Bilingualism, Biliteracy and Metalinguistic Awareness: Word Awareness in English and Japanese Users of Chinese as a Second Language

Benedetta Bassetti*

Birkbeck, University of London

Abstract

Cross-linguistic research shows that some aspects of metalinguistic awareness are affected by characteristics of different writing systems. Users of writing systems that mark word boundaries (such as English) develop word awareness, while users of unspaced writing systems (such as Chinese) do not. Previous research showed that English-speaking users of Chinese as a Second Language (CSL) have higher levels of Chinese word awareness than Chinese monolingual speakers. The present study aimed at disentangling the effects of bilingualism/biliteracy and the effects of L1 writing system characteristics by comparing users of Chinese whose L1 writing system does or does not mark word boundaries. Three groups (Chinese, English-Chinese and Japanese-Chinese) performed two Chinese word segmentation tasks. Results showed significant differences among the groups in mean word lengths and mean level of intra-group agreement. The English group had higher intragroup agreement rates and shorter word lengths than the other two groups; the Japanese group had higher agreement rates than the Chinese group. It is argued that Chinese word awareness as revealed by word segmentation tasks is mostly affected by literacy in a writing system that marks word boundaries, but bilingualism and biliteracy also play a role.

Keywords: word awareness, metalinguistic awareness, Chinese, writing systems, bilingualism, biliteracy, second language writing systems, Japanese

_____

* E-mail address for correspondence: *benedetta@benedetta-bassetti.org*

Introduction

Research has consistently shown that bilingualism facilitates the development of metalinguistic awareness in children (Bialystok, 2001). Still, some aspects of metalinguistic awareness seem only to develop with literacy, and bilingualism has little or no effect on such aspects. Cross-linguistic research has revealed a relationship between the units represented in different writing systems and the metalinguistic awareness of users of these writing systems, showing for instance that users of phonemic writing systems show phonemic awareness whereas users of non-phonemic writing systems show little or no phonemic awareness (Perfetti & Zhang, 1991; Ventura, Kolinsky, Brito-Mendes, & Morais, 2001). It appears that when children learn to read, they learn to identify in the spoken language the same linguistic units they see in writing. In particular, one of these units is the orthographic word, and word awareness does not seem to develop either in illiterate adults or in literate adult speakers of languages whose writing systems do not mark word boundaries (Morais, Bertelson, Cary, & Alegria, 1986; Scholes, 1993a).

The Chinese writing system does not mark word boundaries, and previous research using word segmentation tasks has shown that Chinese literate adults do not develop metalinguistic awareness of words (Hoosain, 1992; Sproat, Shih, Gale, & Chang, 1996; Tsai, McConkie, & Zheng, 1998). Research also shows that English users of Chinese as a Second Language (CSL) outperform native speakers in Chinese word segmentation tasks (Bassetti, 2004, 2005). The present study is aimed at finding out whether English CSL users' higher levels of Chinese word awareness are due to literacy in a word-spaced writing system or to bilingualism. Chinese monolingual speakers were compared with two groups of CSL users: a group whose L1 writing system marks word boundaries and a group whose L1 writing system does not mark word boundaries. If bilingualism, biliteracy or second language instruction lead to increased word awareness, both CSL groups should outperform the Chinese group; if increased word awareness is a consequence of literacy in a word-spaced writing system, only the English group should significantly differ from the Chinese monolingual group.

Writing Systems and Metalinguistic Awareness

Different writing systems represent different linguistic units as discrete orthographic units: for instance, the graphemes of phonemic writing systems represent phonemes whereas the graphemes of syllabic writing systems represent syllables, and some writing systems mark word boundaries with spacing or other symbols whereas others do not (Cook & Bassetti, 2005).

Cross-orthographic research shows that users of different writing systems can identify and manipulate different linguistic units. Illiterate adults may never develop the ability to identify and manipulate certain linguistic units, such as phonemes (Ventura et al., 2001), whereas literate adults and children are generally aware of those linguistic units that are represented in their writing system. For instance, literate Japanese children cannot perform some phonemic awareness tasks (Leong, 1991; Mann, 1986), because their writing system represents morphemes and morae but not phonemes. Native speakers of the same language who are users of different writing systems perform differently in metalinguistic awareness tasks. For instance, Chinese adults who are literate in *hanzi* (Chinese characters, representing monosyllabic morphemes) cannot perform some phonemic awareness tasks (Perfetti & Zhang, 1991), but Chinese adults who learnt *pinyin*, a transcription system based on the roman alphabet, perform much better (Read, Zhang, Nie, & Ding, 1987). Equally, among Kannada-speaking children, those who are literate in the Kannada semi-syllabary cannot perform some phonemic awareness tasks, but blind children who are literate in an alphabetic *braille* perform much better (Padakannaya, 2000). This evidence suggests a relationship between writing systems and metalinguistic awareness. Metalinguistic awareness of at least some linguistic units appears not to be developmental, but to be a consequence of literacy. Furthermore, literacy per se is not sufficient, as only literacy in a writing system that represents certain linguistic units allows the development of metalinguistic awareness of such units.

Various researchers have noticed the relationship referred to here between writing systems' representations of language and metalinguistic awareness. The concept appears with different names, from 'graphic relativity' (Bugarski, 1993) to the 'Model Theory of Literacy' (Olson, 1994), and similar ideas have been proposed by various scholars (Aronoff, 1992; Derwing, 1992; Illic & Sanders, 1988; Scholes, 1993a, 1993b). Olson for instance wrote that 'writing systems provide the concepts and categories for thinking about the structure of spoken language' (p. 68), and that 'the models provided by our script tend to blind us toward other features of language' (p. 89). The term 'orthographic relativity' can be used (in analogy with 'linguistic relativity') as an umbrella term for the effects of writing systems on various aspects of cognition, ranging from metalinguistic awareness to representations of time (Kugelmass & Lieblich, 1979; Tversky, Kugelmass & Winter, 1991), perception of movement (Morikawa & McBeath, 1992) and perception of emotions (Vaid, 1995), among others (for an overview, see Cook & Bassetti, 2005).

