The 9th International Conference on Cognitive Science

Reading and writing in Southeast Asian languages

Heather Winskel*

Psychology, Southern Cross University, Coffs Harbour Campus, Hogbin Drive, Coffs Harbour, 2450, Australia

Abstract

Traditionally, research on reading and writing has focused on a limited number of European languages, in particular English. More recently, there has been a growing interest in conducting research on more diverse languages and scripts. There is a dearth of research conducted on the languages of Southeast Asia. By including these languages and scripts, we can build more comprehensive and representative universal models of reading and writing. In this paper, the characteristics of the languages and writing systems of Southeast Asia will be briefly reviewed. Subsequently, some research that has been conducted on Thai and Malaysian/Indonesian will be focused on. Finally, some suggestions for future research will be made.

Keywords: Reading; Research; Writing; Southeast Asian languages

1. Introduction

Traditionally, research on reading and writing has focused on a limited number of European languages, in particular English, which can be considered to be an "outlier" orthography due to its high degree of inconsistency or irregularity ([1] Share, 2008). More recently, there has been a rapidly growing interest in conducting research on more diverse languages and scripts. There has been quite considerable research focused on East Asian languages, namely Chinese, Japanese and Korean, but much less research has been conducted on the languages of Southeast Asia. By including a broader range of languages and scripts, we can build more comprehensive and representative universal models of reading and writing. At the same time, we can distinguish between what are universal and what are orthography-specific processes and mechanisms in reading.

The aim of this paper is to review some of the reading research conducted on the languages of Southeast Asia. These languages and their writing systems are of growing interest to researchers as they differ significantly and in intriguing ways from the more commonly studied European languages. An important additional goal of the current paper is to promote future research on the rich and diverse languages of Southeast Asia and their corresponding writing systems.

In this paper, initially, the characteristics of the languages and writing systems of Southeast Asia will be briefly reviewed. Subsequently, some of the research that has been conducted on Thai and Malaysian/Indonesian will be focused on. As this region is ideal for research on bilinguals and biscriptals, this area of research will also be examined. Finally, some suggestions for future research will be made.

---

* Corresponding author. Tel.: +61 2 6659 3381; fax: +61 2 6659 3202.
E-mail address: heather.winskel@scu.edu.au

2. Characteristics of the languages and writing systems of Southeast Asia

Southeast Asian languages show striking convergence in terms of structure at all linguistic levels (Enfield, 2005 [2], 2011 [3]; Matisoff, 1973 [4], 2001 [5]). These languages are commonly analytic or isolating, thus, lack inflectional morphology (i.e., do not have agreement, case, gender/number/definiteness on noun phrases, tense-marking on verbs), have a tendency to be monosyllabic, have zero anaphora, rich inventories of sentence final particles, numeral classifiers, verb serialisation, and topic comment structure is favoured (Enfield, 2011 [6]). In addition, lexical tone is an important feature of many of these languages. Thus, most Southeast Asian languages do not inflect for tense, gender and number. In place of inflectional morphology, the Southeast Asian languages typically utilise separate functor words or lexemes.

The writing systems of this region, which include Javanese, Thai, Lao, Khmer, Myanmar, Vietnamese and Tibetan, offer extremely fascinating opportunities for psycholinguistic research. Chinese, English and Tamil are also widely spoken and read in the region. Some of the scripts of Southeast Asia share common features as they are derived from ancient Brahmi script. Furthermore, some of the Southeast Asian scripts do not have interword spaces (e.g., Thai, Lao, Khmer and Myanamar) and others represent tone orthographically (e.g., Thai, Lao, Vietnamese and Myanamar).

