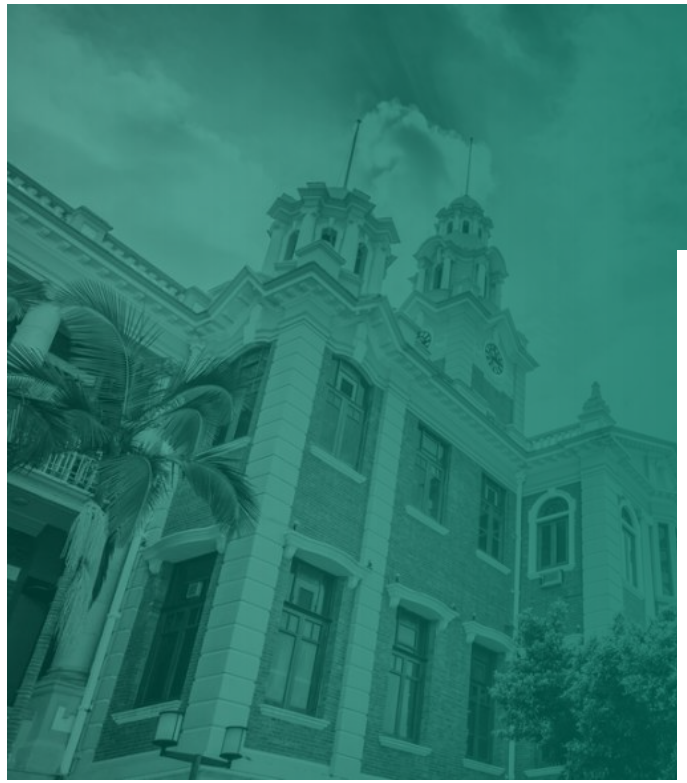
RESEARCH TALK



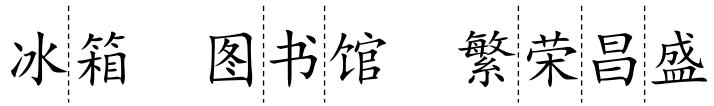
Composition as nonlinear combination in semantic space: A computational characterization of compound processing

Tianqi Wang
The University of Hong Kong

tianqi93@connect.hku.hk
https://tianqi93.github.io/

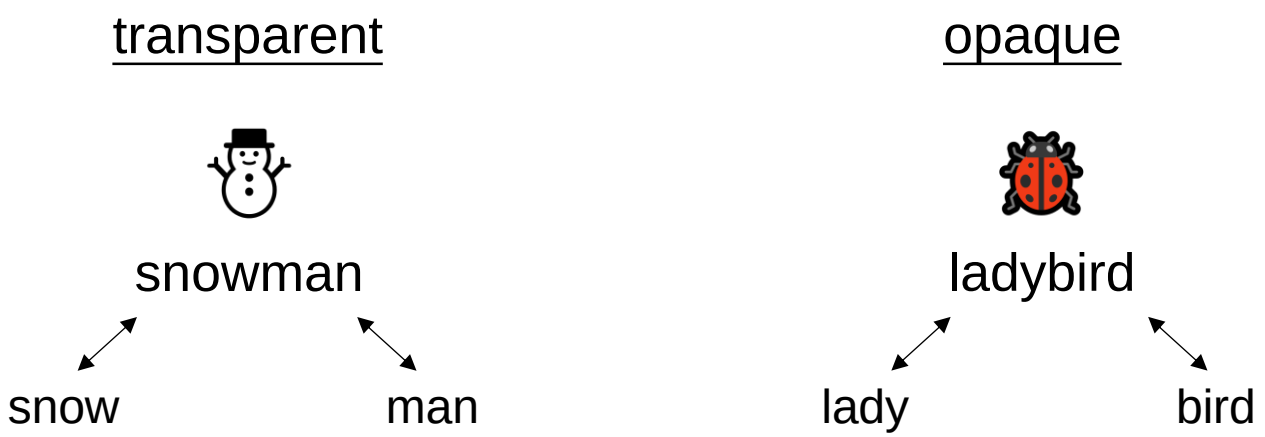
Introduction

- 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 [Tsang et al., 2018; Tse et al., 2017]