The Orthographic Representation of Words and Word Awareness

Many writing systems use spacing or other symbols to mark the boundaries of words. These writing systems represent *orthographic words*, i.e. sequences of graphemes preceded and followed by spacing. Still, the marking of word boundaries is not a necessary feature of a writing system, neither diachronically nor synchronically. Many contemporary writing systems do not mark word boundaries, including some Brahmi-derived writing systems (e.g., Burmese, Thai, Tibetan) and some Chinese-derived writing systems (e.g., Chinese, Japanese, Korean). The orthographic word is also not ontogenetically or phylogenetically necessary. Ontogenetically, word boundaries were not always present, took very long to become established (Andrieux-Reix & Monsonego, 1998), and were sometimes introduced and then eliminated (as in ancient Latin and Greek, Parkes, 1993). Phylogenetically, children acquiring literacy in word-spaced writing systems take about two years to master the use of word boundaries, and children acquiring literacy in unspaced writing systems never develop the ability to segment texts into words (see below).

The English writing system has always been written with interword spacing, although until the 17th century this was inconsistent and partially prosody- rather than syntax-based. On the other hand, the Chinese and Japanese writing systems do not mark word boundaries. Chinese hanzi mostly represent monosyllabic morphemes; spacing is not used. Pinyin (romanised Chinese) is a supplementary writing system used in textbooks, dictionaries, software, etc. Although pinyin uses interword spacing, the rules have not been consistently determined yet and, notwithstanding official guidelines (National Education Commission, 1996), there is still confusion regarding the correct use of spacing. Japanese is a mixed writing system, composed of *kana* (syllabaries) and *kanji* (characters). Although some 19th-century western-educated scholars proposed to introduce interword spacing, the Japanese writing system does not use interword spacing, except for Braille and for texts intended for foreigners or primary school children. For Japanese children, texts are written in kana and spacing delimits phrases, whereas for learners of Japanese texts are romanized, and spacing delimits words (*tango wakachigaki*, 'word-spaced writing'), but there are no conventions for determining word boundaries.

The presence of orthographic words in a writing system determines the presence of word awareness in its users. English-speaking adults show metalinguistic awareness of words in their daily activities, and generally consider words as the units of language *par excellence*. When tested on their word segmentation skills, i.e. the ability to divide a spoken or written text

into words, literate English adults perform correctly (where 'correctly' means in line with orthographic conventions). On the other hand, English preliterate children do not understand that the word is a component of the spoken language (Downing, 1970); cannot distinguish it from phonemes, syllables, sentences and other spoken language elements (Ferreiro, 1997); do not understand that spacing segments text into words (Meltzer & Herse, 1969); cannot say whether prepositions, articles, auxiliaries and compounds are words (Barton, 1985; Tunmer, Bowey & Grieve, 1983); and segment language into words on the basis of meaning or prosody (Berthoud-Papandropoulou, 1978; Ehri, 1975; Holden & McGinitie, 1972; Tunmer et al., 1983).

While word awareness appears in English children around the age of 6, it can be argued that word awareness development and literacy are not just simultaneous, but rather literacy causes word awareness development. Children apply the term 'word' to the written text before applying it to the spoken language (Downing & Leong, 1982; Ehri, 1975; Francis, 1975), and when requested to define 'word', beginning readers refer to letters (Papandropoulou & Sinclair, 1974). Ferreiro (1999) provides an entertaining portrait of Spanish-speaking 2nd-year primary school children trying to understand that the spoken language contains the same number of words as the written language. Furthermore, if word awareness was developmental, all adults should be word-aware. Still, research on English- and Portuguese-speaking illiterate adults shows that they can identify nouns as words, but they do not consider function words as words, cannot segment frequently used phrases, and consider word segmentation a meaningless activity (Morais et al., 1986; Scholes, 1993a); Kannada-speaking illiterate adults also perform like preliterate children in word awareness tasks (Ramachandra & Karanth, 2007) . English-speaking adults with a reading age of 9 or 10 perform better, but do not consider articles and auxiliaries as words (Barton, 1985). While illiterates do not perform well on word segmentation tasks, aliterates (illiterate adults who live in a society without literacy) do not have a term for word or a corresponding concept. In such societies, there is no word for 'word', or if there is it refers to various units of speech, including morphemes, phrases and sentences (Goody, 1977). It appears that word awareness only develops in those who are literate in writing systems that mark word boundaries.

Chinese Word Awareness

Since hanzi represent monosyllabic morphemes, Chinese primary school children develop morphemic awareness (Li, Anderson, Nagy, & Zhang, 2002; Nagy et al., 2002), which does not develop in illiterate adults (Chao, 1968).

In China, hanzi rather than words are used for daily necessities such as calculating text length or checking dictionaries, and in the Chinese linguistic tradition hanzi were the main units of linguistic analysis, until Western linguistics was introduced in the 1950s. Just like for English adults the word is not simply a sequence of letters, so for Chinese speakers the hanzi is not simply an orthographic unit; it is also the building block of the language. The hanzi is what the famous linguist Chao (1968) calls the 'sociological word', i.e. what the Chinese society identifies as the 'word' (the linguistic unit) of their language. Chao repeatedly argued that in Chinese there is no equivalent of the English 'word' (Chao, 1968, 1976a, 1976b), and although a variety of wordhood tests have been developed, even linguists do not agree as to what constitutes a Chinese word (Bassetti, 2004). Given the fact that Chinese laypersons are already aware of a metalinguistic unit, which they use to talk and think about language, they do not need to be aware of words, just as English speakers do not need to be aware of morphemes to talk and think about language (see Ohala, 1992). Recently, mainly as a consequence of new technologies, the word is becoming more and more important in China and the term 'word' (*ci*) has become more widespread. Still, research shows differences between the word awareness of Chinese and English speakers.