3. Characteristics of Thai and some reading research

Thai, a tonal language, has its own distinctive alphabetic script that shares some common characteristics with Indic writing systems, due to common origins. It also has syllabic characteristics. Vowels are not always explicitly specified, instead, inherent vowels can occur with the consonant, as an example an *a* is usually found in words of Sanskrit, Pali, or Khmer origin (e.g., ตลาด /t[a]la:t/ 'market'), whereas an *o* is found in native Thai words (e.g., ลม l[o]m 'wind') (the inherent vowels are in parentheses). In contrast to sequential Roman script, Thai has non-linear characteristics as vowels can be written above, below, or to either side of the consonant as full letters or diacritics, and which commonly combine across the syllable to produce a single vowel or dipthong. There are also five commonly used vowels that precede the consonant in writing but follow it in speech, for example using an English example *epn* would be spoken as *pen*. These types of vowels can operate within the syllable (e.g., แบน <ɛ:bn> 'flat' is spoken as /bɛ:n/) or across the following syllable in a more severely misaligned example (e.g., the word แมลง <ɛ:mlŋ> insect' is spoken as /m(a)lɛ:ŋ/ where (a) represents an inherent vowel). (For more detailed information about the characteristics of Thai see Winskel, 2009 [7], Winskel and Iemwanthong, 2010 [8] or Winskel, in press [9]).

We could expect a processing cost to be associated with these types of misaligned vowel words, where there is a mismatch between the written and spoken forms. In order to investigate this possibility, Winskel (2009) [10] conducted a series of experiments with adults and children. Eye movements of adults reading words with and without misaligned vowels in sentences using the EyeLink II tracking system was conducted. Participants read pairs of words with misaligned and aligned vowel words matched for length and frequency embedded in same sentence frames. In addition, rapid naming data from adults and reading and spelling comparable words in 6;6 to 8;6 years old children was also collected. Results indicated that there was a processing cost associated with reading and spelling the more severely misaligned words where the vowel operates across the syllable, but not in general for the misaligned words where the vowel operates within the syllable in both adults and children. The misaligned words (where the misaligned vowel operates within the syllable) were not significantly different from the aligned control words, where the written and spoken forms match (e.g., มะม่วง /mamuaŋ/ 'mango').

Another line of research, based on this characteristic of Thai, investigated if the privileged role of initial letter position in visual-word recognition is a universal phenomenon. This claim is based on research conducted on European languages (e.g., Chambers, 1979 [11]; Estes, Allmeyer, & Reder, 1976 [12]; Gomez, Ratcliff, & Perea, 2008 [13]; Rayner & Kaiser, 1975 [14]; White, Johnson, Liversedge, & Rayner, 2008 [15]). In European languages, transposed-letter effects tend to be small or negligible when the initial letter is involved, as an illustration, *jugde* closely resembles *judge* while *ujdge* does not. However, we suspected that this might not be the case for Thai with its non-linear configuration of vowels. Letter position coding needs to be quite flexible so that readers can encode the letter positions of words with and without misaligned vowels. In a recent lexical decision experiment, Perea, Winskel and Ratitamkul (2012) [16] found a significant masked transposed-letter priming effect in Thai when the

initial letter was transposed even in very short words (e.g., บท-บาท was faster than พะท-บาท [transposed-letter condition vs. replacement-letter condition]), which does not occur in Roman script.

In order to follow-up these research findings, a more ecologically valid eye tracking paradigm was used to investigate whether the position of transposed letters (initial vs. internal) within a word influences how readily those words are processed during normal silent reading (Winskel, Perea, & Ratitamkul, 2012 [17]). There was no apparent difference in degree of disruption caused when reading internal and initial transposed-letter nonwords. This pattern of findings is in marked contrast with results found in Roman script where, consistent with prior evidence from other paradigms in European languages, there is greater disruption caused by initial than internal transpositions (White et al., 2008). In sum, these experiments indicate that the initial letter position in Thai is not as critical as in Roman script when reading and there is a degree of flexibility in letter position coding in Thai. These results support the view that the orthographic coding scheme is not universal but rather is modulated by the specific characteristics of the language (e.g., Velan & Frost, 2011 [18]).

