

CHAPTER 13: EFFECTS OF WRITING SYSTEMS ON SECOND LANGUAGE AWARENESS: WORD AWARENESS IN ENGLISH LEARNERS OF CHINESE AS A FOREIGN LANGUAGE

Benedetta Bassetto

Much research has shown that second language learners and users read and write their second language writing system differently from its native users, as a consequence of knowing another writing system. A relatively smaller amount of research shows that learners and users of a second language writing system (L2WS) also have a different knowledge of the linguistic units represented by their L2WS, compared with its native users. Native users of different writing systems are affected in their analysis of the spoken language by the linguistic units that their writing system represents as discrete units (by means of graphemes and orthographic conventions). When they learn a second language, they may encounter a L2 writing system that represents different linguistic units as discrete units. In that case, these multi-competent L2WS users may develop a different awareness of the linguistic units in their second language compared with native users of the target language because they know more than one writing system.

The present research shows that English learners of Chinese have different concepts of the Chinese word compared with Chinese natives, as a consequence of knowing both the English and Chinese writing systems. The word is the metalinguistic unit *par excellence* for English speakers, and their encounter with the Chinese writing system, that represents morphemes but not words as discrete units, may lead to a variety of reactions¹. The conflict between a L1 writing system that represents words and a L2 writing system that represents morphemes can be solved by relying on the L1WS to determine word boundaries in the L2, but the impact of the L2WS can affect various aspects of L2 awareness and use. The conflict can be solved in different ways by different L2 learners, ranging from the integration of the two views of language to the complete rejection of the new view of language conveyed by the L2 writing system.

1.1 The first language writing system and second language awareness

Writing systems represent the flow of spoken language as a sequence of distinct linguistic units with clear boundaries. For instance, while phonemes overlap in speech (Lively *et al.*, 1994), they are represented as discrete units (letters) in alphabetic writing systems. But not all writing systems represent the same linguistic units: while the graphemes of alphabetic writing systems represent phonemes, the graphemes of other writing systems represent consonants, syllables or morphemes.

Cross-orthographic research shows that writing systems affect the ability to identify and manipulate linguistic units in their users. In general, literate speakers tend to be aware of those linguistic units that are represented in their writing system. For instance, users of alphabetic writing systems are aware of phonemes, while users of syllabic writing systems are aware of syllables. Awareness of linguistic units is not related to literacy *per se*. Language users who are literate are still not aware of linguistic units that are not represented in their writing system, though present in their speech; for instance, Japanese children cannot perform some tasks that require awareness of phonemes, even though they are literate (Leong, 1991), because their writing system represents morphemes and morae but not phonemes; English adults can perform tasks that require awareness of words, which are represented as individuated units separated by spacing in their writing system, but not tasks requiring awareness of syllables or phrases, whose boundaries are not marked in their writing system (Miller *et*

al., in preparation). Writing systems affect awareness of linguistic units independently of characteristics of the language: this is obvious when comparing native speakers of the same language who are users of different writing systems. For instance, literate Chinese natives, whose writing system represents monosyllabic morphemes, cannot perform some phonemic awareness tasks which can be performed by Chinese natives who learnt *pinyin*, a transcription system based on the Roman alphabet (Read *et al.*, 1987); Kannada-speaking children, whose writing system is a semi-syllabary, cannot perform some phonemic awareness tasks which can be performed by blind Kannada children, who are users of an alphabetic *braille* (Prakash, 2000). This suggests the existence of *orthographic relativity*, whereby language users analyse language differently according to which units are represented in their writing system: phonemes for English speakers, morae for Japanese speakers, morphemes for Chinese, words for English, etc. (see Bugarski's 'graphic relativity', Bugarski, 1993).

If users of different writing systems are aware of different units, what are second language users aware of? Bilingualism helps children develop some aspects of phonological awareness (Bruck and Genesee, 1995), but this does not extend to awareness of phonemes, which is only learnt through exposure to a phonemic writing system. Bilingual children are no better than monolinguals at phoneme substitution or phoneme counting tasks (Bialystok, 2001; Bialystok *et al.*, 2003). But, if they learn to read their L1 writing system and become aware of the linguistic units it represents, they can use this awareness to analyse their L2, and perform differently from, or even better than, literate monolinguals; for instance, Hebrew users of English as a Second Language segment English words into phonemes differently from English natives (Ben-Dror *et al.*, 1995); literate English-Greek bilingual children outperform literate English monolinguals in some English phonemic awareness tasks (Loizou and Stuart, 2003). Bilingualism *per se* does not make L2 users more aware of linguistic units than monolinguals, but once they acquire awareness of a linguistic unit by exposure to one writing system, L2 users can apply this awareness to other languages.

1.2 Word awareness in English and Chinese natives

The present study deals with *word awareness*, that is to say the conscious knowledge of the word as a linguistic unit. Word awareness is demonstrated by the ability to understand and use the term 'word', to identify words in a written or spoken text and to distinguish them from other linguistic units, so that morphemes or phrases are not considered 'words'. According to orthographic relativism, word awareness should only develop in users of those writing systems that represent words as discrete units, and it should not be present in illiterates or in those literates whose writing system does not mark word boundaries.

English is one of the writing systems that mark word boundaries; it represents *orthographic words*, *i.e.* strings of letters preceded and followed by spacing (*interword spacing*). In line with orthographic relativism, literate English adults understand what a 'word' is and can distinguish it from other linguistic units. Using the most widespread test of word awareness, the *word segmentation task*, Miller *et al.* (in preparation) presented a group of English natives with a series of sentences written without interword spacing (such as <icecreamisthemostpopulardessertinsummer>), and asked them to segment the sentences by drawing a line between words. Answers were almost unanimous, showing that English adults understand the meaning of 'word' and can identify them. On the other hand, research repeatedly showed that English preliterate children are *not* word aware: they do not understand what a 'word' is (Downing, 1970),

they do not understand that the spacing between strings of letters in writing separates linguistic units (Meltzer and Herse, 1969), they cannot say whether phonemes, syllables or sentences are words or not (Downing and Oliver, 1974); and when asked to identify words in speech they identify phonemes, sentences or other linguistic units (Ferreiro, 1997). When they learn to read a word-spaced writing system, children then go through a stage where they can count written words but not spoken words (Ferreiro, 1999), and after about two years of literacy, they can consistently identify spoken words the same way as adults. Illiterate English adults also cannot identify words (for instance, they cannot identify the number of words in 'television', 'forever', 'four oxen' or 'the White House') and in general seem to think that dividing speech into words is 'meaningless' (Scholes, 1993).

