Neural dynamics of the compositional process in Chinese compound processing: Analyses on a mega-scale database

Tianqi Wang[1], Xu Xu[2]

[1] The University of Hong Kong [2] Shanghai Jiao Tong University

Know more about him at https://tianqi93.github.io/

Introduction

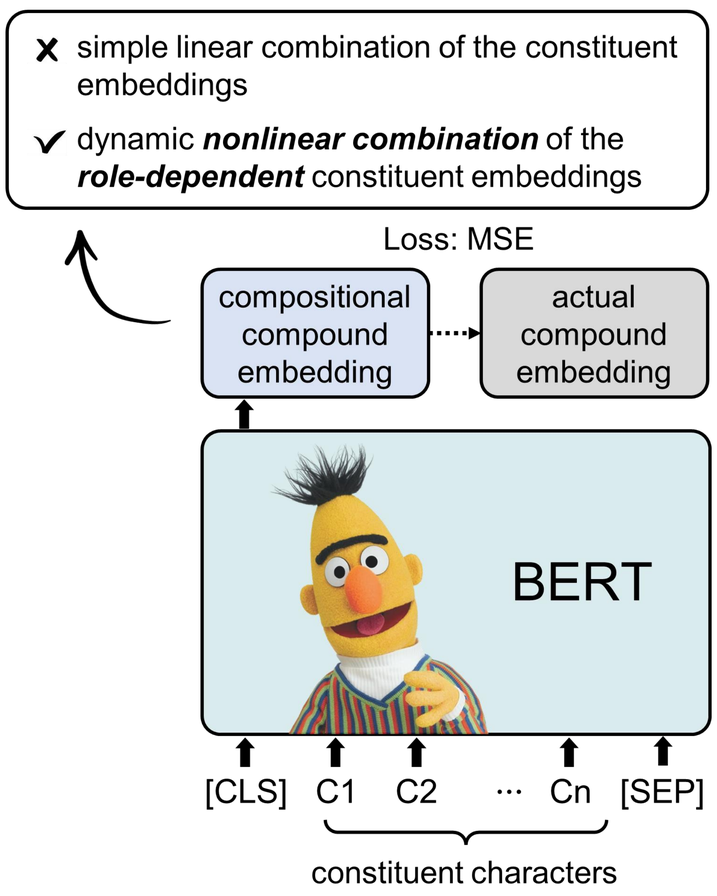
- Most Chinese words are formed by the combination of characters (e.g., 冰箱 refrigerator = 冰 ice + 箱 box)
- Characters are highly salient perceptual units, making morphological segmentation executed without effort
- Previous studies have shown that in Chinese compound recognition, a ***meaning-composition process*** is routinely implemented over the constituent characters
- It is unclear how this compositional process unfolds over time **[Research Question]**

SCAN TO READ OUR RELATED WORK

Method

Computational model

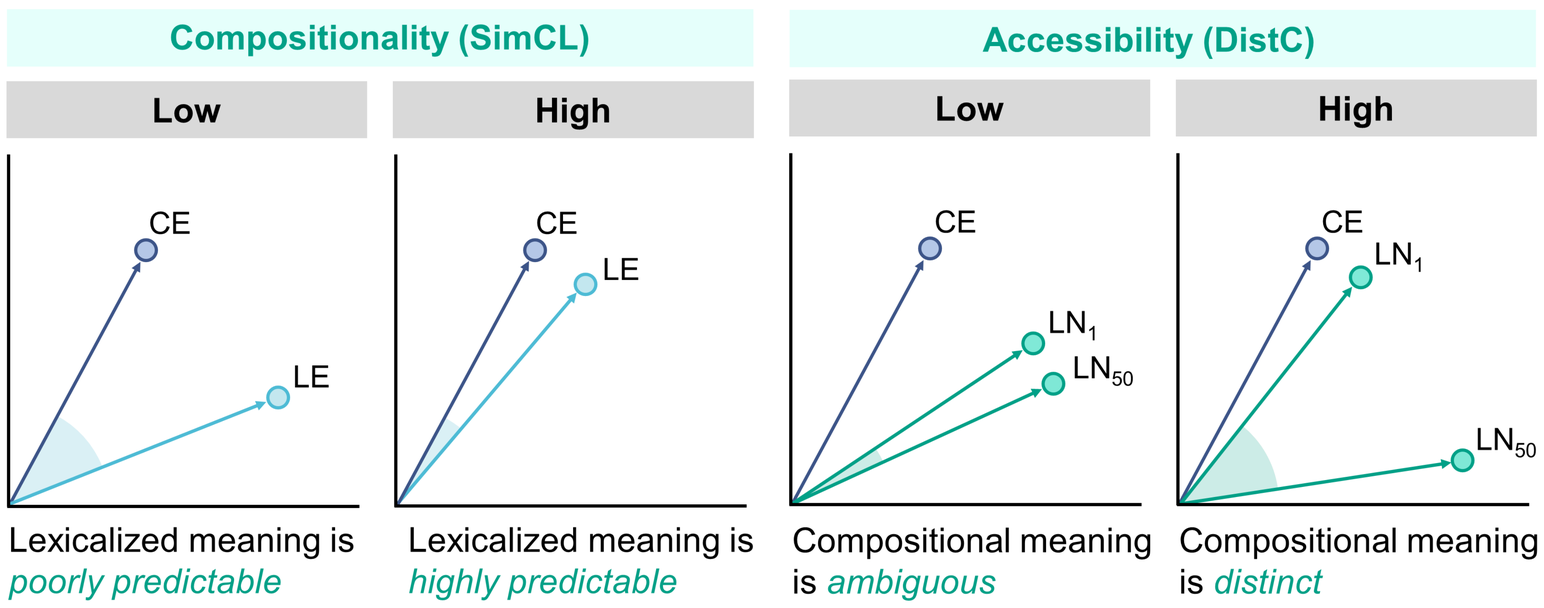
A transformer-based deep neural network is trained to optimally predict the actual compound embedding so as to acquire the ***compounding rules***