Another characteristic of Thai is that it does not have interword spaces (for example, เมื่อวานนี้เขาเอาหนังสือไปห้องสมุดวิทยาลัยในจังหวัดอยุธยา). Winskel, Radach and Luksaneeyanawin (2009) [19] investigated the function of interword spaces and whether inserting spaces between words facilitates reading in Thai. The eye movements of Thai-English bilinguals when reading sentences with high and low frequency target words embedded in same sentence frames with and without interword spaces were examined. Inserting spaces between Thai words shortened reading times on target words, even though Thai script does not naturally include such spaces. Several of our findings suggested that spacing facilitates later word processing rather than word targeting or early lexical segmentation. The addition of interword spaces resulted in shorter refixation measures (gaze duration and total fixation duration), but did not affect first fixation duration. Moreover, first fixation landing positions and landing site distributions were just left of word centre in both the spaced and unspaced conditions. These results in conjunction with the lack of difference found for initial fixation duration, suggest that word targeting and early lexical segmentation is not facilitated (or disrupted) by the insertion of interword spaces, although later word processing including lexical access was facilitated. Thus, there was qualified support for a facilitatory function of interword spaces when reading Thai.

In other research, we have focused on additional features of Thai namely lexical tone, as tone is explicitly represented in Thai script (Winskel & Perea, under review [20]) and the relative contribution played by consonants and vowels in early visual-word recognition (Winskel & Perea, 2013 [21]).

4. Characteristics of Malay/Indonesian and some reading research

A variety of the Malay language is spoken in four Southeast Asian countries, namely Indonesia, Malaysia, Singapore and Brunei (Prentice, 1987 [22]). Malaysian/Indonesian language provides an interesting case study as it uses the same Roman script as English, but in contrast has a high degree of orthographic transparency. Furthermore, the syllable is a highly salient unit as it is predominantly bi- and multi- syllabic and has a simple syllable structure with clear syllable boundaries. The salience of the syllable in Malaysian/Indonesian is reflected in the teaching method adopted, which predominantly focuses on teaching children the correspondences between whole spoken and written syllables rather than the correspondence between letters and phonemes.

Winskel and Widjaja (2007) [23] focused on the grain size predominantly used by children when learning to read and spell in Indonesian. A range of tasks assessing different levels of phonological awareness as well as letter knowledge, reading familiar words and nonwords, and spelling stem and affixed words, were administered to 73 children in Grade 1 and subsequently one year later in Grade 2. The results, in general, indicated that the phoneme was the prominent unit in the early acquisition of reading and spelling in Indonesian, as it was found to be a concurrent predictor of reading for both word and nonword reading for Grade 1 and Grade 2 children. Furthermore, an analysis of word and nonword reading errors revealed that errors were predominantly nonword or phonological errors, which supports this level of processing.

However, the pattern was more complex, as revealed by results from Grade 1 measures as predictors of reading and spelling in Grade 2. The phoneme was a significant predictor of reading for nonword reading but not for reading familiar words or for spelling stem words. Of particular note, letter knowledge was found to play a prominent role for word reading and spelling stem words, as presumably the close correspondence between letter names and sounds enables access to letter-sound relations in these tasks. However, when the task was to spell multisyllabic affixed

words, an awareness of both phonemes and syllables appeared to be advantageous and facilitated this process. The results, in general, indicated that the phoneme was the prominent phonological unit in the early acquisition of reading and spelling in Indonesian, but the syllable also played a significant role, particularly when reading long multisyllabic affixed words. It appears that the transparency of the language and the close correspondence between letter names and sounds facilitates access to the smallest grain size, the phoneme, in Indonesian beginner readers. Moreover, the syllable and morpheme are salient units in Indonesian, consequently they are accessed and utilised by the child, particularly when reading and spelling challenging long, multisyllabic words.