Research on word awareness in Chinese adults has consistently shown lack of word awareness. This research was performed by psychologists and computational linguists, on both children and adults, using *word segmentation tasks*, whereby participants segment a spoken or written text into words. The Chinese psychologist Hoosain (1992) asked 14 Cantonese-Chinese bilinguals to perform a written word segmentation task on 9 sentences and a short text. None of the participants' segmentations coincided either with the researcher's or with anyone else's segmentations. Interestingly, participants often marked whole phrases as words. In another study, 138 Chinese participants segmented a short text (Tsai, McConkie, & Zheng, 1998). Half of the participants had learnt pinyin in school, the other half had not. The pinyin group marked more boundaries (i.e. shorter words) with lower variance (i.e. higher agreement). This shows that native speakers of the same language perform better in word segmentation tasks if they learned a word-spaced supplementary writing system. Miller and colleagues (Miller, Frosch, Kelly, & Zhang, 2001) investigated word segmentation skills in Chinese children, comparing them with a group of American children segmenting an English version of the same sentences. In the American children group, word segmentation accuracy rose from 70% in grade 1 to 95% in grade 3. The Chinese children group also showed some improvement, from 42% to 62%, but after 3 years of literacy they still performed less well than American grade 1 children (although it is not clear what counts as a 'correct' answer, unless this is based on the English translation of the text, which is not a reliable criterion, see Chao, 1968). Research on Chinese word

segmentation was also performed by computational linguists working on Chinese parsing software (Sproat, Shih, Gale, & Chang, 1996). These studies suffer from various limitations, the most notable being that instructions avoid using the word 'word' in the belief that Chinese participants would not understand the task. Still, these studies consistently show low interjudge agreement rates. Since word segmentation is taken to tap into the reader's intuition of what constitutes a word, it is possible to conclude that Chinese speakers are not aware of words.

Bilingualism, Biliteracy and Chinese Word Awareness

Bilingual children develop some aspects of metalinguistic awareness better, or at least earlier, than monolinguals (Cook, 1997), but not those aspects that are determined by literacy. Bilingual children are not better than monolinguals at phonemic awareness tasks (Bialystok, 2001; Bialystok, Majumder, & Martin, 2003; Bruck & Genesee, 1995). Bilinguals who are literate in a writing system that represents phonemes can use this awareness to analyse their other language as well (Ben-Dror, Frost, & Bentin, 1995; Loizou & Stuart, 2003) and may show higher L1 phonemic awareness than monolingual speakers of their L1 (Padakannaya, Rekha, Nigam, & Karanth, 1993), but illiterate bilinguals perform like monolinguals.

With regards to word awareness, bilingualism does not facilitate children's ability to identify words. Bilingualism facilitates another aspect of word awareness, namely lexical arbitrariness, i.e., the ability to separate the noun from its referent: bilingual children can better estimate the length of a word without considering the length of its referent (Yelland, Pollard, & Mercuri, 1993), and can better produce or comprehend sentences where one word is substituted with another (for a review, see Bialystok, 2001). Still, bilingual prereaders are not better than monolinguals at counting words in a text (Ricciardelli, 1992) or at word segmentation tasks and word judgement tasks (Nicoladis & Genesee, 1996), and word counting ability is not affected by bilingualism but only by literacy (Edwards & Christophersen, 1988). Advantages of bilingualism were only found in literate children: Bialystok (1986) compared literate English monolingual and French-English bilingual children, and found that the French-English bilinguals who had only learnt to read French segmented English as well as the English children, and outperformed them in the segmentation of bimorphemic compound words. Along the same lines, Hsia (1992) compared the English word segmentation skills of Chinese-English bilingual children and American children. The bilinguals performed worse than the monolinguals, except one group of bilinguals who were attending primary school and outperformed the English native speakers. It appears that only literacy (and not bilingualism) can

develop this specific aspect of metalinguistic awareness, but once it is developed then bilinguals have an advantage. Indeed, adult bilinguals who are literate in a word-spaced writing system use this as a guideline for determining word boundaries in unwritten second languages, as in the case of writing system developers devising writing systems for previously unwritten languages (Van Dyken & Kutsch Lojenga, 1993).

One study (Bassetti, 2004, 2005) looked at the effects of bilingualism and biliteracy on Chinese words awareness. The study compared Chinese word awareness in monolingual Chinese native speakers and English-speaking users of Chinese as a Second Language using two-word segmentation tasks. Since the English CSL users were native users of a word-spaced writing system, it was predicted that they would show higher word awareness levels than Chinese speakers. Results revealed that English CSL users segmented shorter words and reached higher levels of intragroup agreement on word segmentations. This was due to various differences between groups:

1) The segmentation of function words: Chinese adults often attached function words to content words, treating them as affixes, whereas English CSL users mostly considered function words as words;

2) The segmentation of nominal compounds: Chinese adults often considered nominal compounds containing 3 or more morphemes as single words, whereas English CSL users generally segmented them into two words, mostly corresponding to the number of words in the English translation of the compound;

3) Word segmentation strategies: English CSL users mostly relied on English translation to determine Chinese word boundaries, whereas Chinese adults used a higher number and variety of strategies, based on syntax, semantics and prosody.