Computed metrics

Based on the compositional embeddings (i.e., the vectors generated by the computational model) and the actual embeddings of the compound words, two computed metrics are defined to characterize the end product of the compositional process

- **SimCL:** the cosine between the compositional and lexicalized (actual) compound embeddings
- **DistC:** the SD of the cosine distances of the 50 lexicalized neighbors that are closest to the compositional embedding



CE = compositional embedding LE = lexicalized embedding LN= lexicalized neighbor

Results

Analyses on the ERP megastudy of Chinese word recognition (Tsang & Zou, 2022)

- Higher ***distinctness*** was associated with smaller negativity at the intermediate stage of compound processing, attributable to the fewer number of lexicalized distractors and thus relative ease of semantic access
- Both higher ***distinctness*** and higher ***compositionality*** were associated with attenuated positivity at the later stage of compound processing, thanks to the ease of generating a holistic representation of the compound

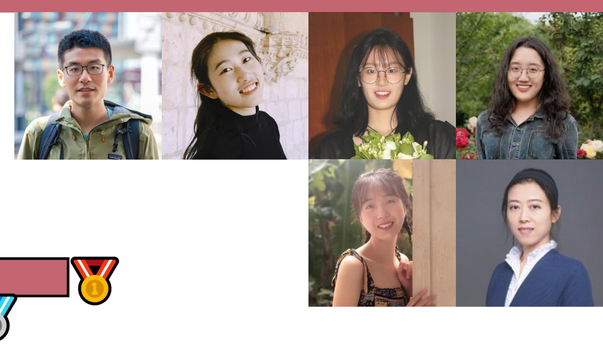
	TW1 0-100	TW2 100-200	TW3 200-300	TW4 300-400	TW5 400-500	TW6 500-600	TW7 600-700	TW8 700-800	TW9 800-900	TW10 900-1000
Intercept	-1.19	0.84	1.83	-1.27	3.15	5.71	4.43	2.57	1.74	1.03
Hemisphere:Right	-0.13	0.01	0.10	0.38	0.38	0.31	0.16	-0.01	-0.05	-0.01
Anteriority:Central	0.01	0.01	-0.01	-0.06	0.00	0.03	0.07	0.07	0.05	0.03
Anteriority:Posterior	-0.66	-1.65	-1.43	-0.26	-0.27	-0.39	-0.80	-1.19	-1.01	-0.70
NumStroke	0.00	-0.01	0.02	0.00	-0.01	-0.01	0.00	0.00	0.01	0.01
LogCD	-0.06	0.00	-0.02	0.29	0.39	0.13	-0.26	-0.26	-0.20	-0.08
C1.LogCD	0.13	0.09	0.14	0.13	-0.25	-0.40	-0.25	-0.06	-0.06	-0.01
C2.LogCD	0.07	0.06	0.22	0.21	-0.20	-0.28	-0.15	-0.09	-0.05	-0.03
C1.NoH	-0.06	-0.06	-0.03	-0.01	0.00	-0.10	-0.22	-0.08	0.04	-0.06
C2.NoH	0.01	0.07	0.01	0.02	0.00	-0.10	-0.13	-0.02	-0.02	0.08
C1.ST	-0.13	-0.07	-0.02	0.00	0.07	-0.06	-0.16	-0.15	-0.07	-0.01
C2.ST	0.03	0.17	0.14	0.19	0.13	-0.07	-0.16	0.00	-0.09	0.04
SimCL	0.90	0.87	0.11	-0.46	-1.35	-1.29	-0.19	0.72	0.39	0.19
DistC	-6.32	-6.51	5.84	7.75	16.00	3.52	-9.68	-11.59	-1.61	2.85