Introduction

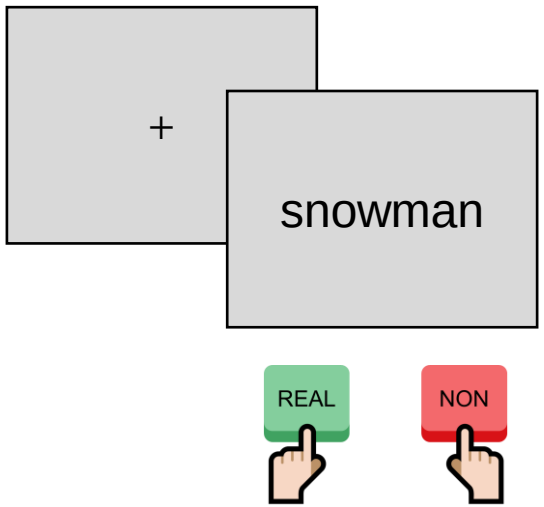
- Most Chinese words are formed by the combination of characters (e.g., 冰箱 refrigerator = 冰 ice + 箱 box)
- Characters are highly salient perceptual units, making morphological segmentation executed without effort [Tsang et al., 2018; Tse et al., 2017]
- The role played by constituents in compound processing has been studied via *semantic transparency* (ST), which produced inconsistent results [Libben et al., 2003; Monsell, 1985; Sandra, 1990]



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 [Tsang et al., 2018; Tse et al., 2017]
- The role played by constituents in compound processing has been studied via *semantic transparency* (ST), which produced inconsistent results [Libben et al., 2003; Monsell, 1985; Sandra, 1990]

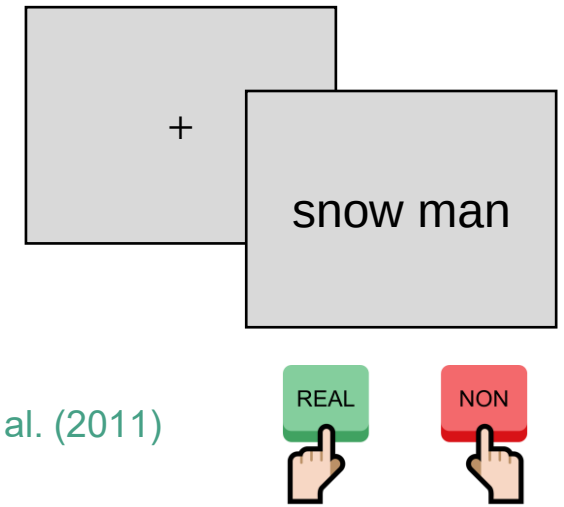
lexical decision task



Libben et al. (2003)

↑ ST ⇒ faster response?

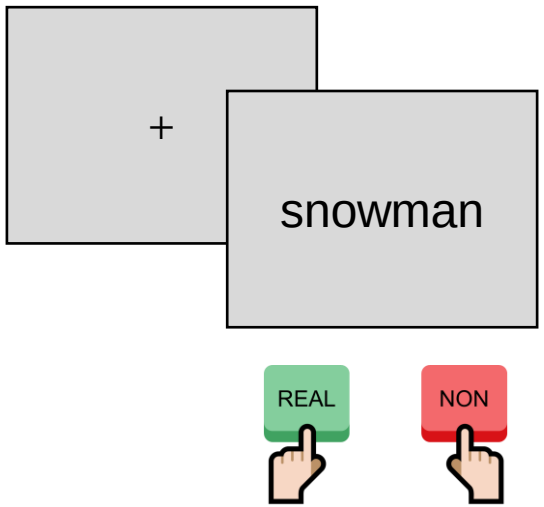
Ji et al. (2011)



Introduction

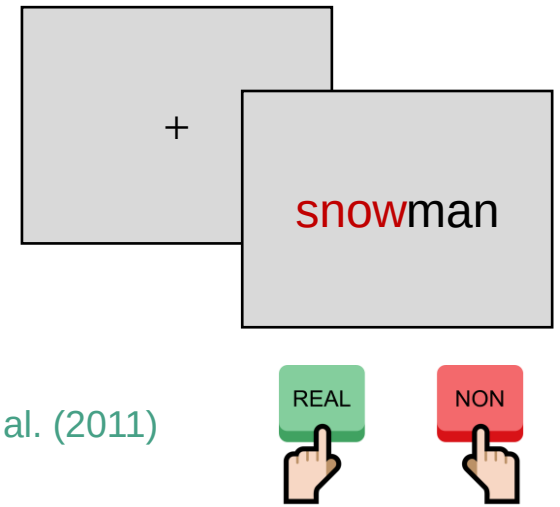
- Most Chinese words are formed by the combination of characters (e.g., 冰箱 refrigerator = 冰 ice + 箱 box)
- Characters are highly salient perceptual units, making morphological segmentation executed without effort [Tsang et al., 2018; Tse et al., 2017]
- The role played by constituents in compound processing has been studied via *semantic transparency* (ST), which produced inconsistent results [Libben et al., 2003; Monsell, 1985; Sandra, 1990]

lexical decision task



↑ST ⇒ faster response?

