Semantic modeling of Chinese characters

Kent Lee

University of Illinois, Urbana-Champaign, & Beckman Institute

BECKMAN INSTITUTE
AT THE UNIVERSITY OF ILLINOIS

Introduction

Chinese characters consist of one or more components, most often:
- One semantic radical ("root"), e.g., 木 'tree,' 口 'mouth,' 火 'fire,' 彳 'footstep'
- Often a phonetic indicator ("stem"): 禾 *hé* → 和 *hé* 'with,' and/or additional semantics. Character semantics may seem clear, but often are abstract to very opaque:

radical	relatively less opaque	relatively more opaque
木	枝 *zhī*, 'branch'	朵 *duǒ* [measure word]
火,	災 *zāi*, 'disaster	熊 *xióng*, 'a bear'
彳	役 *yì*, 'labor'	衔 *xián*, 'hold, harbor'

Character recognition studies show RT effects for general semantic transparency vs. abstractness or opacity; contradictory findings for time course of semantic vs. phonological priming

But…
- Unclear if this is the best or only control for semantic effects in lexical access; specific semantic effects remain poorly understood or untested
- No clear way of quantifying transparency or abstractness, or to know a priori what effects to control for or how to quantify them; better controls needed.

Thus, exploratory techniques are needed to ascertain relevant semantic dimensions of mental representations of characters, and control variables for RT studies

Method

Survey (part 1): Radicals
20 subjects (Hong Kong, Taiwan, Macao) evaluated semantic transparency of the 214 radicals (40 per survey, 7 pt. scale)

Survey (part 2): Radical-character relatedness
3130 characters rated (3000 most freq. characters + 130 low-freq. items for rare radicals) on semantic relatedness of radical to whole character meaning, e.g., 木 cf. 枝 or 朵

Analysis:
Hierarchical linear modeling analysis of similarity ratings ("Sim.")

Part 1: Radical variables:
Sim = char, rad., strokes, rad. frequency, list cond., subject (as random effect), nationality
Predicted Means yield rad. transparency index for Part 2

Part 2: Character variables:
Sim. = nationality, subject, radical, rad. transparency (from Part 1), lexical frequency, #strokes

HLM results

1. Radical abstractness (semantic transparency)

• Subject (random effect)	β = .6625, p=.002	• # strokes	n.s.
• Radical	χ² = 1625.6, p<.0001	• Radical frequency	n.s.
		• List condition	n.s.

Abstractness effects, consistent across raters:

• *High transparency*	59 radicals
• *Medium transparency*	18 sig., 10 marginally sig.
• *Low transparency*	8 sig., 8 marginal
• *Inconsistently rated*	111 radicals, similarity ratings *n.s.*

Predicted means yield a semantic transparency or abstracness index for radicals for HLM Part 2.

2. Radical-character semantic relatedness

• Character	χ² = 3.6, p<.0001	• Radical*	n.s.
• # strokes	β = .1332, p<.0001	• Character frequency	n.s.
• Subject (random ef.)	β = .8679, p=.0011	• Nationality	n.s.
		• Rad. abstractness/trans.*	n.s.

- The non-significance of radicals overall was surprising, but some individual radicals showed consistent effects on relatedness judgments, indicating a need for a subsequent hierarchical non-linear (HNLM) analysis.

3. Discussion
- Radical opacity is an effect independent of stroke density or frequency
- Radical-character consistency in relatedness across characters is independent of lexical frequency and radical transparency, but related to stroke or component density
- Contribution of semantic components to characters can be abstract and inconsistent across characters, serving as cues, albeit weak, for orthographic look-up of characters
- Radical frequency, rad. abstractness, consistency in characters are not necessarily correlated:

	radical	gloss	rad. rating	char.	gloss	rad. consistency	char. relatedness
high freq. radicals	木	'tree, wood'	6.47 (high)	枯	*kū*, 'withered; dull'	6.47	1.90
	力	'strength'	4.97 (med.)	加	*jiā*, 'add'	5.09	4.61
low freq. radicals	女	'woman'	6.05 (high)	委	*wěi*, 'entrust'	4.95	2.55
	亠	'lid'	1.77 (low)	京	*jīng*, 'capital city'	2.00	3.11

Implications for future RT experiments of semantic priming:
- Part 1 yields semantic transparency (opacity) index for radicals
- Part 2 yields a character consistency index, i.e., an individual radical-to-character relatedness index, after partialling out other variables. Part 2 also yields semantic consistency index of radicals, e.g., of a given radical across characters [see top right].
- These can be used as covariates or controls for, and inform the design of, future priming-RT experiments

More sample index values for radical transparency, consistency, and radical-character relatedness (max.=7):

Rad.	Gloss	Rad. trans.	Rad. consistency	Char.	Gloss	Char.-rad. relation
口	'mouth'	6.23	4.41	吵	*chǎo*, 'quarrel'	6.64
				名	*míng*, 'call, name'	4.55
				吊	*dào*, 'hang'	3.83
木	'tree, wood'	6.47	4.58	枝	*zhī*, 'branch'	6.33
				休	*li*, 'chestnut'	4.02
				杰	*jié*, 'outstanding'	2.56
日	'sun; day'	6.30	4.50	星	*xīng*, 'star'	5.73
				旦	*dàn*, 'dawn'	4.05
				署	*shǔ*, 'govt. office'	3.33
糸	'silk'	4.94	3.64	繩	*shéng*, 'rope'	6.73
				紅	*hóng*, 'red'	3.44
				緩	*huǎn*, 'slow, delay'	2.16
匚	'enclosure'	1.70	2.15	區	*qū*, 'region'	1.45
				匹	*pǐ*, 'be equal'	1.44

- Also, non-semantic "meaningless bits" such as 丿, 丶, 乚 rated very low (<1.8)

Conclusions

- Factors relevant to radical and character opacity explored
- Independent effects of radical abstractness and cross-character consistency found
- Indices derived for radical transparency, radical-character relatedness, and radical consistency across characters
- Semantics and semantic variation can be quantified for future semantic priming experiments
- Results indicate some limitations of semantics in character processing (i.e., weak semantic correspondence)

Future analyses

- Further analysis with hierarchical non-linear modeling may yield better radical-character indices
- Multidimensional scaling and cluster analysis will be done to explore specific semantic dimensions relevant to mental representations of radicals and characters
- Influence of non-radical semantic components requries investigation