Unlike the English writing system (but like Thai, Burmese, Tibetan, Japanese and other writing systems), Chinese does not mark word boundaries. Spacing is used to separate Chinese graphemes, the *hanzi* or *zi* (汉字, /xan tsi/). Again confirming orthographic relativism, Chinese natives (both children and adults) are not aware of words; indeed Chinese did not have a term for 'word' until the concept was imported from the West at the beginning of the 20th century (Packard, 1998). When performing a word segmentation task, Chinese natives segment the same text into words differently from each other, are inconsistent with their own previous segmentations, identify whole phrases as words and sometimes do not understand instructions asking them to identify 'words' (Hoosain, 1992; King, 1983; Miller, 2002; Sproat *et al.*, 1996). Interestingly, Chinese natives who learnt pinyin (the Chinese romanization system, which uses interword spacing) segment Chinese texts differently from Chinese natives who only know hanzi (Tsai *et al.*, 1998). This shows that exposure to a word-spaced writing system affects word awareness even among native speakers of the same language. In an interesting cross-linguistic experiment, Miller and his colleagues (in preparation) compared Chinese and English natives' segmentations of the same sentences, presented in Chinese and English respectively. They found that while English natives reached an almost 100% agreement on their word segmentations, Chinese natives had a significantly lower agreement rate (Miller *et al.*, in preparation).

While it appears that users of non-word-spaced writing systems are generally not aware of words, literate Chinese natives might represent a special case because of characteristics of their writing system. Each Chinese grapheme (*hanzi*) represents one morpheme and its corresponding spoken syllable (with very few exceptions). For instance, 爱 represents the morpheme 'to love' and the corresponding syllable /ai/ (in the standard variety of Chinese). Chinese lexical items can be mono- or polymorphemic; in written Chinese they are correspondingly mono- or multi-*hanzi*. For instance: /ai/ (爱, 'to love'); /ai zən/ (爱人, 'spouse'), *etc.* In this way, the writing system assigns one specific written form to each morpheme: while the syllable /ji/ can be written with various *hanzi* (以, 已, 乙, 倚, 蚁, *etc.*), the writing system indicates that the /ji/ in /ji wən/ ('before') is written with the same *hanzi* as the /ji/ in /ji tɛiŋ/ ('already') but not as the /ji/ in /ji ʂən/ ('above-mentioned'). This means that the spoken /ji wən/ can be analysed as the two morphemes 'already-past' and /ji tɛiŋ/ as 'already-pass through', but /ji ʂən/ ('above-mentioned') is not 'already-above', but 'at-above'. The same *hanzi* also often represents more than one morpheme, so that it may represent some that are lexical items and some that are not; when reading, the context of the sentence determines whether a *hanzi* represents a lexical item or a component of a

polymorphemic lexical item. So 生 represents a verb in 她生孩子了 ('she gave birth to a baby'), the second morpheme in 陌生人 ('foreigner'), the third morpheme in 研究生 ('researcher'), and so on. This gives the false impression that 生 represents a lexical item, when in fact it is the written representation of different homophonic morphemes. It is by now clear that the hanzi plays a central role in the Chinese writing system and that its importance and versatility conceal the role of the lexical item.

Going back to language awareness, since their graphemes represent monosyllabic morphemes, Chinese natives can segment language in syllables (Miller *et al.*, in preparation) and, for each syllable, identify the correct hanzi among the many homophonic hanzi that could represent it. For instance, they can say that /ai tɛ^hiŋ/ ('love') contains two syllables, and that the first one is written as 爱 rather than 艾, 碍, 隘, 暖 or other homophonic hanzi. The ability to identify syllables with the corresponding morpheme/hanzi is an important aspect of language awareness for Chinese children acquiring literacy (Li *et al.*, 2002), which illiterates do not have (Chao, 1976). The hanzi is recurrent in Chinese linguistic activities: text length is calculated in hanzi, dictionaries are searched by hanzi, *etc.* (Chao, 1968). Given the importance of the hanzi in their writing system, not surprisingly most Chinese natives think of their language as made of hanzi (T'ung and Pollard, 1982; Hannas, 1997). The status of the hanzi as the metalinguistic unit for Chinese natives is just as salient as the status of the 'word' as the metalinguistic unit for English natives. Hanzi does not mean 'a group of strokes inscribed inside a square', just like the English 'word' does not mean 'a series of letters comprised within two spaces'; hanzi are the linguistic units that everybody is aware of, recognizes and uses to talk about language. And the central role of the hanzi obfuscates the role of the lexical item for Chinese natives, in probably the same way as the central role of the word obfuscates the role of the morpheme for English natives. In this, English and Chinese natives do not differ: they are all aware of the linguistic units that are represented in their writing system.

1.3 Word awareness in L2 users

The evidence reviewed above supports the view that word awareness only develops with literacy in a word-spaced writing system. But does this also apply to L2 users? There is evidence that bilingualism facilitates the development of some aspects of language awareness (Cook, 1997). Are L2 users aware of words in the absence of literacy? The answer seems to be negative: bilingual prereaders are not better than monolinguals at counting words in a text (Ricciardelli, 1992) or at word segmentation and word judgement tasks (Nicoladis and Genesee, 1996); word counting in bilingual children is positively affected only by their literacy, not by their bilingualism (Edwards and Christophersen, 1988). For instance, although preliterate American children performed better than Chinese-English bilingual children in English word segmentation, a group of Chinese-English bilingual children learning to read English in the first year of primary school outperformed the American native speaker children who could not read (Hsia, 1992). Word awareness acquired through exposure to a writing system can be used to analyse another language: French-English bilingual children who are literate only in French can segment English texts in words as well as literate English children (and even perform better in the segmentation of bimorphemic compound words such as 'snowman') (Bialystok, 1986). It is clear that the ability to segment a text into words, or to decide whether something is a word, only develops

with literacy in a word-spaced writing system; but once word awareness is acquired via one writing system, it can be used to analyse another language even in the absence of literacy in that language, and then bilinguals can even enjoy an advantage over literate monolinguals.

