**Comprehension – Behavioral Studies: The Sapir-Whorf Hypothesis**

**HISTORY AND CONTROVERSY**

The Sapir-Whorf hypothesis refers to the claim that features of one's language affect one's thoughts (linguistic determinism), so that different languages foster different conceptual systems (linguistic relativity). This lemma reviews experimental tests of this hypothesis in Chinese.

The hypothesis is so named because of notions expressed in Sapir (1929) and Whorf (1939) (cf. Hill and Mannheim 1992). In its first modern application to Chinese (cf. Kwan 2001), Hockett (1954) cites the greater emphasis on manner in Chinese than in English verbs of holding and breaking, as well as the lack of number marking in Chinese, but he provides no evidence that these differences affect thought.

The Sapir-Whorf hypothesis remains controversial (Gumperz and Levinson 1996), in part because researchers differ in their views on its content and value. Pinker (2007) distinguishes ten distinct interpretations of linguistic determinism, among them "thinking for speaking" (Slobin 1996), in which speakers habitually encode concepts in ways that make them easier to translate into speech. However, it has proven difficult to demonstrate empirically that such conceptual habits are active in the absence of linguistic processing. Putatively nonlinguistic tasks may still trigger the use of language as a silent mental tool (e.g. activating plausible category names). Nevertheless, if language structure does affect behavior, this fact has scientific and practical importance regardless of the mechanism.

Linguistic relativity poses the additional challenge of inferring causation from correlation. Speakers of different languages also differ in non-linguistic experience, which may influence both language and concepts (Ji et al. 2004). For example, Ross et al. (2002)
found that speaking Chinese led Chinese-English bilinguals to emphasize group memberships in self-descriptions more than when speaking English: each language primed the social assumptions of the associated culture, despite the lack of linguistic features marking them.

**EXPERIMENTAL TESTS OF THE SAPIR-WHORF HYPOTHESIS IN CHINESE**

Experiments have tested potential linguistic influences on cognition along a continuum from lexical to grammatical.

In a study testing for lexical effects, Brown (1986) found that English speakers draw causal inferences from verbs matching their associated adjectives (e.g. subject-oriented *attract/attractive* vs. object-oriented *abhor/abhorrent*). However, when the materials were translated into Chinese, which does not have such adjective associates, speakers in Hong Kong had the same causal interpretations. These results support universal constraints on verbal semantics, not linguistic relativity. Frank et al. (2000) were led to similar conclusions regarding types of shame distinguished lexically in Chinese but not English (e.g. 羞恥 *xiūchǐ* vs. 慚愧 *cánkuì*). English-speaker descriptions of various shame-related scenarios fit categories corresponding to those lexicalized in Chinese, again supporting universal semantics.

In another lexical study, Hoffman et al. (1986) gave English monolinguals and Chinese-English bilinguals descriptions fitting an English or Chinese personality schema (e.g. the same person may be described as 世故 *shìgù* in Chinese or less efficiently in English as worldly and somewhat reserved). Novel descriptions derived from schemas in the testing language tended to be misidentified as having already appeared, showing that the schemas were mentally active. However, the authors admit that these results are consistent with language influencing behavior in the task rather than thought per se.
Moving into semi-closed class vocabulary, Huang (1999) examined two time systems in Chinese. Chinese months are named sequentially after numbers (e.g. 十二月 shíèr yuè 'December', literally 'twelve month'), but the 節氣 jìeqì system uses twenty-four lexically distinct units (e.g. 大雪 dàxuě 'heavy snow'). Speakers fluent in both systems judged if a given month or jìeqì unit occurs three or five units before or after another. While jìeqì judgments were slower when units were further apart in time, month judgments were equally fast, as if participants used arithmetic. These results again seem to show language being used as a tool in the task.

