an idea through its sound. Ideophones are often onomatopoeic, but can also use sound to represent light, color, speed, or any other aspects of an idea.

Interaction effect
An effect in which one independent variable influences the effect of another independent variable on the dependent variable. For example, an interaction effect between the number of morphemes and spacing style would mean that monomorphemic words would show a certain difference between word spacing and syllable spacing (e.g. no difference), whereas polymorphemic words would show a different pattern (e.g. an advantage for syllable spacing).

Intersyllable spaces
Spaces between syllables

Interword spaces
Spaces between words

Latin square design
A Latin square is a two dimensional array, with the same number of rows and columns, where every value occurs exactly

once in each row and each column. In experimental design, a Latin square design is useful for making sure that each participant receives each treatment exactly once.

Linear regression
A statistical approach to model the effect of one or more independent variables on a (scalar) dependent variable, using a sum of linear functions.

Major syllable
A syllable that allows all the normal phonotactic possibilities for a language.

Marginally significant
For this thesis, marginal significance is defined as having a p-value of 0.1 or less. (See Significant, p-value.)

Minor syllable
Syllables that are phonologically bound to a major syllable, and are in some way phonologically reduced relative to the major syllable. Typically, they show reduced contrast in tone, vowel quality, consonant clusters, or final consonants.

Morpheme
The smallest component of a word that has semantic meaning. Thus a monomorphemic word has one unit with semantic meaning. For example, the English word "carpet" has two syllables but is a single semantic unit. Polymorphemic words, such as "carpool", contain more than one semantic unit.

Morphological decomposition route
A route for cognitive processing during reading in which readers determine the meaning of a word by first determining the meaning of the individual morphemes, and then using that information to derive the meaning of the whole word.

Morphosyllabic
Relating to a one-syllable morpheme, such that the syllable and morpheme boundaries coincide.

Multicollinearity
A phenomenon that occurs when two or more independent variables in a statistical model are highly correlated. This makes the coefficient estimates for these variables highly

suspect, although it does not affect the estimates for other, non-correlated parameters in the model.

Multilevel model with crossed random effects

A statistical model in which each data point is embedded within a hierarchical structure containing different levels, and in which at least one level contains two or more random effects. For this thesis, Level 1 would be each individual reading time observation. Level 2 would contain the two random effects of reader and linguistic item (that is, a word or a sentence). In other words, it is assumed that each reader will have his or her own idiosyncrasies that will influence their reading time, not all of which can be included and accounted for in the model. Similarly, each word or sentence will have its own idiosyncrasies. Since each observation time belongs to both a reader and an item, but readers do not "belong" to items nor items to readers, these effects both operate on the same level and are therefore "crossed". For the readers in Thailand who read each sentence separately, there exists another level, Level 3, containing the "story" variable, within which each sentence is embedded. See Baayen et al. (2008) for more information on multilevel models in psycholinguistic research.

Multilevel ordinal logistic regression

A type of multilevel model designed for noncontinuous dependent variables. "Ordinal" means that the dependent variable (for this thesis, comprehension score) can have multiple discrete, ordered values, such as 0, 0.5, or 1.

Null hypothesis

In statistical testing, the null hypothesis refers to a default position that there is no relationship between two variables of interest.

Optimal

For the purposes of this thesis, an orthography is described as "optimal" for reading if it results in the greatest readability (that is, the fastest reading time and greatest comprehension).

Since comprehension did not prove to vary in the results for this thesis, optimal spacing styles for reading relate to reading speed alone. An orthography is "optimal" for writing if it takes the least effort for writers to learn and use.

Optimal viewing position

A location to the left of center of a word (for left-to-right scripts), where readers of languages with interword spaces tend to land after a saccade, which enables the greatest visual access to the word overall.

Orthography

A standard system for writing a language, including choice of script; choice of symbols for phonemes, syllables, morphemes and/or words; choice of spacing and other punctuation; and any other choices needed to represent the language in writing in a standardized way.

P-value

The *p*-value is defined as the probability, under the assumption that the null hypothesis is true, of obtaining a result equal to or more extreme than what was actually observed.

Parafovea

The area just outside the fovea, but still close enough to perceive certain key forms and other information helpful to reading.

Paraorthography

The set of symbols in an orthography that do not represent sounds directly, but instead help readers understand how graphemes are joined and separated to form larger linguistic units, and how the overall meaning should be interpreted. The paraorthography includes punctuation, (small) spaces between graphemes, spaces between words, paragraph divisions, section headers, chapter breaks, and so on.

Percentage difference

The percentage difference between two values is defined as the difference between the values divided by the average of the two values. In this thesis, any percentage comparisons word spacing and syllable spacing reading times or speeds are based on this method of calculating the percentage difference.

Periphery
The area outside the parafovea, where only the most basic forms are perceived in reading.

Phoneme
The smallest unit of sound which represents meaningful contrast.