2. The present study

Word awareness develops in English natives as a consequence of learning to read English, and does *not* develop in literate Chinese natives. Since English represents orthographic words and Chinese does not, literate English speakers might use their word awareness to analyse L2 Chinese. Their concept of the Chinese word could therefore be different from that of native Chinese speakers. Given that, once word awareness is acquired, it can be used to analyse a second language, do English learners of Chinese as a Foreign Language (CFL) apply their word awareness to identifying words in Chinese? Do they have a different concept of the Chinese word compared with Chinese natives?

In order to investigate this question, a Chinese word segmentation task was given to a group of English CFL learners and a group of Chinese natives. On the basis of previous findings, two hypotheses were proposed: 1) English CFL learners will mark shorter words; and 2) English CFL learners will show higher levels of intragroup agreement on their word segmentations. The first hypothesis was proposed because previous research had shown that Chinese natives who learnt the pinyin romanization system (which represents orthographic words) marked more word boundaries (*i.e.* shorter words) than Chinese natives who did not learn it (Tsai *et al.*, 1998); it was reasoned that English CFL learners' prolonged exposure to the English writing system should have even stronger effects than a limited exposure to pinyin and should result in shorter words than those marked by pinyin-literate Chinese natives. The second hypothesis was proposed because previous research had shown that English natives segmenting English words reach almost 100% agreement, but Chinese natives segmenting Chinese have much lower levels of agreement (Miller *et al.*, in preparation); it was reasoned that if L1 word awareness can be used to analyse the second language, English CFL learners who reach such high levels of agreement in their first language should reach higher levels of intragroup agreement on Chinese segmentation than Chinese natives.

The two hypotheses were tested by means of two word segmentation tasks (a text and a sentence segmentation task respectively), whereby participants were asked to segment the materials into words. For both tasks, a one-factor between-subjects quasi-experimental design was used to test the effects of the first language writing system (English and Chinese) on average word length and on intragroup agreement rates.

2.1 Participants

60 English-speaking learners of Chinese as a Foreign Language (CFL) were recruited at various British universities. They were users of English as an L1 and as an L1 writing system, enrolled in third- or fourth-year Chinese language courses. 90% of respondents rated their own Chinese reading skills as good or proficient.

The 60 Chinese natives were native users of the standard variety of Chinese and of the Chinese writing system. They were matched to the English group in terms of educational background, had all learnt pinyin (the Chinese romanisation system) in school and most of them knew at least one additional Chinese language besides

Standard Chinese, as is the norm in the People's Republic of China. Since knowledge of English could affect the results, they were given an English vocabulary test (Schmitt *et al.*, 2001) to check that their knowledge of English was non-existent or minimal.

2.2 Materials and procedure

Participants were given a set of printed materials containing two texts (for the text segmentation task) and 9 sentences (for the sentence segmentation task). The written instructions invited them to draw a square around each word in the text, and a final questionnaire included demographic information. The Chinese texts were two short descriptive passages taken from a Chinese encyclopaedic dictionary (Cihai Bianju Weiyuanhui, 1989). In total they were 342-hanzi long and contained 300 valid word boundaries (hanzi not followed by a punctuation mark). The 9 sentences were taken from a previous study (Hoosain, 1992) and consisted of 7 hanzi (with 6 valid word boundaries) each. Materials were judged by a native Chinese language teacher as appropriate for the target L2 learners. The hanzi in the text were highly frequent: 99% belonged to the 'frequent' category in a hanzi frequency dictionary (Shanghai Jiaotong Daxue, 1988). In the final questionnaire, 95% of the English respondents reported that the Chinese materials were not difficult.

3. Results

Results from the text segmentation task revealed that the effect of L1 Writing System was significant. The average word length, *i.e.* the average number of hanzi per word, was significantly different between the two groups, with English learners of Chinese as a Foreign Language showing a significantly shorter average word length compared with the Chinese natives ($M= 1.79$, $SD=.14$ and $M=2.78$, $SD=.81$ respectively, see Figure 1).

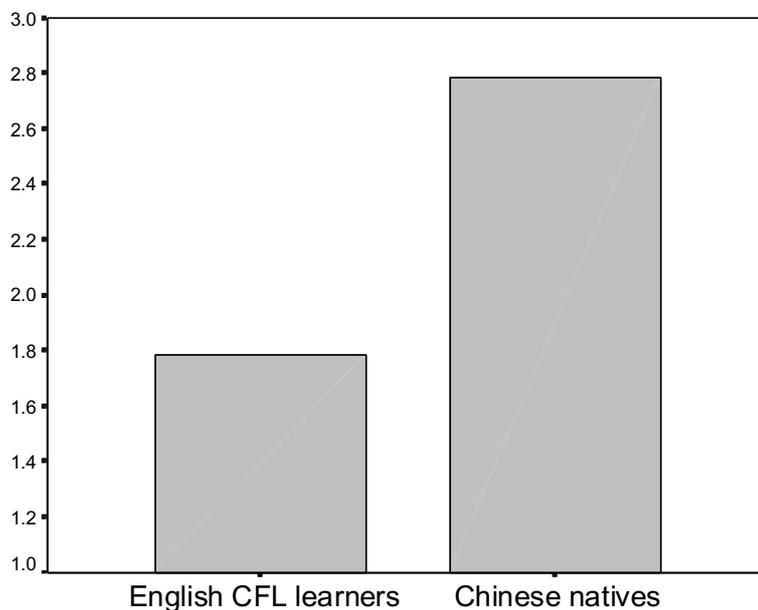


Figure 1 Average word length (in hanziⁱⁱ) by group

An independent group t-test revealed a significant difference between the two group means ($t_{1,118} = -9.397$, $p < .001$). In line with the first hypothesis, the English CFL learners segmented text into shorter words compared with Chinese natives.

Results from the sentence segmentations revealed a significant effect of L1 Writing System on intragroup agreement rates, with English learners of Chinese as a

Foreign Language showing a significantly higher agreement rate than Chinese natives. The agreement rate on each sentence was calculated by means of an Index of Commonality, which expresses the frequency of agreements as a proportion of the total number of comparisons as a figure ranging from 0 to 1. The average agreement rate for the English CFL group was .65 ($SD=.18$); for the Chinese group it was .24 ($SD=.05$), showing that English CFL learners agreed on significantly more sentence segmentations than Chinese natives. (As the Index of Commonality was based on agreement on each segmentation of the whole sentence, the levels of agreement are relatively low.)

