Behavioral Studies: Processing of Chinese Compounds

INTRODUCTION

Chinese compounds are words containing two or more root morphemes. Compound processing involves the real-time lexical access of compounds in comprehension and production by normal adults. This lemma focuses on behavioral studies (for neurological studies, see Lexical process: Morphology); the focus on compounds is necessitated by the paucity of processing studies on Chinese affixation (cf. Myers et al. 2006). All studies reviewed here were conducted on Mandarin unless otherwise noted.

Chinese compounds are notoriously difficult to distinguish from syntactic phrases, partly for orthographic reasons. While English orthography marks word boundaries (with spaces) but not morpheme boundaries, Chinese orthography does the reverse. This causes Chinese readers to disagree about how text should be segmented into words (詞 *cí*), as confirmed experimentally by Hoosain (1992) with speakers in Hong Kong. The fuzziness of the word notion in Chinese calls into question the dominant experimental paradigm in the study of lexical access, the lexical decision task. In this task, participants must decide if a presented form is a real word or not, but if Chinese speakers do not have clear intuitions about wordhood, this task seems inappropriate (Hung et al. 1999).

Nevertheless, there is considerable evidence that there is a word level in Chinese lexical access. Mattingly and Xu (1993) found that Chinese readers were faster to detect characters in two-character strings when they were real words. This word superiority effect implies that even in Chinese, word representations are obligatorily activated in reading. Taft (2003) found that Chinese readers showed different response patterns when performing lexical decisions on characters (vs. pseudo-characters) depending on whether they were asked to evaluate them as characters (字 *zì*) or as words (詞 *cí*); only in the latter case did they tend to reject bound morphemes. Words are key to fluent reading: Feng (2006) reports that the mean length of eye saccades (jumps) while reading Chinese is approximately 1.5 words, approximately the same length as for English reading despite orthographic differences, and Hsu and Huang (2000) found that inserting spaces between characters within words slows the reading of Chinese text. Finally, in Chinese, just as in all other languages that have been studied, common words are recognized more quickly and accurately than rare ones (Myers 2006, Cai and Brysbaert 2010).

READING CHINESE COMPOUNDS

Most experiments on Chinese compound processing have examined reading. Such studies have tested the influence of lexical statistics, the meaning and syntactic category of characters, and compound-internal structure.

One lexical statistic affecting compound access is morpheme frequency, which is expected to affect word access when morphemes become active. In practice, character frequency is usually used as a proxy. Only recently have psycholinguists begun to develop statistical tools to quantify the character polysemy that makes morpheme frequency difficult to compute (Galmar and Chen 2010).

Taft et al. (1994) matched the whole-word frequency of two-character compounds while manipulating character frequency. Participants were faster to judge compounds as real words if both characters were common than if one was rare. This pattern suggests that during compound recognition, component morphemes are obligatorily activated. However, rather than seeing such results as demonstrating the decomposition of compounds into morphemes, as might be the case for a language like English, Myers (2006) argues that they instead reflect the composition of compounds out of characters during reading.

Taft et al. also found that when both characters were rare, compounds were responded to just as quickly as compounds composed of two common characters. Myers (2006) argues that this kind of result may follow from character transition probability, which increases word cohesiveness, facilitating word access. A common compound composed of rare characters must have high transition frequency, since these characters tend to appear together.

Transition probability is closely related to neighborhood density, or the number of words that differ from a compound by only one character. As Tsai et al. (2006) found, the access of a compound may benefit from its having many neighbors, given that compounds in dense neighborhoods were not only recognized faster in a lexical decision task, but also skipped more frequently and fixated for shorter durations during sentence reading (suggesting that they were easier to process).

Similarly, Huang et al. (2006) found that high frequency compounds showed facilitation from neighborhood density. However, they found that with low-frequency targets, neighbors acted as competitors, inhibiting lexical access in dense neighborhoods. They also established that first characters had a stronger neighborhood effect than second characters, consistent with a character-by-character word composition process.

Huang et al. (2011) found a similar pattern of results when neighborhoods were defined morphemically (e.g. distinguishing 花 *huā* 'flower' in 花園 *huāyuán* 'flower garden' from 花 *huā* 'spend' in 花錢 *huāqián* 'spend money'), and Bai et al. (2011) found that lexical competition can also arise from words sharing the same characters in reversed order (e.g. 事故 *shìgù* 'accident' and 故事 *gùshì* 'story'), though in both studies the effects were more robust in the neurological measure (evoked potentials).