Libben et al. (2003)



Ji et al. (2011)

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 [Tsang et al., 2018; Tse et al., 2017]
- The role played by constituents in compound processing has been studied via *semantic transparency* (ST), which produced inconsistent results [Libben et al., 2003; Monsell, 1985; Sandra, 1990]
- Psycholinguists started to reconceptualize ST from the *compositional perspective*

✔ The *predictability* of the compound meaning given the *combination* of the constituents' meanings

Marelli & Luzzatti (2012)

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 [Tsang et al., 2018; Tse et al., 2017]
- The role played by constituents in compound processing has been studied via *semantic transparency* (ST), which produced inconsistent results [Libben et al., 2003; Monsell, 1985; Sandra, 1990]
- Psycholinguists started to reconceptualize ST from the *compositional perspective*
- How does this *combinatorial process* modulate Chinese compound processing? **[RQ]**

✔ The *predictability* of the compound meaning given the *combination* of the constituents' meanings

Marelli & Luzzatti (2012)

How to derive a meaning representation for the compositional compound meaning?

A computational model to derive compositional compound meaning

English

Compounding rules

	modifier		head
airport	air	+	port
snowman	snow	+	man

Computational model

modifier matrix *M* ***u*** **head matrix *H*** ***v*** ***c***

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

compositional

$$M \cdot \begin{matrix} \textit{girl} \\ \textit{tea} \\ \vdots \\ \textit{milk} \end{matrix} + H \cdot \begin{matrix} \textit{friend} \\ \textit{bag} \\ \vdots \\ \textit{man} \end{matrix} = \begin{matrix} \textit{girlfriend} \\ \textit{teabag} \\ \vdots \\ \textit{milkman} \end{matrix}$$

CAOSS framework (Marelli et al., 2017)

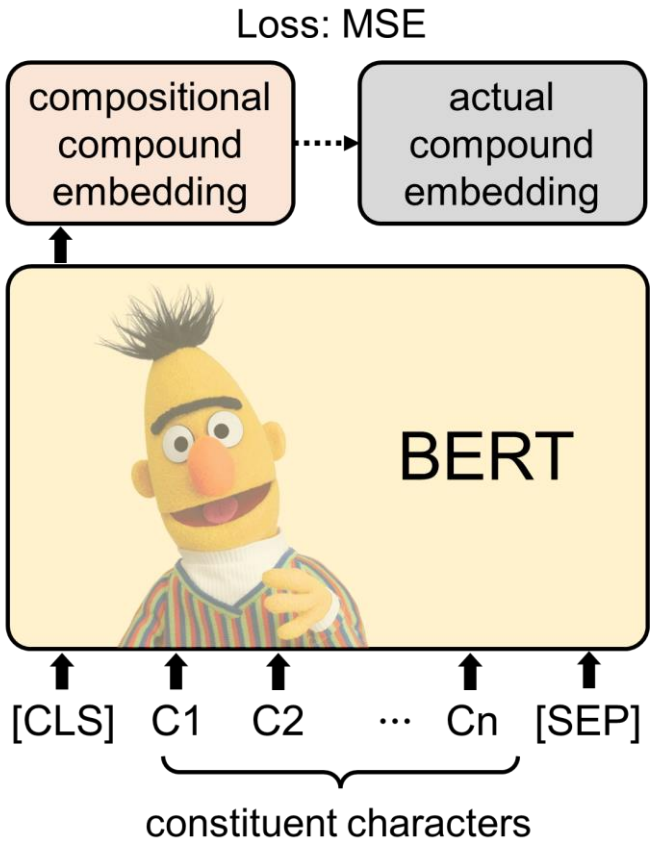
A computational model to derive compositional compound meaning