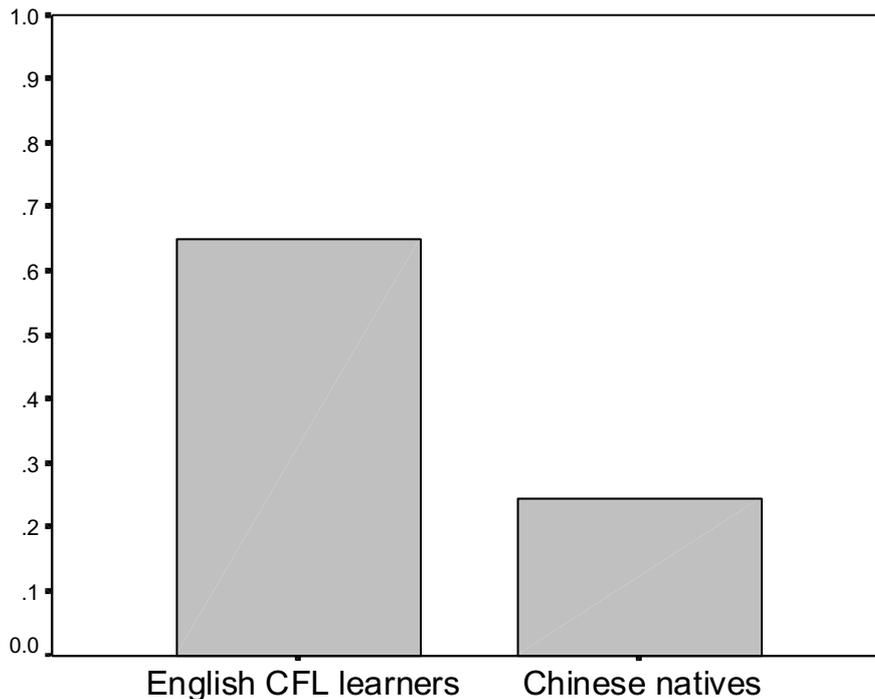


Figure 2 Intragroup agreement rates by group

A repeated measures t-test by item revealed that the difference between the two groups was significant ($t_{1,8} = 6.83, p < .001$), showing that, in line with the second hypothesis, the English CFL learners had a higher intragroup agreement rate than the Chinese natives.

3.1 Analysis of results

Since both groups were segmenting the same materials in the same language and writing system, differences can only be attributed to differences in word awareness, and not to differences between the languages and/or writing systems being segmented, as could be the case with cross-linguistic comparisons. These results show that English learners of Chinese have a different approach to Chinese word segmentation from Chinese natives, and agree more with each other's approach to identifying words. But on the other hand they are also affected by the Chinese language and writing system. The English group was far from the almost 100% agreement that English natives show when segmenting English materials. This is due to characteristics of the Chinese writing system, notably the lack of interword spacing and the important role of the morpheme/hanzi.

In line with the hypotheses, exposure to a first language writing system that marks word boundaries resulted in shorter words and higher agreement rates in the

segmentation of a second language, but it is not clear why this should be so. In order to understand why English CFL learners identify shorter Chinese words than Chinese natives, further analyses were performed on the linguistic units marked as words by both groups.

3.2 Reasons for different word lengths

Two main differences between the Chinese and the English groups seem to have led to differences in word lengths. English CFL learners mostly treated function words as words, while Chinese natives considered them as both words and affixes (affixed to the preceding, or sometimes following, content word). English CFL learners also segmented nominal compounds in smaller units, while Chinese natives considered them as single words. For instance, let us consider the following 7-hanzi phrase:

Chinese text:	十	七	世	纪	的	欧	洲
Transcription:	/ʃi	tɕʰi	ʃi	tɕi	tr	ou	tʃou/
Hanzi meaning:	Ten	seven	age	epoch	<i>de</i>	Europe	continent
English translation:	The Europe of the Seventeenth century						

This is how it was segmented by most English CFL learners (dots represent where participants drew word boundaries):

十七•世纪•的•欧洲 ('seventeenth century *de* Europe', 4 words).

Most Chinese natives segmented it as follows:

十七世纪的•欧洲 ('seventeenth-century-*de* Europe', 2 words)

or

十七世纪•的•欧洲 ('seventeenth-century *de* Europe', 3 words).

The results from the Chinese group are in line with previous findings that Chinese natives consider compounds and phrases as words and attach function words to content words (Hoosain, 1992; King, 1983). T-test comparisons were performed on the two groups' treatment of *de* (the most frequent function word in Chinese) as a word and on the treatment of 4-hanzi nominal compounds as words, and both differences were statistically significant (Figures 3 and 4). Obviously when nominal compounds are considered one word and function words are affixed to content words, the average word will be longer than when nominal compounds are segmented and function words considered words.