The researcher concluded that English CSL users are affected by characteristics of both their first and second language writing systems. Because of the effects of the English writing system, they consider function words as words, segment nominal compounds, and reach higher levels of intragroup agreement than the Chinese group. At the same time, because of characteristics of the Chinese writing system, they find it difficult to segment Chinese into words and do not reach high levels of intragroup agreement. On the other hand, Chinese adults behave under many respects like English preliterates: they do not consider function words as words, they do not segment compounds and phrases in smaller words, they often lack self-consistency in segmenting the same word twice, and they rely on prosody to identify word boundaries. These are all characteristics of preliterates'

segmentations and also appear in Old English manuscripts. It appears that Chinese word segmentation is difficult for all Chinese language users, but those Chinese language users who are bilingual and literate in a word-spaced writing system perform more in line with English speakers segmenting English.

Rationale

Previous research shows that English users of Chinese as a Second Language (CSL) perform word segmentation tasks differently from Chinese monolinguals (Bassetti, 2004, 2005). If the difference between these two groups' word segmentation performance is due to the English CSL users' literacy in a word-spaced writing system, then a group of CSL users with an unspaced L1 writing system should perform in the same way as Chinese monolinguals. If, on the other hand, the English CSL users' performance is due to bilingualism, biliteracy or second language instruction, then all CSL users should perform in line with English CSL users, regardless of whether their other writing system marks word boundaries or not. Japanese CSL users are an ideal group to test the effects of L1 writing system on L2 word awareness, because of the way written Japanese marks boundaries. In written Japanese there is no spacing and word boundaries are not marked, so Japanese CSL users should segment words like Chinese native speakers. On the other hand, script alternation in written Japanese may represent other boundaries. Noun phrases and compounds words of Chinese origin (e.g., 'People's Republic of China' or 'seventeenth century') are written as strings of kanji preceded and followed by kana, and in general lexical morphemes are written in kanji and grammatical morphemes in kana. For this reason, if L1 orthographic conventions affect L2 word awareness, Japanese CSL users should consider noun phrases as single words (as Chinese native speakers do, and unlike English CSL users), and should segment grammatical morphemes as single words (as English CSL users do).

In order to ascertain the role of orthographic background on word awareness, this study compared three groups: Chinese native speakers; English CSL learners, whose L1 writing system uses interword spacing; and Japanese CSL learners, whose L1 writing system does not mark word boundaries. Their word awareness was tested by means of two written word segmentation tasks. On the basis of the evidence reported above, it was hypothesised that:

1) Orthographic background will affect word length, with English users of Chinese as a Second Language (CSL) segmenting shorter words, and Japanese CSL users and Chinese monolinguals segmenting longer words. Japanese and Chinese groups may not differ, because their writing

systems do not mark interword boundaries, or they may differ because of the Japanese participants' bilingualism.

2) Orthographic background will affect agreement on word segmentations, with English CSL users reaching higher levels of intra-group agreement on word boundaries than either Japanese CSL users or Chinese native users.

Method

Participants

As indicated above, three groups with different orthographic background were compared: Chinese monolinguals, English-speaking users of Chinese as a Second Language (CSL), and Japanese CSL users. The Chinese and English-speaking groups were randomly-selected subsets of the larger groups tested in Bassetti (2004, 2005).

Twenty-five monolingual Chinese speakers were recruited in China (mean age = 24; M = 11, F = 24). They were native speakers of the standard variety of Chinese (*putonghua*), were literate in both the Chinese writing system and its romanization system (*pinyin*), and most of them (67%) also spoke another Chinese language. While finding Chinese participants in that age range with no knowledge of English is impossible, these participants knew very little English. Most of them rated their knowledge as extremely poor or very poor on a 5-point scale, and the group scored on average 9% on a standardized English vocabulary test (Schmitt, Schmitt, & Clapham, 2001). Nobody knew other non-Chinese languages. All participants had a high school diploma.

Twenty-five English-speaking users of Chinese as a Second Language (CSL) were recruited in Great Britain (mean age = 22, M = 9, F = 16). They were native speakers of English, and most of them (92%) knew at least another language; 20% of them knew another language that does not mark word boundaries (Japanese, Korean or Cantonese). They were enrolled in third- or fourth-year university Chinese language courses and had learnt *pinyin*. All but one rated their Chinese as 'good' or 'proficient'.

Twenty-five Japanese users of Chinese as a Second Language were recruited in Japan and China (mean age = 28; M = 10, F = 11; 4 participants did not provide biographical information). They were native speakers of Japanese, mostly rated their English proficiency as 'poor' or 'very poor' (68%, 'average' = 26%, one participant rated herself as 'good'), and knew no other foreign languages (one participant knew Spanish). They had studied Chinese on average for 1.5 years (excluding one participant who had studied

it for 20 years), they had learnt pinyin, and mostly rated their Chinese proficiency as 'poor' (74%, 'average' = 26%).

Materials and Procedure

Participants received a questionnaire containing instructions, a Sentence Segmentation Task, a Text Segmentation Task, and questions about demographic and linguistic information. The Sentence Segmentation Task consisted of nine 7-hanzi Chinese sentences taken from Hoosain (1992). Sentences had simple structures and high-frequency hanzi. The Text Segmentation Task consisted of two short descriptive passages on neutral topics, already used in Bassetti (2005). The text was 342-hanzi long, and 99% of hanzi belonged to the 'most frequent' or 'frequent' category according to a hanzi frequency list (SJDHB, 1988). English and Japanese participants also received a list of hanzi with glosses in their first language to help them understand the Chinese text. Participants were instructed to draw a square around each string of hanzi that constituted a word.