Morpheme activation predicts that semantically opaque compounds like 花生 *huāshēng* 'peanut' should be processed differently from transparent compounds like 花園

huāyuán 'flower garden', since only in the former do the meanings of the component morphemes compete with that of the whole word. Consistent with this prediction, Mok (2009) found a stronger word superiority effect for character detection in compounds that contained at least one semantically opaque morpheme as compared with fully transparent compounds. This suggests that the word-level representation of opaque compounds is activated more strongly than that of transparent compounds, in order to remain accessible despite competition from the component morphemes.

Further evidence of word-level dominance in opaque compounds is that the meaning of such compounds may become active before that of their component morphemes. Liu and Peng (1997) found that when target compounds for lexical decisions appeared immediately after prime compounds, opaque primes (e.g. 草率 *cǎoshuài* 'careless', literally 'grass-command') facilitated responses to targets related to the prime's whole-word meaning (e.g. 馬虎 *mǎhǔ* 'careless') but not to targets related to component morphemes (e.g. 樹木 *shùmù* 'tree', related to 草 *cǎo* 'grass'). Only with a longer delay between prime and target was morphemic facilitation observed.

If characters are activated during compound access, common component characters should facilitate the recognition of transparent compounds, but inhibit the recognition of opaque ones, where the semantics of characters and words compete. Just such a pattern of results was reported by Peng et al. (1999). The syntactic categories of morphemes and compounds may also compete: Hsu et al. (2004) found faster lexical decisions for nominal compounds and verbal compounds if composed, respectively, of nouns and verbs, rather than the reverse.

Distinguishing morpheme activation from character activation, and morphemic priming from priming via whole-word semantics, present serious challenges. One of the most thorough attempts to address them is Zhou et al. (1999). In a series of visual lexical decision experiments, prime and target compounds were matched in whole-word semantic relatedness, while prime-target relations were varied in terms of characters, character meanings, and syllables. Thus for any given target (e.g. 華貴 *huáguì* 'luxury', literally 'splendid-valuable') the researchers contrasted primes sharing a character with the target but not character meaning (e.g. 華僑 *huáqiáo* 'overseas Chinese', literally 'China-abroad') against primes sharing both character and character meaning, and thus presumably a morpheme (e.g. 華麗 *huálì* 'gorgeous', literally 'splendid-beautiful'). While both character and morpheme priming were found, the latter effect was stronger (syllable priming was not detected at all). As the above sample materials show, however, it is possible to question whether whole-word semantic relatedness between primes and targets was fully matched across conditions.

Compound processing involves not just characters, morphemes, and words, but also their structured relationships. In particular, not only does Chinese productively form modifier-head compounds like 書店 *shūdiàn* 'bookstore', a structure familiar from languages like English, but unlike English, also coordinative compounds like 父母 *fùmǔ* 'parents' (literally 'father-mother'). There is some evidence that this structural difference is associated with differences in processing.

Zhang and Peng (1992) examined character frequency effects in modifier-head and coordinative compounds in lexical decision. They found that in modifier-head compounds, a facilitative effect was found only for the first character, consistent with the character-by-character activation implied by other studies (e.g. Huang et al. 2006, reviewed above). For coordinative compounds, however, both character frequencies were equally facilitative, which makes sense given the equally important roles of both morphemes.

Liu and McBride-Chang (2010) found a different sort of contrast in a primed lexical decision study on written Cantonese compounds. Prime-target pairs crossed whole-word semantic relatedness with structure (modifier-head vs. coordinative). For modifier-head primes, the familiar whole-word semantic priming only occurred if the target also had modifier-head structure, while for coordinative primes, whole-word semantic priming was only found if the target had a modifier-head structure. The researchers interpret the results as implying that the modifier-head structure is inherently more informative than the coordinative structure, perhaps because the morphemes in coordinative compounds have a highly predictable relationship (related or opposite meanings), which is not true for modifiers and heads.

The processing of modifier-head relations was the explicit focus of Ji and Gagné (2007). In a sense-nonsense judgment task on written Chinese compounds, transparent modifier-head targets (e.g. 雪人 *xuěrén* 'snowman': MADE-OF relation) were preceded by transparent primes that shared a target character (morpheme) but differed in modifier-head relation (e.g. 雪球 *xuěqiú* 'snowball': MADE-OF relation, vs. 雪鏟 *xuěchǎn* 'snow shovel': FOR relation). They found that prime-target pairs sharing modifier-head relation were responded to more quickly than pairs that did not, suggesting that such relations are active in compound processing. Moreover, unlike English (Gagné 2001), this effect was found even if the primes and targets shared heads rather than modifiers. The researchers interpret this difference as a side-effect of the lack of word boundaries in Chinese orthography. This make it impossible for a reader to identify a character as modifier until a following head is reached, which in turn makes heads more important for defining modifier-head relations than is the case in English.