Lee and Wheldall (2011) [24] further investigated word reading, letter knowledge and phonological awareness in 46 Grade 1 Malaysian children. Eleven of these children were identified as low-progress readers. Stepwise multiple regressions revealed that the syllable was the most influential predictor but the phoneme also played a significant role in word reading. Children's reading performance on words with different syllable structures was also examined. Words with a simple open CV syllable structure were found to be easier to decode than words with digraphs, diphthongs, or the vowel *e*. As the complexity of syllabic structure increased, there was a corresponding decline in performance. The position of phonemes in a word was also found to affect word recognition performance. Words with a digraph at the end (e.g., *batang*) were easier to decode than words with a digraph at the beginning (e.g., *syarikat*). Moreover, words with two vowel graphemes belonging to different syllables appearing together in the middle of a word (e.g., *soal*) or at the end of a word (e.g., *tua*) proved problematic to beginner readers, due to confusion over the location of the syllable boundary. Furthermore, it was found that shorter stem words were easier to read than longer multisyllabic words with derivational affixes (Winskel & Lay Wah, in press [25]).

Interestingly, the Indonesian study (Winskel & Widjaja, 2007 [26]) did not reveal the same degree of prominence of the syllable in reading and spelling as the studies conducted on Malaysian (Rickard Liow & Lee, 2004 [27]; Lee & Wheldall, 2011 [28]). In fact the Indonesian study highlighted the phoneme as being the more prominent unit. One feasible explanation for this variation in results is due to differences in correspondence between the names of letters and the sounds they make. In Indonesian there is a direct correspondence, whereas in Malaysian the names of the letters are similar to those of the English alphabet and hence do not directly correspond. The close correspondence between letter names and letter sounds in Indonesian appears to facilitate access to the smallest grain size, the phoneme in beginner Indonesian readers. In contrast in Malaysian, where there is not a direct correspondence, the syllable plays a more prominent role than in Indonesian. This points to letter knowledge playing a significant role in learning to read and spell and the grain size that is accessed by beginner readers. Letter-name knowledge appears to help children acquire the alphabetic principle, that is, that written graphemes stand for phonemes in speech (Treiman, Tincoff, & Richmond-Welty, 1996 [29]). This effect is enhanced in languages such as Indonesian or Turkish (Öney & Durgunoğlu, 1997[30]), where letter names and letter sounds correspond to a much greater extent than English, as the names of the letters contain the phoneme that the letter typically represents.

5. Reading in bilinguals and biscriptals

There are excellent opportunities to examine processing and reading strategies in bilinguals and biscriptals in the Southeast Asian region. For example, Susan Rickard Liow and colleagues (Rickard Liow, in press [31]) have investigated the processes that contribute to spelling development in English for three groups of bilingual children (English-Mandarin, Mandarin-English and Malay-English) in Singapore. A series of empirical studies explored the processes that contribute to spelling in English for these three subtypes of bilingual children who all follow the same curriculum. Early spellings are speech-based (Jalil & Rickard Liow, 2008 [32]), and the three main languages (English, Mandarin, and Malay) are dissimilar in terms of phonology as well as orthography. These differences in the characteristics of the languages spoken and scripts learned by the different subtypes of bilinguals have been found to influence both phonological representation and spelling development.

6. Conclusion

In general, most models and theories of reading are based on a very limited number of languages and orthographies, predominantly European languages and Roman script, thus, it is essential to have data on a greater number of languages and their scripts. Southeast Asian region offers rich opportunities for future research on reading and writing in the different languages of the region. Future research could focus on reading in adults, the

development of reading in children, including reading in bilinguals and biscriptals and specific groups of children or adults.

References

[1] Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, 134, 584-615.

[2] Enfield, N. J. (2005). Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology, 34*(1), 181 206.

[3] Enfield, N. J. (2011). Dynamics of human diversity in mainland Southeast Asia: Introduction. In N. J. Enfield (Ed.), *Dynamics of human diversity: The case of mainland Southeast Asia* (pp. 1-8). Canberra: Pacific Linguistics.

[4] Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Ed.), *Southern California occasional papers in Linguistics, No. 1* (pp. 72-95). Los Angeles: University of Southern California.

[5] Matisoff, J. A. (2001). Genetic versus contact relationship: prosodic diffusability in Southeast Asian languages. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics* (pp. 291-327). Oxford: Oxford University Press.