Results

The analysis of word length revealed significant effects of orthographic background. In line with the first hypothesis, the English group segmented shorter words than either the Japanese or Chinese groups. The mean word length was calculated for each participant as the mean number of hanzi per word in the Text Segmentation Task. The English CSL users segmented the shortest words, with a mean word length of 1.78 hanzi ($SD = .12$), followed by the Japanese CSL users ($M = 2.40$, $SD = .94$) and the Chinese monolingual group ($M = 2.52$, $SD = .58$). A one-way between-groups Analysis of Variance revealed that orthographic background had a significant effect on word length, $F_{2, 72} = 9.53$, $p < .001$, $\omega^2 = .25$. Bonferroni post-hoc comparisons showed that word length was significantly different between the Chinese monolinguals group and the English CSL users group ($p < .001$), and between the English CSL users and the Japanese CSL users ($p < .005$). Chinese monolinguals and Japanese CSL users did not differ significantly.

The results of the Sentence Segmentation Task revealed that orthographic background significantly affected intragroups agreement rates, with the English-speaking group reaching the highest agreement rate, followed by the Japanese group and then by the Chinese group. For each group an intragroup agreement rate was calculated for each sentence in the Sentence Segmentation Task using an index of commonality, which expresses the

frequency of agreements as a proportion of the total number of comparisons, and yields a figure ranging from 0 ('complete disagreement') to 1 ('perfect agreement'). The mean agreement rates were .26 (SD = .07) for the Chinese monolingual group, .43 (SD = .20) for the Japanese CSL group and .70 (SD = .16) for the English CSL group. The levels of agreement are relatively low because the index represents the mean level of agreement calculated for each whole-sentence segmentation that was proposed by at least one participant in the group. The intragroup agreement rates for each sentence were entered as cases in a one-way repeated-measures Analysis of Variance with orthographic background as independent variable. Results showed that orthographic background significantly affects intragroup agreement, $F_{2, 16}$ = 29.82, p < .001, ω^2 = .14. Bonferroni post-hoc comparisons revealed that the Chinese monolingual group had a significantly lower intragroup agreement rate than either the English group or the Japanese group (both p < .001, r = .88 and r = .44 respectively), and the Japanese group had a lower agreement rate than the English group (p < .05, r = .83).

The difference in the mean word length between the English and Japanese group was mostly due to their different treatment of nominal compounds. The text contained various nominal compounds (e.g., 创建时期, 'foundation period'). Nominal compounds composed of 4 hanzi were the most frequent ones in the text (with 16 occurrences), and they all translated as two English orthographic words (except 中国研究, which could translate as either 'sinology' or 'Chinese research'). An analysis revealed that 4-hanzi compounds were marked as one single word by the majority of the Chinese monolinguals (M = 60%, ranging from 40% to 80% across different compounds), by one third of the Japanese participants (M = 33%, ranging from 20% to 52%), and by only 5% of the English participants (ranging from 0% to 12%). The number of participants who considered each of the 16 nominal compounds as one word was entered into a one-way repeated-measures ANOVA with orthographic background as the independent variable. Results revealed a significant effect, $F_{2, 30}$ = 135.76, p < .001, ω^2 = .84. Bonferroni post-hoc tests revealed that all groups differed significantly from each other (for all comparisons, p < .001). The Chinese group considered nominals as single words more often than the Japanese group (r = .82) and the English group (r = .95), and the Japanese group considered nominals as single words more often than the English group (r = .81).

Discussion

The results show that users of different first language writing systems produce different word segmentations of the same language, producing words of different lengths and with different levels of intragroup agreement. Their

performance was in line with the hypotheses. The English group segmented the shortest words with the highest agreement levels, and the Chinese group segmented the longest words with the lowest agreement levels. The Japanese group produced the same word length as the Chinese group, with levels of agreement higher than the Chinese group but lower than the English group. It appears that the same language is analysed differently depending on the concept of word of different groups of users, and that users of different first language writing systems produce different linguistic analyses of the same language.

The results are due to an interaction between characteristics of the Chinese language and writing system and the word awareness of the different groups of Chinese language users. On the one hand, the Chinese language and writing system play an important role in determining the Chinese word segmentation performance of all groups. Chinese is difficult to segment into words, and previous research on native speakers' word segmentation consistently revealed low levels of intragroup agreement. The groups in this study also reached relatively low levels of intragroup agreement, which is particularly striking coming from groups of just 25 participants, and especially considering that English speakers reach almost ceiling levels when segmenting texts in their first language. It appears that L2 users of Chinese cannot simply apply their L1 word awareness to the segmentation of Chinese texts in the way French-English bilingual children do (Bialystok, 1986). On the other hand, participants' L1 word awareness also plays an important role. If the characteristics of the Chinese language and writing system were determining, all groups should perform in the same way. The significant differences across groups show that the main factor in word segmentation of a text are not the characteristics of the language to be segmented, but the concept of word of the person performing the segmentation. This discussion will therefore aim at isolating factors that may determine differences in word segmentation performance across the three groups, looking first at the effects of the L1 writing system, then at the effects of bilingualism and biliteracy, and finally at the effects of the first language.

Effects of Orthographic Background

This study confirms that the main factor in the development of word awareness is literacy in a word-spaced writing system. Previous research had already revealed that those who know a word-spaced writing system segment shorter words and reach higher levels of agreement in Chinese, whether they are Chinese speakers who are literate in the word-spaced romanization system (Tsai et al., 1998) or CSL users whose L1 writing system marks word boundaries with spacing (Bassetti, 2004, 2005). The present study shows that