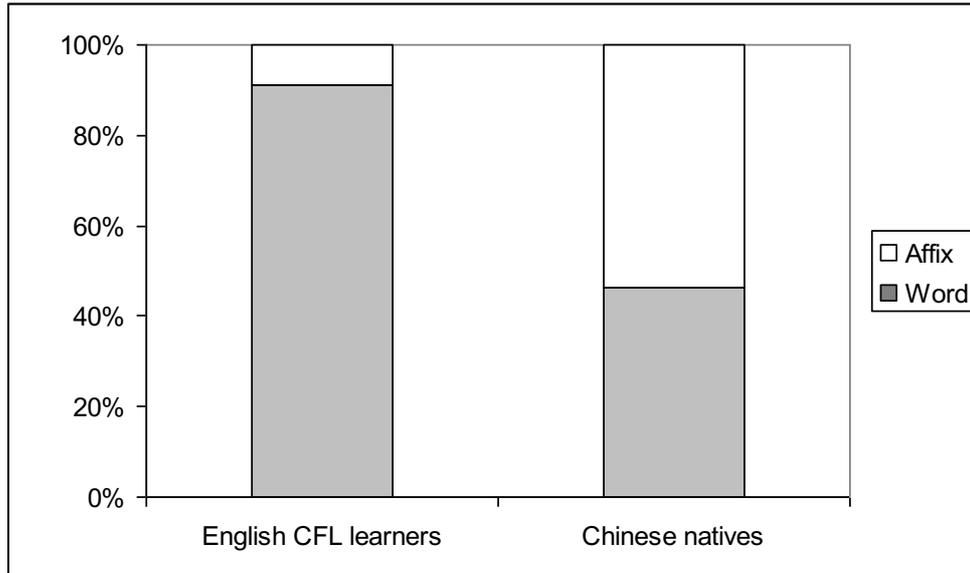


Figure 3 Segmentation of *de* by group

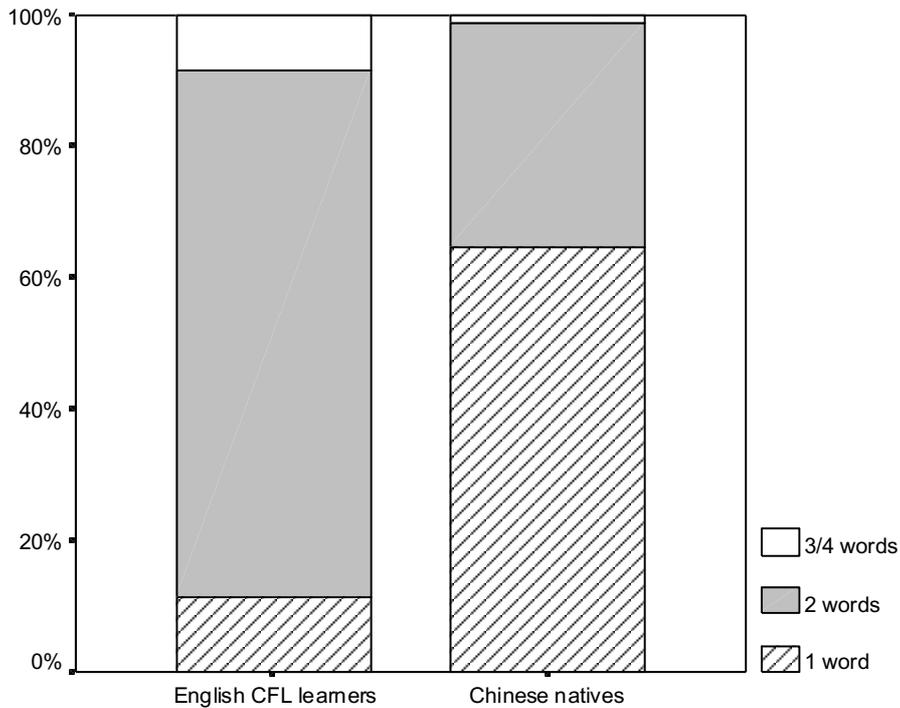


Figure 4 Segmentation of 4-hanzi nominal compounds by group

3.3 Reasons for different agreement rates

Going back to the short phrase presented above, the distinction between the two groups' segmentations was not as clear-cut as it looked above. At closer view, the segmentation patterns of the Chinese group were much more complex. While 83% of Chinese participants considered 欧洲 ('Europe') as one word, another 12% considered 十七世纪的欧洲 all as one word ('17th-century-*de*-Europe'), and the remaining 5% considered 的欧洲 (*de*-Europe) as one word. Regarding 十七世纪的 (*Seventeenth century de*), as many as 5 different segmentations were suggested:

十七世纪的
十七世纪•的
十七•世纪的
十七•世纪•的
十•七•世纪•的.

Obviously such a variety of segmentations on such a short string explains the high levels of disagreement on word segmentation in the Chinese group, in line with previous findings in the literature (Hoosain, 1992). With regards to the English CFL learners group, although 85% of English participants segmented the phrase as 十七•世纪•的•欧洲 ('Seventeenth century *de* Europe'), it is worth noting that another 10% considered 十七世纪 ('Seventeenth-century') as one word, in line with the Chinese natives' segmentation, and the remaining 5% segmented 'seventeen' in two words, 十 and 七 ('ten' and 'seven'). Compared with the English group, the Chinese group presented a wider variety of segmentation, with a lower percentage of participants agreeing on one main segmentation, but interestingly the English group also presented some minority segmentations. This explains the low intragroup agreement rate of the Chinese group, and the higher but still relatively low level of agreement in the English group.

Besides differing on the levels of intragroup agreement, the two groups also differed in the levels of self-consistency (intrajudge agreement). When the text contained the same lexical item twice, the English CFL learners tended to treat the same hanzi or hanzi strings in the same way throughout, but Chinese natives treated the same hanzi or hanzi strings differently in the same text. This characteristic of the Chinese natives' performance had already been noted in the literature with reference to the segmentation of *de* and of the negation *bu* (King, 1983). In the present research, this lack of self-consistency appeared both in the segmentation of *de* (which occurred 12 times in the text) and in the segmentation of nominal compounds (two compounds occurred twice each). The English CFL learners showed significantly higher self-consistency in the segmentation of both *de* and the repeated compounds, but they too were not 100% self-consistent.

But why do Chinese participants show lower levels of intragroup agreement and self-consistency in word segmentation? This is because the two groups' approaches to word segmentation are different. The Chinese use a higher number of word segmentation strategies and a wider variety of them compared with the English CFL learners.

3.4 Word segmentation strategies

In the final questionnaire, participants were asked to describe their word segmentation criteria. An analysis of the answers revealed quantitative and qualitative differences between the two groups. English CFL learners applied fewer and less varied segmentation criteria, and while some criteria were common to both groups, others were only mentioned by one group or the other.

The following are typical descriptions of how an English CFL learner segments Chinese text into words:

- 'Whether in English it is a word or not'
- 'Counted English equivalent as one word + Chinese grammatical particles as one word.'

Translation into English was the most frequently reported strategy in the English CFL group, reported by 36% of respondents, 47% of whom indicated it as their only strategy. This could partly explain the higher levels of intragroup agreement and self-consistency.