Chinese commonly uses a vertical metaphor for time (e.g. 上個月 shàng gè yuè 'previous month', literally 'up a month'), whereas English favors a horizontal metaphor (e.g. 'the month before'). To see if the vertical time metaphor has truly restructured Chinese cognition, Boroditsky (2001) tested speakers exclusively in English. Bilingual Chinese-English speakers (but not English monolinguals) made faster 'earlier'/later' judgments about the relative order of months after making overtly spatial judgments on vertical primes.

However, these effects could not be replicated, in either Chinese (Chen 2007) or English (January and Kako 2007); Boroditsky et al. (2011) concede their own replication difficulty. Moreover, Chen (2007) notes that even in Chinese, horizontal time metaphors far outnumber vertical ones, and January and Kako (2007) ask why brief training caused even English speakers to show vertical priming, when years of English experience barely influenced the Chinese-English bilinguals.

Boroditsky et al. (2011) (see also Miles et al. 2011) tested the same claim with a new task. Participants saw a sequence of two pictures (e.g. photographs of the same person at different ages) and judged whether the second was 'earlier' or 'later'. Chinese but not English speakers were faster if the 'earlier' response key was above the 'later' key. Similarly, Furhman
et al. (2011) found that Chinese speakers were more likely than English speakers to point above or below a reference point to indicate relative time.

Yet ascribing cognitive effects solely to the vertical metaphor is difficult given that cognition is also affected by the vertical orientation of Chinese text (Chan and Bergen 2005). The effect of text orientation on temporal judgments was confirmed by Chen and O'Seaghdha (forthcoming), who found that only participants from Taiwan (where vertical text is common) behaved as described in Boroditsky et al. (2011); participants from China (where vertical text is rarer) behaved like the English-speaking participants.

Chinese classifiers represent another semi-closed lexical class. Zhang and Schmitt (1998) found that Chinese speakers, but not English speakers, judged noun pairs as more similar if they shared the same classifier. In a more ambitious study, Saalbach and Imai (2007) presented Chinese and German speakers with pairs of nouns related taxonomically (categorically), thematically (relationally), or via shared classifier. Although two out of four tasks showed stronger classifier effects for Chinese speakers, the Germans also grouped classifier-related nouns together, suggesting that classifiers encode universal semantic features.

Although shape is a universal semantic feature, experience with shape classifiers may make Chinese speakers more sensitive to it. Kuo and Sera (2009) found that in classifying objects, Chinese speakers were influenced by shape more than English speakers, though both groups primarily classified taxonomically and functionally. Imai et al. (2010) found similar results for Chinese and German children in a classification task, but in other tasks, both groups favored either shape or taxonomic categories.

Such classifier effects seem to require linguistic processing. Gao and Malt (2009) had Chinese speakers memorize nouns in sentence contexts. Noun recall fell more into classifier-defined clusters (compared with English speakers), but only when nouns had
appeared in sentences with overt classifiers. Huang and Chen (2011) replicated this pattern in a different task. In an eye-tracking study, Huettig et al. (2010) found that Mandarin speakers looked at pictures of objects sharing the same classifier as an object named in an auditorily presented sentence, but only when the classifier was also presented. Tsang and Chambers (2011) report related results in Cantonese.

One function of Chinese classifiers is to individuate nouns, leading some to propose that all Chinese nouns are inherently mass. Consistent with this, Li et al. (2009a) found that Chinese speakers tended to group novel entities with same-material substances more often than English speakers, who tended to classify them as objects. However, this effect appeared only in linguistic tasks (interpreting names or deixis), not in cognitive tasks (rating objecthood/substancehood). Similarly, Barner et al. (2009) found that Chinese-English bilinguals made material-based generalizations only when tested in Chinese. Further undermining the notion that all Chinese nouns are mass, Li et al. (2009b) found that Mandarin-learning infants had the cognitive capacity to recognize plurality even before acquiring the plurality marker [men].