users of a non-word-spaced L1 writing system mark longer words than users of a word-spaced L1 writing system, mostly because they consider complex nominals as single words. This is in line with findings from illiterate adults who consider noun phrases as single words (Scholes, 1993a). Japanese CSL users segmented nominal compounds as single words significantly more often than the other bilingual group because they are not literate in a word-spaced writing system. Japanese CSL users' segmentations of nominal compounds is probably a consequence of another characteristic of their L1 writing system. Although the Japanese writing system does not use spacing to separate orthographic units, it creates orthographic units by alternating its various scripts. A typical Japanese text contains sequences of kanji (morphemic graphemes), *hiragana* (syllabic graphemes used for native words) and *katakana* (syllabic graphemes used for loans from languages other than Chinese). Nominal compounds tend to be written as sequences of kanji, preceded and followed by hiragana or punctuation marks, as in 日本語能力試験, 'Japanese proficiency test'. If compounds are of non-Chinese foreign origin, they appear as sequences of katakana graphemes, sometimes separated by dots in correspondence with the interword spaces in the original compound (e.g., オン・ザ・ロック, 'on the rocks'). Sometimes kanji and katakana compounds are used, preceded and followed by hiragana or punctuation, e.g. 漢字リスト 'kanji list'. Since in written Japanese nominal compounds are mostly represented as orthographic units, this could also contribute to explain why Japanese CSL users tend to consider nominal compounds as single words more often than English CSL users. The difference in word segmentation performance between the two bilingual groups therefore appears to be a consequence of their orthographic backgrounds. In particular, differences in word segmentation can be attributed to the presence of interword spacing in English and its absence in Japanese, and also possibly to the different treatment of compound nominals, which are written with interword spacing in English and as single orthographic units in Japanese.

Effects of Bilingualism and Biliteracy

Although in terms of word length the Japanese group performed differently from the English group and in line with the Chinese group, their level of intragroup agreement was significantly higher than the Chinese group, albeit lower than the English group. Since the Japanese CSL users were literate in two writing systems that do not mark word boundaries, this difference cannot be attributed to their L1 writing system. A possible explanation is that, unlike the Chinese group, the Japanese group was bilingual, biliterate and had undergone second language instruction.

Previous research has repeatedly shown that bilingualism does not facilitate the development of word awareness (word counting, word segmentation) in children (Edwards & Christophersen, 1988; Nicoladis & Genesee, 1996; Ricciardelli, 1992). This study mostly confirms these findings, but not entirely. If bilingualism and biliteracy resulted in heightened word awareness regardless of the languages and writing systems involved, the English and Japanese group should have behaved the same, in terms of both word length and intragroup agreement. Instead, the large differences between the two bilingual groups confirm that bilingualism does not have a determining impact on word awareness. On the other hand, the Japanese group reached significantly higher levels of intragroup agreement than the Chinese monolingual group. This means that bilingualism, although it is not the main factor, at least has an effect. Faced with the same Chinese text as monolingual native speakers of Chinese, the Japanese users of L2 Chinese analyse it differently. This could be due to their bilingualism, biliteracy and/or or to second language instruction. It appears that, while the main factor in Chinese word awareness is the orthographic background, users of Chinese who know another language and writing system perform differently from Chinese monolinguals.

The next question is whether differences in word segmentation performance could be partly due to participants' first languages, as well as to their first language writing systems. Interestingly, the pattern of results seems not to reflect characteristics of languages. From the viewpoint of traditional morphological typology, Chinese and Japanese are further away on the synthetic and fusional continua, with English in between. If Chinese word segmentation was affected by morphological characteristics of languages, English CSL users should behave more like Chinese native speakers, and Japanese CSL users should differ more from Chinese native speakers. The results of this study do not confirm these predictions. These results are therefore better explained as consequences of orthographic rather than linguistic continua. The English writing system segments texts the most, with spacing after each orthographic word; the Chinese writing system segments texts the least, with no markers for either word or phrase boundaries. Japanese is in between, because it has no word boundary markers, but script alternation creates sequences of same-script graphemes that are longer than English orthographic words but shorter than the units delimited by punctuation.

A final question is whether the English CSL users' performance is to be considered 'better' than the Chinese and Japanese participants' performance. It is not possible to say which segmentation is correct, because not even linguists can agree on word boundaries in Chinese. If levels of intragroup agreement are considered evidence of higher word awareness, then the English participants outperformed the other two groups. On the other hand,

the English participants' reliance on their first language to segment Chinese may have negative consequences. In particular, English CSL learners who expect Chinese to be composed of words may have more difficulty in dealing with novel hanzi combinations. Such novel combinations, which are often used by Chinese writers, may be considered by CSL users as words 'that are not words at all' (as in Hannas, 1997, p. 180). As Bassetti (2004) showed, Chinese language textbooks often add to the confusion and reinforce learners' view that Chinese is made of words. Furthermore, some English-speaking CSL users strongly believe that Chinese should be written with interword spacing and that Chinese speakers should become aware of words in the way English speakers are (Bassetti, 2004). In conclusion, it is not possible to say whether one group of participants in the present study performed 'better', but it can be argued that the effects of English orthography on English CSL users may not be limited to their metalinguistic awareness but may extend to their actual use of the language.

Conclusion

Chinese, English and Japanese users of Chinese analysed the same materials in the same language and writing system in a different way. These differences can only be attributed to their word awareness, although the low levels of intragroup agreement in all groups is a consequence of characteristics of the Chinese language and writing system. It appears that language users analyse the same piece of text differently, depending on their metalinguistic awareness. It is argued that the differences in Chinese word awareness of the Chinese, English and Japanese groups are mainly a consequence of their literacy experiences. The lack of interword spacing in Chinese and Japanese and its presence in English seems to be the main cause for differences in Chinese word segmentation. Other factors include bilingualism (including biliteracy and second language instruction), as well as specific characteristics of each writing system, such as whether compounds are written as one or two orthographic units. Future research could look at other bilingual groups with different L1 writing systems, for instance German CSL users, whose L1 writing system has interword spacing but treats nominal compounds as single orthographic words, or Arabic CSL users, whose L1 writing system has interword spacing but compounds postpositions and nouns in single orthographic words. Such research would further confirm that different groups of users look at the same language through the lens of the orthographic conventions of their L1 writing system.