The Chinese group reported more varied and complex word segmentation strategies, which included various criteria, sometimes organised in a sequence as in the following example:

'I use the following stages: 1) I first segment the sentence in subject and predicate; 2) I then segment each part into the smallest units according to the word's meaning and word class, but at the same time I consider completeness of meaning, I don't simply segment according to word class, for instance: 学生生活 ['student-life'] and 锻炼身体 ['body-building'] [make one word]; 3) Finally, I rely on intuition, and the rhythm when I read it.' [all translations by the author]

Although not all the Chinese participants gave such elaborate answers, some of their answers contained as many as five different word segmentation criteria, while the overwhelming majority of English participants (81%) reported only one criterion. Using more than one criterion naturally leads to more varied segmentations.

Besides reporting different numbers of criteria, the two groups also reported using different criteria. For instance, as mentioned above, the most frequently used criterion in the English group was translation into English; this hardly ever occurred in the Chinese group (who had no or minimal knowledge of English). The most striking difference is the overwhelming use of syntactic strategies by the Chinese group (dividing subject, verb and object; dividing nouns, verbs, adverbs and adjectives; *etc.*) reported by 45% of respondents. While a couple of English CFL learners reported using 'grammar', they did not explain how they used it.

Arguably the most interesting difference is that the Chinese group reported the use of prosodic strategies for word segmentation. For instance, one of the Chinese respondents wrote:

'I segment according to the spoken intonation. [...] When we speak, there are always some pauses, and I use these pauses to segment.'

A variety of prosody-based criteria were reported, including pauses in speaking or in reading, intonation and rhythm. While such criteria were reported by 18% of Chinese respondents, the English group did not report them at all (only one respondent reported using the 'tempo of the text'). This prosodic approach could explain why *de* was not considered a word (phonologically it behaves as a clitic). It could also explain its apparently inconsistent treatment by the Chinese group, which could be determined by prosodic context, as proposed by King (1983).

Sometimes the two groups reported using similar criteria, but from different viewpoints. If all criteria are classified by type of strategy (semantic strategies, syntactic strategies, intuition, *etc.*), it appears that both groups mostly used semantic strategies, which include considering the meaning of each hanzi, considering words as units of meaning, analysing the sentence meaning, *etc.* But while English CFL learners mostly looked at the meaning of single hanzi or words, the majority of Chinese respondents were preoccupied with the meaning of longer units and stressed the importance of keeping units of meaning together within the same word. For instance, one Chinese participant wrote:

'I do a complete segmentation depending on the meaning of the whole sentence, I don't just mechanically segment into the smallest words. That way, one loses coherence and completeness of meaning.'

This probably explains why the Chinese group did not segment nominal compounds into smaller words.

Segmentation strategies can explain the differences in word length and intragroup agreement between the two groups, as well as differences in the two groups' self-consistency. When only one strategy is used, and it mostly consists of English translation (as is the case with English CFL learners), the same hanzi or hanzi string will be segmented in the same way by different participants and by the same participant on different occasion. When different participants apply different criteria, and each participant employs more than one criterion (so that different criteria can take priority in different contexts), this leads to the more varied segmentations seen in the Chinese group.

4. Discussion

It appears that English learners of Chinese as a Foreign Language have a different concept of the Chinese word compared with Chinese natives. The two groups do not only differ in how they identify words, they differ in their view of what constitutes a Chinese word: for most Chinese natives it is a syntactic unit, while for most English CFL learners it is the equivalent of an English word; for both groups a word is a unit of meaning that cannot be further segmented, but for English CFL learners this means a mono- or disyllabic unit, while for Chinese natives this includes longer compounds and phrases; for Chinese natives it is also a prosodic unit that can be identified by means of pauses and intonation units, a possibility that never occurs to English CFL learners.

But the Chinese word awareness of English CFL learners is not simply a consequence of cross-orthographic influence. They can use their L1 English word awareness to analyse L2 Chinese; but, unlike French-English bilingual children, who use their awareness of French words to segment English (Bialystok, 1986), CFL learners cannot simply 'transfer' their L1 word awareness because they are affected by characteristics of the Chinese language and writing system. English CFL learners obtain lower intragroup agreement rates and self-consistency when segmenting Chinese than they would obtain in segmenting English texts. This is due to the Chinese writing system: it is partly due to its lack of interword spacing, and partly to the centrality of the morpheme/hanzi that imposes itself on CFL learners as well. Those English CFL learners who considered 十七世纪 ('Seventeenth-century') as one word have developed a concept of word (or at least of the Chinese word) which is different from the concept of word in monolingual English natives and in line with the Chinese concept of an unbreakable unit of meaning; those who divided 十七 'seventeen' into two words ('ten' and 'seven') were influenced by the important role of the morpheme/hanzi; in both cases, their segmentations are not the results of simply translating into English. The percentage of English participants who at least occasionally showed such 'Chinese-style' segmentations testifies the impact of the Chinese language and writing system on their concept of the Chinese word. The effects of the Chinese writing system also surface in the descriptions of their word segmentation criteria: there is the difficulty of deciding what a word is in the absence of interword spacing ('Difficulty in deciding whether to split up names, esp. names of centres, e.g. 人口研究中心 [Population Research Centre] all one word? Or 3 separate ones?') and there is the centrality of the hanzi as a unit of meaning; one felt 'Each character is a word; however there are many two character phrases that are words'. And in fact a few CFL learners reported difficulty in deciding what constitutes a Chinese word: 'A difficult question to answer';

'I don't really know!'. Even the authority of reference tools becomes questionable when tools in the two languages differ: while one learner showed a typical reliance on the authority of dictionaries: 'If I know I can find it in the dictionary it must be a word', another noted: 'It's difficult because 中华人民共和国 [People's Republic of China] will probably appear as ONE word in the dictionary'.

4.1 The multi-competent L2 user and language awareness

These results support the theory of multi-competence, that is the knowledge of two or more languages in one mind (Cook, 1991). Literate L2 users not only have two or more languages in their minds, they also have two or more writing systems (see discussion in the introduction to this book). Their use and their knowledge of their languages and writing systems are different from the use and knowledge of native users of the target language and writing system, and are influenced by the two (or more) languages and the two (ore more) writing systems in the multi-competent L2 user's mind. In this way, the multi-competent English learner/user of Chinese as a Foreign Language has a different knowledge of the Chinese language from Chinese natives.