THE PROCESSING OF SPOKEN CHINESE COMPOUNDS

Although the lexical access literature has traditionally focused on reading, there have been a number of important findings in both the recognition and production of spoken Chinese compounds.

Modality (e.g. reading vs. listening vs. speech) is expected to have dramatic effects on Chinese compound access. Chinese morphemes are almost always monosyllabic and the syllable inventory is highly restricted in size. As Packard (1999) points out, the resulting morphemic homophony makes it implausible to suppose that listeners access spoken Chinese compounds via their morphemes (the way readers seem to access written words via characters), since such a strategy would require them to activate all possible morphemic homophones and search the lexicon for their combinations.

Given this insight, the recognition of spoken compounds should not require morpheme activation, and thus show no morpheme (i.e. character) frequency effects. Indeed, when Zhou and Marslen-Wilson (1994) ran a series of lexical decision experiments on spoken compounds, systematically varying the frequencies of characters and words (and syllables), no character frequency effects were detected, though word frequency effects remained as robust as ever. The lack of character frequency effects is particularly striking given that they tested semantically transparent compounds, where morpheme activation should benefit word access.

However, morphemes do seem to become activated eventually in the recognition of spoken compounds. Using materials similar to those tested by Zhou et al. (1999) except spoken, Zhou and Marslen-Wilson (1995) ran a series of primed auditory lexical decision experiments, where primes and targets were presented with varying time lags, from short (within the same trial) to long (prime and target in separate trials). They found that when primes and targets shared a character with the same meaning (i.e. shared a morpheme), targets were responded to more quickly than controls. Moreover, unlike character or syllable priming on their own, this effect remained even with a long lag between prime and target, consistent with previous research suggesting that long-lasting priming is a hallmark of true morphemic activation (Feldman 2003).

The production of spoken compounds also involves processes quite different from both reading and spoken word recognition. A simple example of such a difference is the irrelevance of purely orthographic information in producing spoken compounds. This was confirmed by Bi et al. (2009), who found production effects due to shared orthographic elements (phonetic components of characters) only when compounds were read aloud, but not when they were produced in picture naming or in response to semantically related prompts.

Surprisingly, however, characters may not influence compound production even when they represent morphemes. Janssen et al. (2008) failed to find character frequency effects on production latencies for Chinese compounds in picture naming, nor did Chen and Chen (2006) find such effects when compounds were prompted by semantically related words. Chen and Chen (2006, 2007) also found no additional implicit priming across compounds, trained as a set, that shared both the first character and the first syllable, beyond that found for compounds sharing the first syllable only. These results contrast with the facilitative morpheme frequency and priming effects found in Dutch compound production (Roelofs 1996), though Janssen et al. (2008) found no morpheme frequency effects in English compound production either.

Chen and Chen suggest that phonological reasons explain why morphemes need not be activated in preparing the phonological form of Chinese compounds. In English (and Dutch), resyllabification tends not to apply between roots, so the /b/ in *crab* does not act as an onset in the compound *crabapple*, unlike the /b/ in the suffixed *crabby*. In Chinese, by contrast, true compounds like 牙醫 *yáyī* 'dentist' and disyllabic morphemes like 螞蟻 *mǎyǐ* 'ant' show no systematic phonological differences.

Nevertheless, other studies suggest that Chinese speakers do active morphemes at some stage of compound production. One such study is Bates et al. (2003), which involved the naming of the same set of pictures by speakers of seven languages, including English and Chinese. For most of these languages, when speakers reused the same word for more than one picture, suggesting that the pictures were difficult to name, they chose a common word, but Chinese speakers tended to produce low-frequency (compound) words in this situation. This suggests that Chinese speakers respond to lexical retrieval difficulty by productively coining words, which tend to be rarer than the ready-made words preferred by speakers of other languages.

A different study with similar implications is Perry and Zhuang (2005). Chinese speakers were asked to name pictures of objects that have both monosyllabic and disyllabic names (e.g. 'elephant', named either 大象 *dàxiàng* or 象 *xiàng*). Speakers were more likely to choose the disyllabic name if the pictures appeared in a list with other words with disyllabic names. The authors interpret this as showing that word choice can be influenced by phonology, but as Myers (2010) points out, an alternative interpretation is that what is influenced is the choice to apply the productive morphological operation of morphemic truncation.