References

Andrieux-Reix, N., & Monsonego, S. (Eds.) (1998). Segments graphiques du français: Pratiques et normalisations dans l'histoire. Special issue of *Langue française*, no 119.

Aronoff, M. (1992). Segmentalism in linguistics: The alphabetic basis of phonological theory. In P. Downing, S. D. Lima & M. Noonan (Eds.), *The Linguistics of Literacy* (pp. 71-82). Amsterdam: John Benjamins.

Barton, D. (1985). Awareness of language units in adults and children. In A. W. Ellis (Ed.), *Progress in the Psychology of Language* (Vol. 1, pp. 187-205). Hillsdale, N.J.: Lawrence Erlbaum.

Bassetti, B. (2004). *Second language reading and second language awareness in English-speaking learners of Chinese as a Foreign Language.* Unpublished PhD thesis, University of Essex, Colchester, UK.

Bassetti, B. (2005). Effects of writing systems on second language awareness: Word awareness in English learners of Chinese as a Foreign Language. In V. J. Cook & B. Bassetti (Eds.), *Second Language Writing Systems* (pp. 335-356). Clevedon, UK: Multilingual Matters.

Ben-Dror, I., Frost, R., & Bentin, S. (1995). Orthographic representation and phonemic segmentation in skilled readers: A cross-language comparison. *Psychological Science,* 6, 3, 176-181.

Berthoud-Papandropoulou, I. (1978). An experimental study of children's ideas about language. In A. Sinclair, R. J. Jarvella & W. J. M. Levelt (Eds.), *The Child's Conception of Language* (pp. 55-65). New York: Springer-Verlag.

Bialystok, E. (1986). Children's concept of word. *Journal of Psycholinguistic Research,* 15, 13-32.

Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. Cambridge: Cambridge University Press.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics,* 24, 27-44.

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language,* 22, 307-324.

Bugarski, R. (1993). Graphic relativity and linguistic constructs. In R. J. Scholes (Ed.), *Literacy and Language Analysis* (pp. 5-18). Hillsdale, N.J.: Lawrence Erlbaum.

Chao, Y. -R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chao, Y. -R. (1976a). The logical structure of Chinese words. In: Chao, Y. -R. (Ed.), *Aspects of Chinese Sociolinguistics* (pp. 260-284). Standford: Stanford University Press.

Chao, Y. -R. (1976b). Rhythm and structure in Chinese word conceptions. In: Y. -R. Chao (Ed.), *Aspects of Chinese Sociolinguistics* (pp. 275-292). Standford: Stanford University Press.

Cook, V. J. (1997). The consequences of bilingualism for cognitive processing. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 279-299). Mahwah, N.J.: Lawrence Erlbaum Associates.

Cook, V. J., & Bassetti, B. (2005). Introduction to researching Second Language Writing Systems. In V. J. Cook & B. Bassetti (Eds.), *Second Language Writing Systems* (pp. 1-67). Clevedon, UK: Multilingual Matters.

Derwing, B. L. (1992). Orthographic aspects of linguistic competence. In P. Downing, S. D. Lima & M. Noonan (Eds.), *The Linguistics of Literacy* (pp. 193-210). Amsterdam: John Benjamins.

Downing, J., & Leong, C. K. (1982). *Psychology of Reading*. New York: MacMillan.

Downing, J. A. (1970). Children's concepts of language in learning to read. *Educational Research,* 12, 106-112.

Edwards, D., & Christophersen, H. (1988). Bilingualism, literacy and meta-linguistic awareness in preschool children. *British Journal of Developmental Psychology,* 6, 235-244.

Ehri, L. C. (1975). Word consciousness in readers and prereaders. *Journal of Educational Psychology,* 67, 204-212.

Ferreiro, E. (1997). The word out of (conceptual) context. In C. Pontecorvo (Ed.), *Writing Development: An Interdisciplinary View* (pp. 47-60). Amsterdam: John Benjamins.

Ferreiro, E. (1999). Oral and written words. Are they the same units? In T. Nunes (Ed.), *Learning to Read: An Integrated View from Research and Practice* (pp. 65-76). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Francis, H. (1975). *Language in Childhood: Form and Function in Language Learning*. London: Elek.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Hannas, W. C. (1997). *Asia's Orthographic Dilemma*. Honolulu, HI: University of Hawai'i Press.

Holden, M. H., & McGinitie, W. H. (1972). Children's conceptions of word boundaries in speech and print. *Journal of Educational Psychology,* 63, 551-557.

Hoosain, R. (1992). Psychological reality of the word in Chinese. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language Processing in Chinese* (pp. 111-130). Amsterdam: Elsevier Science Publishers.

Hsia, S. (1992). Developmental knowledge of inter- and intraword boundaries: Evidence from American and Mandarin Chinese-speaking beginning readers. *Applied Psycholinguistics,* 13, 3, 341-372.

Illic, I., & Sanders, B. (1988). *The Alphabetization of the Popular Mind*. San Francisco, CA: North Point Press.

Kugelmass, S., & Lieblich, A. (1979). The impact of learning to read on directionality in perception: A further cross-cultural analysis. *Human Development,* 22, 406-415.

Leong, C. K. (1991). From phonemic awareness to phonological processing to language access in children developing reading proficiency. In D. J. Sawyer & B. J. Fox (Eds.), *Phonological Awareness in Reading: The Evolution of Current Perspectives* (pp. 217-254). New York: Springer-Verlag.

Li, W., Anderson, R. C., Nagy, W., & Zhang, H. (2002). Facets of metalinguistic awareness that contribute to Chinese literacy. In W. Li, J. S. Gaffney & J. L. Packard (Eds.), *Chinese Children's Reading Acquisition: Theoretical and Pedagogical Issues* (pp. 87-106). Dordrecht: Kluwer Academic Publishers.