[6] see [3]

[7] Winskel, H. (2009). Reading in Thai: The case of misaligned vowels. *Reading and Writing. An Interdisciplinary Journal, 22*, 1-24.

[8] Winskel, H., & Iemwanthong, K. (2010). Reading and spelling acquisition in Thai children. *Reading and Writing. An Interdisciplinary Journal, 23*, 1021-1053.

[9] Winskel, H. (in press). Learning to read and write in Thai. In H. Winskel & P. Padakannaya (Eds.). *South and Southeast Asian Psycholinguistics.* Cambridge: Cambridge University Press.

[10] see [7]

[11] Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior, 18*, 225-241.

[12] Estes, W. K., Allmeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics, 19*, 1-15.

[13] Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136*, 389-413.

[14] Rayner, K., & Kaiser, J. S. (1975). Reading mutilated text. *Journal of Educational Psychology, 67*, 301-306.

[15] White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1261-1276.

[16] Perea, M., Winskel, H., & Ratitamkul, T. (2012). On the flexibility of letter position   coding during lexical processing: The case of Thai. *Experimental Psychology. 59*(2), 68-73.

[17] Winskel, H., Perea, M., & Ratitamkul, T. (2012). On the flexibility of letter position coding during lexical processing: Evidence from eye movements when reading Thai. *Quarterly Journal of Experimental Psychology,* 65(8), 1522-36.

[18] Velan, H., & Frost, R. (2011). Words with and without internal structure: what determines the nature of orthographic and morphological processing? *Cognition, 118*, 141-156.

[19] Winskel, H., Radach, R., & Luksaneeyanawin, S. (2009). Eye movements when reading spaced and unspaced Thai and English: A comparison of Thai-English bilinguals and English monolinguals. *Journal of Memory and Language, 61*, 339-351.

[20] Winskel, H., & Perea, M. (2013). Letter position coding of consonants and vowels during normal reading: Evidence from parafoveal previews in Thai. *Journal of Cognitive Psychology, 25*(1), 119-130.

[21] Winskel, H., & Perea, M. (under review). Does tonal information affect the early stages of visual-word processing in Thai? *Quarterly Journal of Experimental Psychology.*

[22] Prentice, D.J. (1987). Malay (Indonesian and Malaysian). In B. Comrie (Ed.) *The world's major languages.* London: Croom Helm.

[23] Winskel, H., & Widjaja, V. (2007). Phonological awareness, letter knowledge and literacy development in Indonesian beginner readers and spellers. *Applied Psycholinguistics, 28*, 21-43.

[24] Lee, L. W., & Wheldall, K. (2011). Acquisition of Malay word recognition skills: Lessons from low-progress early readers. *Dyslexia, 17*(1), 19-37.

[25] Winskel, H., & Lee, L. W. (in press). Learning to read and write in Indonesian/ Malaysian: A transparent alphabetic orthography. In H. Winskel & P. Padakannaya (Eds.). *South and Southeast Asian Psycholinguistics.* Cambridge: Cambridge University Press.

[26] see [23]

[27] Rickard Liow, S. R. J., & Lee, L. C. (2004). Metalinguistic awareness and semi-syllabic scripts: Children's spelling errors in Malay. *Reading and Writing. An Interdisciplinary Journal, 17*, 7-26.

[28] see [24]

[29] Treiman, R., Tincoff, R, & Richmond-Welty, E. D. (1996). Letter names help children to connect print and speech. *Developmental Psychology, 32*, 505-514.

[30] Öney, B., & Durgunoğlu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. *Applied Psycholinguistics, 18*, 1-15.

[31] Rickard Liow, S. (in press). Diversity in bilingual children's spelling skill development: The case of Singapore. In H. Winskel & P. Padakannaya (Eds.). *South and Southeast Asian Psycholinguistics.* Cambridge: Cambridge University Press

[32] Jalil, S., & Rickard Liow, S. J. (2008). How does home language influences early spellings? Phonologically plausible errors of diglossic Malay children. *Applied Psycholinguistics, 29*, 535-552.