Chinese

Compounding rules

	C1	C2		
subordinate	黑	板	modifier	head
coordinate	城	市	head	head
verb-object	扫	地	Not clear	
verb-resultative	说	明		
subject-predicate	晚	安		

Computational model



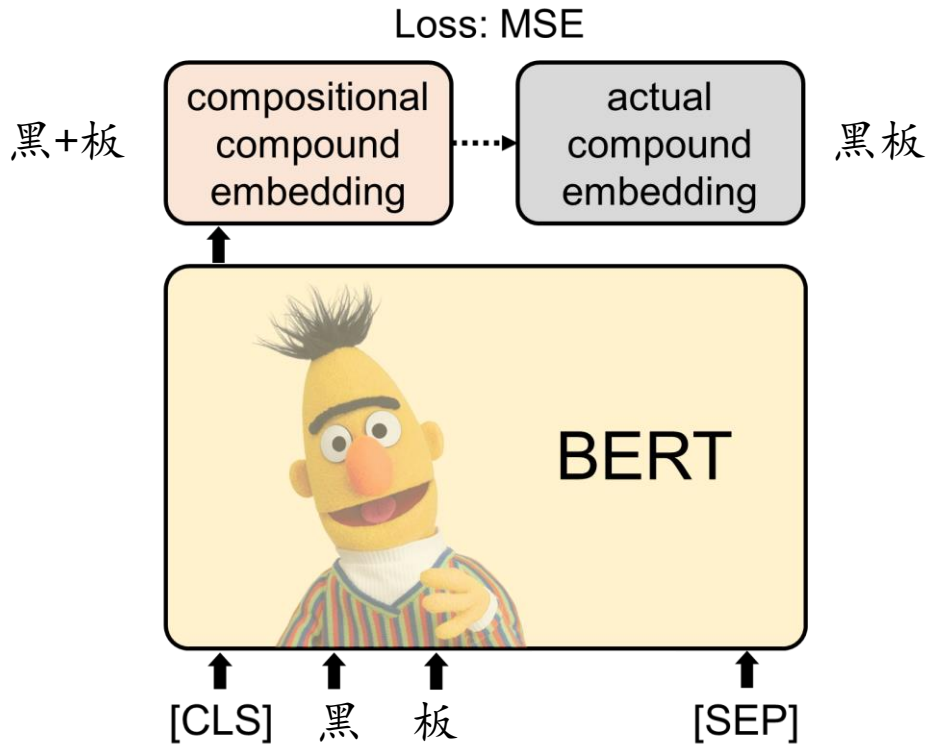
A computational model to derive compositional compound meaning

Chinese

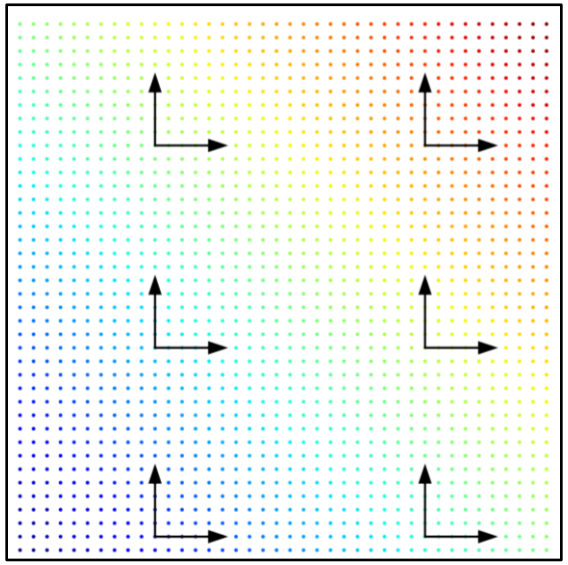
Compounding rules

	C1	C2		
subordinate	黑	板	modifier	head
coordinate	城	市	head	head
verb-object	扫	地	Not clear	
verb-resultative	说	明		
subject-predicate	晚	安		