Notes. The numbers refer to the fixed effects estimates. Shaded cells = corrected $p < 0.050$. TW = time window, NumStroke = number of strokes, CD = contextual diversity, NoH = number of homophones, ST = semantic transparency, C1 = first character, C2 = second character. The six regions were dummy-coded.

A primed lexical decision megastudy for Chinese characters

Tianqi Wang[12], Leqi Zhou[1], Hongyu Wang[1], Weining Liu[1], Yuqing Ding[1], Xu Xu[1]

[1] Shanghai Jiao Tong University [2] The University of Hong Kong

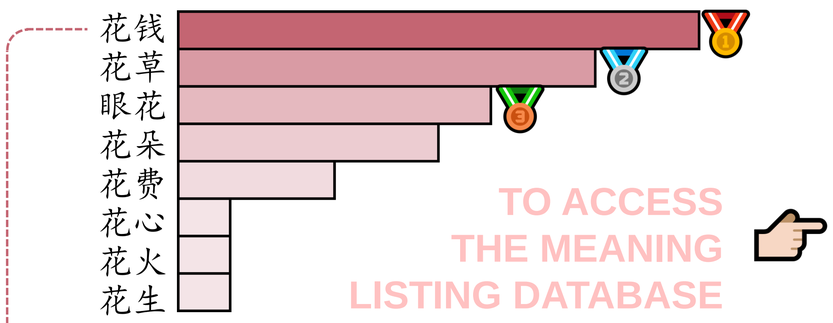


How character recognition is influenced by word-level representations

Stimuli

- **1,864** target characters
- Prime-target pairs in the ***priming*** condition were constructed from the most frequent word formations as collected in a meaning listing study (Chen et al., 2024)
- Primes in the ***baseline*** condition were not any of the top 5 character associate of the target

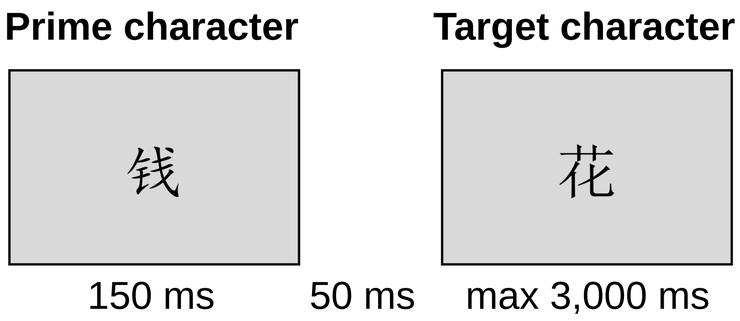
The target character is 花



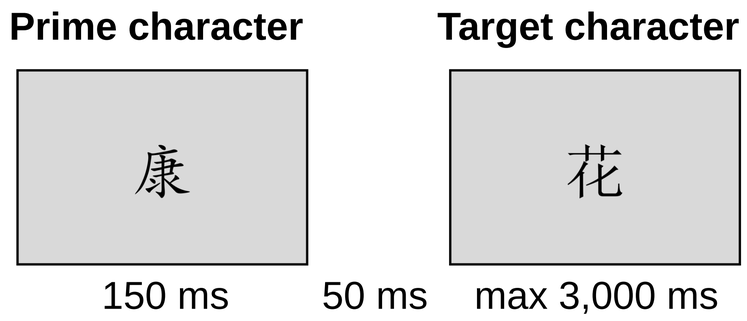
Participants

- **145** participants (101 females, mean age = 22.97±1.94)