Unlike English, Chinese numbers encode a decimal system (e.g. 十二 shí’èr ‘ten-two’ for ‘twelve’, 二十 èrshí ‘two-ten’ for ‘twenty’). Miura et al. (1988) found that this decimal system affected how Chinese children counted with blocks representing tens and ones, as compared with English-speaking children. Geary et al. (1996) replicated these results. However, not only are they consistent with language being used as a tool (including overt counting), but they have minimal practical implications: Wang and Lin (2009) found no difference in computation ability between Chinese and American students, while in other areas of mathematics the Chinese advantage increases over years of schooling, suggesting roles for teaching or learning strategies.

Moving to grammatical morphemes, Chen and Su (2011) studied the effect of gender,
marked in English but not Chinese (他 tā traditionally represents both 'he' and 'she', with the character 她 tā 'she' a recent invention). English participants identified the biological gender of people in stories, whether by listening or reading, more quickly and accurately than Chinese participants. While this discovery is intriguing (especially given the sharply distinguished gender roles in traditional Chinese culture), note that all pronouns were presented overtly, making it unnecessary to posit Chinese-specific gender-encoding habits independent of Chinese language processing.

Another grammatical difference between English and Chinese is that only the former marks tense. Chen et al. (2012) asked Chinese speakers to judge whether pictures showing one stage of an action matched Chinese sentences marked for prospective (準備要 zhēn bèi yào), progressive (正在 zhèngzài), or completive aspect (剛剛 gānggāng). Accuracy was higher for participants with greater English proficiency; monolingual English speakers, given the task in English with tense markings, were most accurate. The authors conclude that experience with English tense makes it easier to decompose events into stages. However, the authors acknowledge that participants may have described the pictures to themselves (language as a tool), and that broader cross-cultural differences in encoding event structure (Ji et al. 2004) may have played a role.

Finally, in one of the most notorious applications of the Sapir-Whorf hypothesis in any language, Bloom (1981) claimed that the lack of overt counterfactual marking in Chinese (cf. the English subjunctive) made it difficult for Chinese speakers to interpret if-then structures with false premises. However, Chinese speakers actually process counterfactuals quite reliably (Feng and Yi 2006) and Bloom's effects have repeatedly failed to replicate using more natural materials (Au 1983, Liu 1985, Wu 1994). Culture again seems to play a role: Lardiere (1992) found that speakers of Arabic, which explicitly marks counterfactuals, also balked at Bloom's counterfactuals, perhaps, Lardiere speculates, because of the prominence
of rote learning in both Chinese and Arabic educational traditions.

References


Boroditsky, Lera, Orly Fuhrman and Kelly McCormick, "Do English and Mandarin speakers think about time differently?", *Cognition* 118/1, 2011, 123-129.


Chen, Jenn-Yeu and Padraig G. O'Seaghdha, "Do Mandarin and English speakers think about


Gumperz, John J. and Stephen C. Levinson, "Introduction to Part I", in: John J. Gumperz and


Huang, Wei, "Reasoning about conventional time as a function of conventional time systems", *Memory and Cognition* 27/6, 1999, 1080-1086.


Kuo, Jenny Yi-chun and Maria D. Sera, "Classifier effects on human categorization: The role of shape classifiers in Mandarin Chinese", *Journal of East Asian Linguistics* 18/1, 2009, 1-19.


Li, Peggy, Yarrow Dunham and Susan Carey, "Of substance: The nature of language effects on entity construal", *Cognitive Psychology* 58/4, 2009a, 487-524.

Li, Peggy, Tamiko Ogura, David Barner, Shu-Ju Yang and Susan Carey, "Does the conceptual distinction between singular and plural sets depend on language?", *Developmental Psychology* 45/6, 2009b, 1644-1653.


Miura, Irene T., Chungsoon C. Kim, Chih-Mei Chang and Yukari Okamoto, "Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons", *Child Development* 59/6, 1988, 1445-1450.

Ross, Michael, W. Q. Elaine Xun and Anne E. Wilson, "Language and the bicultural self", 


James Myers
Summary

The Sapir-Whorf hypothesis refers to the claim that features of one's language affects one's thoughts (linguistic determinism), so that different languages foster different conceptual systems (linguistic relativity). This lemma reviews experimental tests of this hypothesis in Chinese.