English CFL learners have a different concept of the Chinese word, compared with natives. But what is actually happening in the minds of these L2WS users? Do they have a different concept of the Chinese word coexisting with their concept of the English word, or is their overall concept of word changing? In addition to the results from this research, informal conversations with other CFL users revealed interesting cases: an Italian CFL user reported discovering the existence of morphemes when she started studying Chinese and then applying the same concept to her first language; an English CFL user, asked to translate some Chinese sentences into English, wrote all the English compounds corresponding to two hanzi in the Chinese text (such as 桌布, <tablecloth>) as two separate English orthographic words (<table cloth>) and wrote: 'Im not sure what constitutes a word really – for example table cloth – is that two words in English? Is it the same in Chinese?'. While this is of course anecdotal evidence, it is still interesting.

On the other hand, other CFL users think that words are self-evident and universal units of language analysis. They criticize the Chinese natives' view of language as made of hanzi and their habit of putting hanzi together to create new words that do not exist in dictionaries (e.g., Hannas, 1997). The negative view of the Chinese lack of word awareness also creeps into Chinese language textbooks, as in the following (co-authored by a CFL user): 'Most Chinese still think of their language as consisting of characters rather than words' (T'ung and Pollard, 1982: 2). Although the authors explain that this view facilitates Chinese reading, and proceed to teach both spoken words and hanzi, they do not explain why the Chinese think their language is made of hanzi, and the use of the word 'still' implies that this view is incorrect rather than different. And while some CFL users work to produce word-based Chinese reference tools (e.g. the ABC Dictionaries series: DeFrancis, 1996; or the word index to a hanzi dictionary, Mair, 2003), other CFL users comment that these word dictionaries might be more difficult to use than hanzi dictionaries (Light, 1998). And while some CFL users fight to get romanised library catalogues written in words rather than syllables, according to the 'rational aggregation of Sinitic syllables into words', others find syllable-based catalogues easier to search or to produce because they have 'absolutely no faith in [their own] ability to separate the words correctly' (see Chinese Kenyon archives, 2000).

It appears that there is much variability in how English-speaking CFL users react to the impact of a different writing system and related views of language. While reliance on L1 word awareness is their main approach to identifying L2 words, CFL learners show signs of developing a new concept of the Chinese word different from the concept of the English word, as a consequence of exposure to the Chinese writing system.

4.2 Orthographic relativism and Chinese word awareness

These research findings support orthographic relativism, the view that writing systems affect their users' views of language. Previous research showed that native speakers of the same language analyze their first language differently if they learnt to read it through the medium of different writing systems (Prakash, 2000; Read *et al.*, 1987); the present study shows that L1 and L2 users show different awareness of the same linguistic units in the same language if they were exposed to different first language writing systems.

But how many of these differences can be attributed to the participants' L1 writing system, rather than to bilingualism? Given the lack of orthographic conventions for word boundaries, both English and Chinese natives trying to segment Chinese find themselves in the same situation as preliterates or illiterates. But while the word segmentations of Chinese natives are reminiscent of the word segmentations of English preliterates, those of English CFL learners are not. This is evident in at least three aspects:

- 1) English preliterates mostly do not consider function words as words, but either attach them to the following content word or ignore them altogether; Chinese literates also often considered function words to be affixes, and interestingly a small percentage of them ignored function words altogether, *i.e.* did not mark them as either words or parts of words;
- 2) Both English preliterates and Chinese literates often do not segment compounds and phrases in smaller words;
- 3) Both English preliterates and Chinese literates sometimes rely on prosody to identify words: stress units for English children; intonation groups, potential pauses, *etc.* for Chinese literates.

English CFL learners mostly do not show such features of preliterates' word segmentation: function words are mostly considered words, compounds and phrases are segmented and prosodic clues are not taken into account. Interestingly, this is in line with the spacing conventions of the English writing system, where function words are represented as orthographic words – 'in', 'on' and 'the' (unlike some function words in the Arabic and Hebrew writing systems, see Bauer, 1996), compounds are variable, going from 'table napkin' to 'table-knife' to 'timetable' (unlike in the Dutch writing system), and prosodic boundaries are not reflected in spacing conventions (unlike in the Thai or Khmer writing systems, see Coulmas, 1999; Diller, 1996). It therefore appears that previous experience of learning a word-spaced writing system is at least partly the cause of the differences between the Chinese and English participants' view of the Chinese word. Of course this parallel between Chinese adults and English preliterates cannot be taken as evidence that the differences between Chinese and English users of Chinese are due to their respective writing systems. To demonstrate this causal link it would be necessary to compare CFL learners with different L1 writing system backgrounds. The next step could be a comparison of English and Japanese CFL users, because the Japanese writing system does not mark word boundaries with spacing,

some of its graphemes (*kanji*) represent morphemes, and the alternation of morphemic and syllabic graphemes (*kanji* and *kana*) segments the written text differently from English orthographic conventions. In any case, it is interesting to note that the L1WS orthographic conventions are sometimes also present in the word segmentations of professional linguists; when trying to identify word boundaries for previously unwritten languages, linguists sometimes rely on the orthographic conventions of English or French (see criticism in Van Dyken and Kutsch Lojenga, 1993).

In conclusion, native users of English and Chinese have different concepts of the Chinese word and different approaches to word segmentation. English CFL learners, while mostly relying on their English word awareness, also show effects of their L2 writing system. The presence of two writing systems in the mind of these multi-competent L2 and L2WS users may lead to a new awareness of the Chinese word, and possibly of the word.

Acknowledgments

The study reported here formed part of the author's doctoral thesis (Bassetti, 2004), which was supported by a Postgraduate Studentship by the Economic and Social Research Council.