Loizou, M., & Stuart, M. (2003). Phonological awareness in monolingual and bilingual English and Greek five–year–olds. *Journal of Research in Reading, 26,* 1, 3-18.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition,* 24, 65-92.

Meltzer, H. S., & Herse, R. (1969). The boundaries of written words as seen by first graders. *Journal of Reading Behavior,* 1, 3-14.

Miller, K., Frosch, C., Kelly, M. K., & Zhang, H. (2001). *What words look like: Preschool children's understanding of orthographic regularities in Chinese and English*. Unpublished manuscript.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition,* 24, 45-64.

Morikawa, K., & McBeath, M. (1992). Lateral motion bias associated with reading direction. *Vision Research,* 32, 6, 1137-1141.

Nagy, W. E., Kuo-Kealoha, A., Wu, X., Li, W., Anderson, R. C., & Chen, X. (2002). The role of morphological awareness in learning to read Chinese. In W. Li, J. S. Gaffney & J. L. Packard (Eds.), *Chinese Children's Reading Acquisition: Theoretical and Pedagogical Issues* (pp. 59-83). Dordrecht: Kluwer Academic Publishers.

Nicoladis, E., & Genesee, F. (1996). Word awareness in second language learners and bilingual children. *Language Awareness,* 5, 2, 80-90.

Ohala, J. J. (1992). The costs and benefits of phonological analysis. In P. Downing, S. D. Lima & M. Noonan (Eds.), *The Linguistics of Literacy* (pp. 211-238). Amsterdam: John Benjamins.

Olson, D. R. (1994). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge: Cambridge University Press.

Padakannaya, P. (2000). Is phonemic awareness an artifact of alphabetic literacy?! Poster presentation, ARMADILLO - 11, Texas, October 13-14.

Padakannaya, P., Rekha, D., Nigam, R., & Karanth, P. (1993). Phonological awareness, orthography, and literacy. In R. J. Scholes (Ed.), *Literacy and Language Analysis* (pp. 55-70). Hillsdale, N.J.: Lawrence Erlbaum.

Papandropoulou, I., & Sinclair, H. (1974). What is a word? Experimental study of children's ideas on grammar. *Human Development,* 17, 4, 241-258.

Parkes, M. B. (1993). *Pause and Effect: An Introduction to the History of Punctuation in the West*. Berkeley: University of California Press.

Perfetti, C. A., & Zhang, S. (1991). Phonological processes in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 17, 4, 633-643.

Ramachandra, V., & Karanth, P. (2007). The role of literacy in the conceptualization of words: Data from Kannada-speaking children and non-literate adults. *Reading and Writing*, 20, 173-199.

Read, C. A., Zhang, Y., Nie, H., & Ding, B. (1987). The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition,* 24, 31-44.

Ricciardelli, L. A. (1992). Bilingualism and cognitive development in relation to threshold theory. *Journal of Psycholinguistic Research,* 21, 4, 301-316.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing,* 18, 1, 55-81.

Scholes, R. J. (1993a). On the orthographic basis of morphology. In R. Scholes (Ed.), *Literacy and Language Analysis* (pp. 73-95). Hillsdale, N.J.: Erlbaum.

Scholes, R. J. (Ed.). (1993b). *Literacy and Language Analysis*. Hillsdale, N.J.: Lawrence Erlbaum.

SJDHB (Shanghai Jiaotong Daxue Hanzi Bianmazu, & Shanghai Hanyu Pinyin Wenzi Yanjiuzu) (Eds.) (1988). *Han zi xin xi zi dian* [A Dictionary of Chinese characters information]. Beijing: Kexue Chubanshe.

Sproat, R., Shih, C., Gale, W., & Chang, N. (1996). A stochastic finite-state word-segmentation algorihm for Chinese. *Computational Linguistics,* 22, 3, 377-404.

National Education Commission and National Language Commission (1996). *Han yu pin yin zheng ci fa ji ben gui ze* [Basic rules for Hanyu pinyin orthography]. Beijing: Zhongguo biaozhun chubanshe.

Tsai, C. -H., McConkie, G. W., & Zheng, X. J. (1998). Lexical parsing by Chinese readers. Paper presented at the Advanced Study Institute on Advances in Theoretical Issues and Cognitive Neuroscience Research of the Chinese Language, University of Hong Kong. Available at http://research.chtsai.org/papers/cht_asi.html. Last accessed: 20 June 2007, from www.geocities.com/hao510.

Tunmer, W. E., Bowey, J. A., & Grieve, R. (1983). The development of young children's awareness of the word as a unit of spoken language. *Journal of Psycholinguistic Research,* 12, 567-594.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology,* 23, 4, 515-557.

Vaid, J. (1995). Script directionality affects nonlinguistic performance: Evidence from Hindi and Urdu. In I. Taylor & D. R. Olson (Eds.), *Scripts and Literacy: Reading and Learning to Read Alphabets, Syllabaries and Characters* (pp. 295-310). Dordrecht: Kluwer Academic Press.

Van Dyken, J. R., & Kutsch Lojenga, C. (1993). Word boundaries: key factors in orthography development. In R. L. Hartell (Ed.), *Alphabets of Africa* (pp. 3-20). Dakar: UNESCO-Dakar Regional Office.

Ventura, P., Kolinsky, R., Brito-Mendes, C., & Morais, J. (2001). Mental representations of the syllable internal structure are influenced by orthography. *Language and Cognitive Processes,* 16, 4, 393 - 418.

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics,* 14, 4, 423-444.

Acknowledgments

Benedetta Bassetti is Lecturer in Applied Linguistics at Birkbeck, University of London. She holds a BA (Rome) in Oriental Languages, an MA (London) in Applied Linguistics and a PhD (Essex) in Applied Linguistics. Her research interests are second language writing systems and bilingualism and thought.