SUMMARY

While Chinese compounds behave as units at some point in processing, the activation and integration of their component morphemes depend in complex ways on a wide variety of factors, including lexical frequency, semantic transparency, compound-internal structure, and modality. The last two factors are underexplored, however, as is the processing of Chinese affixation.

References

Bai, Chen, Shushan Cai and Petra B. Schumacher, "Reversibility in Chinese word formation influences target identification", *Neuroscience Letters* 499, 2011, 14-18.

Bates, Elizabeth, Simona D'Amico, Thomas Jacobsen, Anna Székely, Elena Andonova, Antonella Devescovi, Dan Herron, Ching Ching Lu, Thomas Pechmann, Csaba Pléh, Nicole Wicha, Kara Federmeier, Irini Gerdjikova, Gabriel Gutierrez, Daisy Hung, Jeanne Hsu, Gowri Iyer, Katherine Kohnert, Teodora Mehotcheva, Araceli Orozco-Figueroa, Angela Tzeng and Ovid Tzeng, "Timed picture naming in seven languages", *Psychonomic Bulletin & Review* 10 /2, 2003, 344-380.

Bi, Yanchao, Tao Wei, Niels Janssen and Zaizhu Han, "The contribution of orthography to spoken word production: Evidence from Mandarin Chinese", *Psychonomic Bulletin & Review* 16/3, 2009, 555-560.

Cai, Qing and Marc Brysbaert, "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles", *PLoS ONE* 5/6, 2010, 1-8.

Chen, Jenn-Yeu and Train-Min Chen, "Form encoding in Chinese word production does not involve morphemes", *Language and Cognitive Processes* 22/7, 2007, 1001-1020.

Chen, Train-Min and Jenn-Yeu Chen, "Morphological encoding in the production of compound words in Mandarin Chinese", *Journal of Memory and Language* 54, 2006, 491-514.

Feldman, Laurie Beth, "What the repetition priming methodology reveals about morphological aspects of word recognition", in: Jeffrey S. Bowers and Chad J. Marsolek, eds., *Rethinking Implicit Memory*, Oxford: Oxford University Press, 2003, 124-138.

Feng, Gary, "Eye movement in Chinese reading: Basic processes and cross-linguistic differences", in: Ping Li, Li Hai Tan, Elizabeth Bates and Ovid J. L. Tzeng, eds., *The Handbook of East Asian Psycholinguistics Vol. 1: Chinese*, Cambridge: Cambridge

University Press, 2006, 187-194.

Gagné, Christina L., "Relation and lexical priming during the interpretation of noun-noun combinations", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27/1, 2001, 236-254.

Galmar, Bruno and Jenn-Yeu Chen, "Identifying different meanings of a Chinese morpheme through semantic pattern matching in augmented minimum spanning trees", *The Prague Bulletin of Mathematical Linguistics* 94, 2010, 15-34.

Hoosain, Rumjahn, "Psychological reality of the word in Chinese", in: Hsuan-Chih Chen and Ovid J. L. Tzeng, eds., *Language Processing in Chinese*, Amsterdam: Elsevier Science Publishers, 1992, 111-130.

Hsu, Jean Ching-Fen, Ovid J.-L. Tzeng and Daisy L. Hung, "Syntactic effects on constituent components in word recognition: Evidence from compound nouns and compound verbs in Chinese", *University of Maryland Working Papers in Linguistics* 13, 2004, 173-193.

Hsu, Sheng-Hsiung, and Kuo-Chen Huang, "Interword spacing in Chinese text layout", *Perceptual and Motor Skills*, 91, 2000, 355-363.

Huang, Hsu-Wen, Chia-Ying Lee, Jie-Li Tsai and Ovid J.-L. Tzeng, "Sublexical ambiguity effect in reading Chinese disyllabic compounds", *Brain & Language* 117, 2011, 77-87.

Huang, Hsu-Wen, Chia-Ying Lee, Jie-Li Tsai, Chia-Lin Lee, Daisy L. Hung and Ovid J.-L. Tzeng, "Orthographic neighborhood effects in reading Chinese two-character words", *NeuroReport* 17/10, 2006, 1061-5.

Hung, Daisy L., Ovid J. L. Tzeng and C.-Y. Ho, "Word superiority effect in the visual processing of Chinese", in: Ovid J. L. Tzeng, ed., *Journal of Chinese Linguistics Monograph Series No. 13: The Biological Bases of Language*, 1998, 61-95.

Janssen, Niels, Yanchao Bi and Alfonso Caramazza, "A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds", *Language* *and Cognitive Processes* 23, 2008, 1191-1223.