References

- Bassetti, B. (2004) Second language reading and second language awareness in English-speaking learners of Chinese as a Foreign Language. PhD thesis, University of Essex, Colchester, UK.
- Bauer, T. (1996) Arabic writing. In P.T. Daniels and W. Bright (eds.) *The World's Writing Systems* (pp. 559-564). Oxford: Oxford University Press.
- Ben-Dror, I., Frost, R. and Bentin, S. (1995) Orthographic representation and phonemic segmentation in skilled readers: a cross-language comparison. *Psychological Science* 6 (3), 176-81.
- 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. and Martin, M.M. (2003) Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics* 24, 27-44.
- Bruck, M. and 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, NJ: Lawrence Erlbaum.
- Chao, Y.-R. (1968) *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chao, Y.-R. (1976) Rhythm and structure in Chinese word conceptions. In Y.-R. Chao (ed.) *Aspects of Chinese Sociolinguistics* (pp. 275-92). Stanford: Stanford University Press.
- Chinese Kenyon archives (2000) Archives of the 'Chinese@kenyon.edu mailing list', <http://lbis.kenyon.edu/services/archive/~chinese/2576.html>
- Cook, V.J. (1991) The poverty of the stimulus argument and multi-competence. *Second Language Research* 7 (2), 103-117.
- Cook, V.J. (1997) The consequences of bilingualism for cognitive processing. In A.M.B. de Groot and J.F. Kroll (eds.) *Tutorials in Bilingualism*:

- Psycholinguistic Perspectives* (pp. 279-299). Mahwah, NJ: Lawrence Erlbaum Associates.
- Coulmas, F. (1999) *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell Publishers.
- DeFrancis, J. (1996) *ABC Chinese-English Dictionary*. Honolulu: University of Hawaii Press.
- Diller, A. (1996) Thai and Lao writing. In P.T. Daniels and W. Bright (eds.) *The World's Writing Systems* (pp. 457-466). Oxford: Oxford University Press.
- Downing, J.A. (1970) Children's concepts of language in learning to read. *Educational Research* 12 106-12.
- Downing, J.A. and Oliver, P. (1974) The child's concept of a word. *Reading Research Quarterly* 9 (4), 568-82.
- Edwards, D. and Christophersen, H. (1988) Bilingualism, literacy and meta-linguistic awareness in preschool children. *British Journal of Developmental Psychology* 6, 235-244.
- Ferreiro, E. (1997) The word out of (conceptual) context. In C. Pontecorvo (ed.) *Writing Development: An Interdisciplinary View* (pp. 47-60). Amsterdam: 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.
- Hannas, W.C. (1997) *Asia's Orthographic Dilemma*. Honolulu: University of Hawaii Press.
- Hoosain, R. (1992) Psychological reality of the word in Chinese. In H.-C. Chen and 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), 341-72.
- King, P.L. (1983) Contextual factors in Chinese pinyin writing. PhD thesis, Cornell University.
- Leong, C.K. (1991) From phonemic awareness to phonological processing to language access in children developing reading proficiency. In D.J. Sawyer and B.J. Fox (eds.) *Phonological Awareness in Reading: The Evolution of Current Perspectives*. New York: Springer-Verlag.
- Li, W., Anderson, R.C., Nagy, W. and Zhang, H. (2002) Facets of metalinguistic awareness that contribute to Chinese literacy. In W. Li, J.S. Gaffney and J.L. Packard (eds.) *Chinese Children's Reading Acquisition: Theoretical and Pedagogical Issues* (pp. 87-106). Dordrecht: Kluwer Academic Publishers.
- Light, T. (1998) Review of *ABC Chinese English-Dictionary* [sic] edited by John DeFrancis with Bai Yuqing, Victor H. Mair, Robert M. Sanders, E-tu Zen Sun, and Yin Binyong. *Journal of the Chinese Language Teachers Association* 33 (1), 115-6.
- Lively, S.E., Pisoni, D.B. and Goldinger, S.D. (1994) Spoken word recognition: research and theory. In M.A. Gernsbacher (ed.) *Handbook of Psycholinguistics*. San Diego: Academic Press.
- Loizou, M. and Stuart, M. (2003) Phonological awareness in monolingual and bilingual English and Greek five-year-olds. *Journal of Research in Reading* 26 (1), 3-18.

- Mair, V.H. (2001) *ABC Dictionary of Sino-Japanese Reading*. Honolulu: University of Hawaii Press.
- Meltzer, H.S. and Herse, R. (1969) The boundaries of written words as seen by first graders. *Journal of Reading Behavior* 1, 3-14.
- Miller, K. (2002) Children's early understanding of writing and language: the impact of characters and alphabetic orthographies. In W. Li, J.S. Gaffney and J.L. Packard (eds.) *Chinese Children's Reading Acquisition: Theoretical and Pedagogical Issues* (pp. 17-29). Dordrecht: Kluwer Academic Publishers.
- Miller, K., Chen, S.-Y. and Zhang, H. (in preparation) Where the words are: judgments of words, syllables, and phrases by speakers of English and Chinese.
- Nicoladis, E. and Genesee, F. (1996) Word awareness in second language learners and bilingual children. *Language Awareness* 5 (2), 80-90.
- Packard, J.L. (1998) Introduction. In J.L. Packard (ed.) *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese* (pp. 1-34). Berlin, New York: Mouton de Gruyter.
- Prakash, P. (2000) Is phonemic awareness an artefact of alphabetic literacy?! Poster presentation, ARMADILLO - 11, Texas (October 13-14, 2000).
- Read, C.A., Zhang, Y., Nie, H. and 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. and Clapham, C. (2001) Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18, 55-81.
- Scholes, R.J. (1993) On the orthographic basis of morphology. In R. Scholes (ed.) *Literacy and Language Analysis* (pp. 73-95). Hillsdale, NJ: Erlbaum.
- SJDHBZ, Shanghai Jiaotong Daxue Hanzi Bianmazu (eds.) (1988) *Han zi xin xi zi dian* [Dictionary of Chinese character information]. Beijing: Kexue Chubanshe.
- Sproat, R., Shih, C., Gale, W. and Chang, N. (1996) A stochastic finite-state word-segmentation algorithm for Chinese. *Computational Linguistics* 22 (3), 377-404.
- Tsai, C.-H., McConkie, G.W. and 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 (http://www.geocities.com/hao520/research/papers/cht_asi.htm.)
- T'ung, P.-C. and Pollard, D.E. (1982) *Colloquial Chinese*. New York: Routledge.
- Van Dyken, J.R. and 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.

Endnotes

ⁱ Whether the word is a valid linguistic construct or not is totally irrelevant here; a construct does not need to be scientifically valid in order to affect people's thinking. Indeed, astrological signs are used by some people to interpret human behaviour, regardless of whether they are considered valid constructs by psychologists.

ⁱⁱ Words cannot be shorter than one hanzi.