Priming



Baseline



Information available in the database

Properties at the character level

	Priming		Baseline	
Character frequency	PRM.Prime.Freq	Target.Freq	BSL.Prime.Freq	Target.Freq
Number of strokes	PRM.Prime.NumStroke	Target.NumStroke	BSL.Prime.NumStroke	Target.NumStroke
Number of word formations	PRM.Prime.NWF	Target.NWF	BSL.Prime.NWF	Target.NWF
Number of meanings	PRM.Prime.NoM	Target.NoM	BSL.Prime.NoM	Target.NoM
Number of pronunciations	PRM.Prime.NoP	Target.NoP	BSL.Prime.NoP	Target.NoP

Properties at the word level

Prime position

Prime-target associative strength — Frequency of 🥇 to the total frequency of responses to the target character

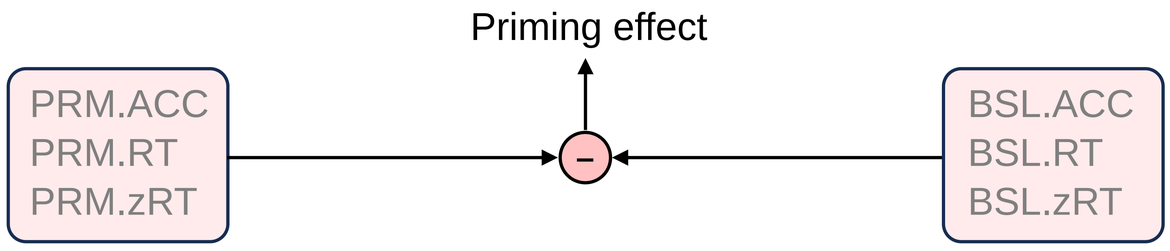
Frequency of the word formed

Response measures

Accuracy rate

Response time

Standardized response time



Notes. PRM = priming condition, BSL = baseline condition

Item-level analyses

	Priming effect zRT		Priming effect ACC	
	Beta	*t*	Beta	*t*
Step 1				
Target.LogFreq	-0.047	-2.868 **	-0.010	-3.405 ***
Target.NumStroke	-0.005	-2.584 **	0.000	-1.226
Target.LogNWF	-0.074	-2.808 **	-0.016	-3.516 ***
Target.LogNoM	0.021	0.490	0.017	2.290 *
Target.LogNoP	-0.054	-0.698	-0.028	-2.079 *
PRM.Prime.LogFreq	0.001	0.064	-0.001	-0.262
PRM.Prime.NumStroke	0.000	-0.157	0.000	-1.201
PRM.Prime.LogNWF	-0.052	-2.098 *	-0.003	-0.642
PRM.Prime.LogNoM	-0.008	-0.226	-0.006	-0.964
PRM.Prime.LogNoP	-0.054	-0.694	0.031	2.311 *
BSL.Prime.LogFreq	0.011	0.793	0.003	1.431
BSL.Prime.NumStroke	0.002	0.783	0.001	2.248 *
BSL.Prime.LogNWF	-0.039	-1.715	-0.003	-0.776
BSL.Prime.LogNoM	-0.025	-0.699	0.005	0.842
BSL.Prime.LogNoP	0.023	0.304	0.008	0.596
R^2 change	0.107		0.111	
Step 2				
Word.LogFreq	-0.014	-1.325	-0.001	-0.411
R^2 change	0.000		0.000	
Step 3				
Prime.Position	-0.082	-5.391 ***	-0.003	-1.239
PrimeTarget.Association	0.137	2.250 *	0.032	3.057 **
R^2 change	0.017		0.005	

Trial-level analyses

	zRT		ACC	
	Estimate	*t*	Estimate	*t*
Target.LogFreq	-0.171	-15.366 ***	0.315	6.926 ***
Target.NumStroke	0.017	9.687 ***	-0.016	-2.246 ***
Target.LogNWF	-0.210	-10.263 ***	0.630	7.445 ***
Target.LogNoM	0.036	0.990	-0.086	0.567
Target.LogNoP	0.065	0.981	0.113	0.409
Prime.LogFreq	0.044	4.617 ***	-0.210	-4.893 ***
Prime.NumStroke	0.001	0.897	-0.011	-1.569
Prime.LogNWF	-0.013	-0.803	0.088	1.218
Prime.LogNoM	-0.030	-1.118	-0.138	-1.184
Prime.LogNoP	0.023	0.420	0.075	0.305
Condition	-0.179	-40.152 ***	0.436	15.946 ***
Block.Order	-0.021	-32.129 ***	0.017	4.577 ***

- In Chinese, word formations establish the contexts in which characters cooccur with each other
- Facilitation in the priming condition may come from the automatic spreading activation between associated characters
- Chinese readers are sensitive to the position of characters within words formed