Ji, Hongbo and Christina L. Gagné, "Lexical and relational influences on the processing of Chinese modifier-noun compounds", *The Mental Lexicon* 2/3, 2007, 387-417.

Liu, Ying and Dan-ling Peng, "Meaning access of Chinese compounds and its time course", in: Hsuan-Chih Chen, ed., *Cognitive Processing of Chinese and Related Asian Languages*, Hong Kong: The Chinese University Press, 1997, 219-232.

Liu, Phil D. and Catherine McBride-Chang, "Morphological processing of Chinese compounds from a grammatical view", *Applied Psycholinguistics* 31, 2010, 605-617.

Mattingly, Ignatius G. and Yi Xu, "Word superiority in Chinese", *Haskins Laboratories Status Report on Speech Research*, SR-113, 1993, 145-152.

Mok, Leh Woon, "Word-superiority effect as a function of semantic transparency of Chinese bimorphemic compound words", *Language and Cognitive Processes* 24/7-8, 2009, 1039-1081.

Myers, James, "Processing Chinese compounds: A survey of the literature", in: Gary Libben and Gonia Jarema, eds., *The Representation and Processing of Compound Words*, Oxford: Oxford University Press, 2006, 169-196.

Myers, James, "Chinese as a natural experiment", *The Mental Lexicon* 5/3, 2010, 423-437.

Myers, James, Yu-chi Huang and Wenling Wang, "Frequency effects in the processing of Chinese inflection", *Journal of Memory and Language*, 54/3, 2006, 300-323.

Packard, Jerome L., "Lexical access in Chinese speech comprehension and production", *Brain and Language* 68, 1999, 89-94.

Peng, Danling, Ying Liu, and Chunmao Wang, "How is access representation organized? The relation of polymorphemic words and their morphemes in Chinese", in: Jian Wang, Albrecht Inhoff, and Hsuan-Chih Chen, eds., *Reading Chinese Script: A Cognitive Analysis*, Mahwah, New Jersey: Lawrence Erlbaum Associates, 1999, 65-89.

Perry, Conrad and Jie Zhuang, "Prosody and lemma selection", *Memory & Cognition* 33/5, 2005, 862-870.

Roelofs, Ardi, "Morpheme frequency in speech production: Testing WEAVER", in: Geert E. Booij and Jaap. van Marle, eds., *Yearbook of Morphology 1996*, Dordrecht, The Netherlands: Kluwer Academic Press, 1996, 135-154.

Taft, Marcus, "Morphological representation as a correlation between form and meaning", in: Assink, Egbert, M. H. and Dominiek Sandra, eds., *Reading Complex Words: Cross-Language Studies*, Amsterdam: Kluwer, 2003, 113-138.

Taft, Marcus, Jong-Tsun Huang and Xiaoping Zhu, "The influence of character frequency on word recognition responses in Chinese", in: Hsing-Wu Chang, Jong-Tsun Huang, Chih-wei Hue, and Ovid J. L. Tzeng, eds., *Advances in the Study of Chinese Language Processing, Vol. 1*, Taipei: Department of Psychology, National Taiwan University, 1994, 59-73.

Tsai, Jie-Li, Chia-Ying Lee, Ying-Chun Lin, Ovid J. L. Tzeng and Daisy L. Hung, "Neighborhood size effects of Chinese words in lexical decision and reading", *Language and Linguistics* 7/3, 2006, 659-675.

Zhang, Biyin and Danling Peng, "Decomposed storage in the Chinese lexicon". in: Hsuan-Chih Chen and Ovid J. L. Tzeng , eds., *Language Processing in Chinese*, Amsterdam: North-Holland, 1992, 131-149.

Zhou, Xiaolin and William D. Marslen-Wilson, "Words, morphemes, and syllables in the Chinese Mental lexicon", *Language and Cognitive Processes* 9/3, 1994, 393-422.

Zhou, Xiaolin and William D. Marslen-Wilson, "Morphological structure in the Chinese mental lexicon", *Language and Cognitive Processes* 10/6, 1995, 545-600.

Zhou, Xiaolin, William D. Marslen-Wilson, Marcus Taft and Hua Shu, "Morphology, orthography, and phonology in reading Chinese compound words", *Language and*

Cognitive Processes 14/5/6, 1999, 525-565.

James Myers

Summary

This lemma focuses on the real-time lexical access of Chinese compounds in comprehension and production by normal adults, as studied through behavioral experiments.

Index terms

Cantonese

characters

compounds

lexical access

morphemes

processing

reading

speech

words