Phonotactics
The permitted syllable structure of a language, and which sounds are permitted for which elements of the syllable. It can also relate to what types of structures are allowed in minor vs. major syllables, and what types of syllables are allowed to join into words.

Presyllable
A minor syllable occurring before the major syllable in a word.

Pseudohomophone
Nonwords with one or more graphemes changed, but with the same sound as a real word. An English example would be "brane" instead of "brain".

Psycholinguistic grain size
The phonological level to which readers most easily map symbols for different orthographies, depending on which phonological levels are most saliently and consistently available for that orthography. A fuller definition is found in Section 2.4.8.

Readability
The ease of reading text in a given orthography, considering both reading time and comprehension, with emphasis given to reading time.

Reduplication
Reduplication is a morphological process in which a word or part of a word is repeated. Reduplication of nouns often indicates plurality, and reduplication of verbs often indicates repeated action, though not always. **Full reduplication** repeats the entire word, whereas **partial reduplication** repeats only

part of the word. For isolating languages in Mainland Southeast Asia, partial reduplication usually means that the word is repeated but with a change in the vowel or tone.

Result of practical importance
It is possible that statistical analysis shows the difference between reading speed for two different spacing regimes to be statistically significant. However, statistics cannot confirm whether or not the difference is large enough to be of practical importance. For this, those with experience in the field must make a subjective judgment based on their experience as to whether the difference is meaningful or not.

Saccade
A movement of the eyes while reading, in which the eyes jump from one focal point to another.

Saccade landing position
The location on which the eyes focus after a saccade.

Script
A set of symbols and associated conventions, often shared in common by orthographies for multiple languages, used for writing a language or group of languages. For instance, the English orthography uses the Latin script, along with most other European languages, while the Arabic script is used to write Arabic, Persian, and many other languages.

Scriptura continua
Text which does not use spaces or any other segmentation device to mark linguistic boundaries, but runs graphemes together in succession.

Segmentation, text segmentation
The division of text according to linguistic boundaries by using spaces, hyphens, apostrophes, periods, commas, or other punctuation.

Semantic opacity, semantic transparency

The extent to which the meaning of a compound word is easily discernible from the meaning of its constituents. A compound whose meaning is a simple sum of its parts is semantically transparent, while one that is not is semantically opaque. Semantic opacity and transparency are not dichotomous but rather ends on a spectrum.

Sesquimorphemic

A word that has two morphemes, one of which can occur in other words and has a known meaning that is related to the meaning of the whole word, and one of which is meaningless or has uncertain meaning. The meaningless or uncertain morpheme may be a fossilized morpheme.

Sesquisyllabic

A word or morpheme with one major syllable and one minor syllable.

Significant

For the purposes of this thesis, a "significant" or "statistically significant" result is one for which the result of a statistical significance test (also known as "p value") is less than .05. In other words, I use an α of .05 for this thesis. To say a result is significant at the .05 level means that there is a 1 in 20 chance that the null hypothesis is really true and the (apparently extreme) result obtained was by chance. See also "Result of practical importance".

Statistical power

The statistical power of a test is the probability that it correctly rejects the null hypothesis when there actually is a real relationship between two variables. In general, a test with a larger sample size has greater statistical power, and tests with less error also have greater power.

Syllabary

A syllabary is a writing system in which each character or grapheme represents a syllable in spoken language. Unlike alphasyllabaries, there is no predictable relationship between syllables that share common consonants or vowels, but each syllable has its own distinct symbol. Examples are the Japanese *hiragana* system and the Yi syllabary of China.

Syllable-spaced
A syllable-spaced text or orthography uses spaces to mark syllable boundaries.

Tetragraph
Four graphemes (e.g. letters) that represent a single phoneme.

Tone sandhi
Alternation of tone due to phonological or morphological processes in a language. For Hmong Daw, there is no phonological tone sandhi described by linguists, and all tone sandhi is morphological. This means that an expression showing tone sandhi between two syllables is defined as a compound word.

Visual acuity benefit
The benefit gained from having a whole word in the foveal region of the retina, where sensitivity is the greatest. Longer words that extend into the parafovea must typically be read in more than one fixation, slowing down reading and making a morphological decomposition route more likely (Bertram & Hyönä 2003).

Whole word route
A route for cognitive processing during reading in which readers determine the meaning of a word by accessing the entire word in a mental lexicon, rather than by processing it in smaller chunks first.

Wilcoxon signed-rank test
A non-parametric statistical test designed to compare two related samples, to see if their means differ. It is mainly used as an alternative to the paired Student's t-test when data is not normally distributed.

Word-spaced
A word-spaced text or orthography uses spaces to mark word boundaries.

Writing system	A system for writing a language. This can be general and refer to a type of system, such as "alphabetic writing systems," or to the system used to write a particular language. It includes the characters or symbols used, as well as the way these symbols are put together in writing. While essentially synonymous with orthography (and used synonymously in this thesis), the term "writing system" focuses more on the general properties of the system used for writing, whereas "orthography" tends to evoke the concept of standardization.