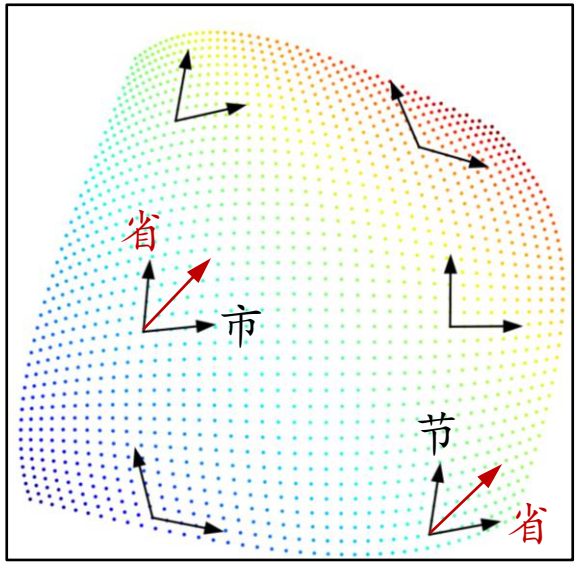
Computational model



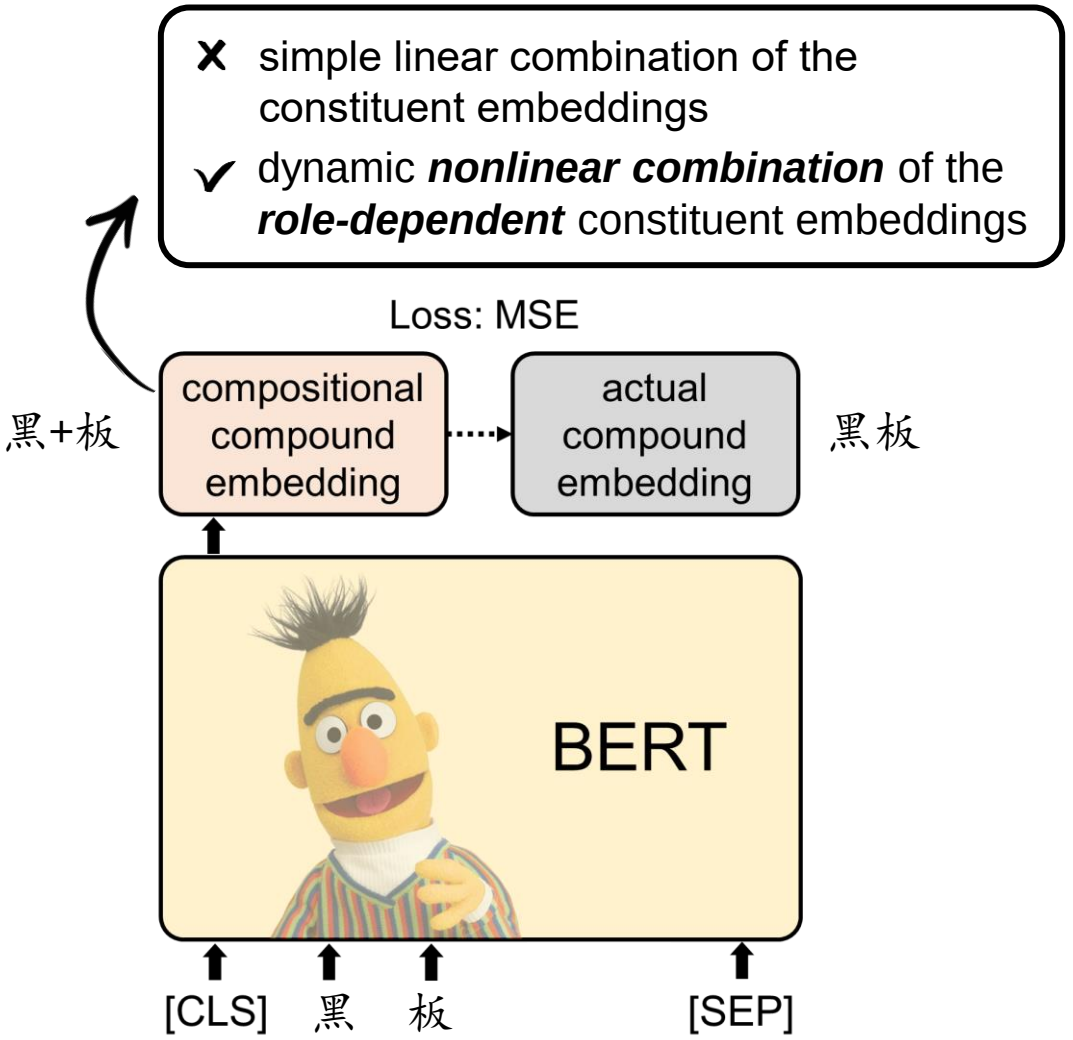
A computational model to derive compositional compound meaning



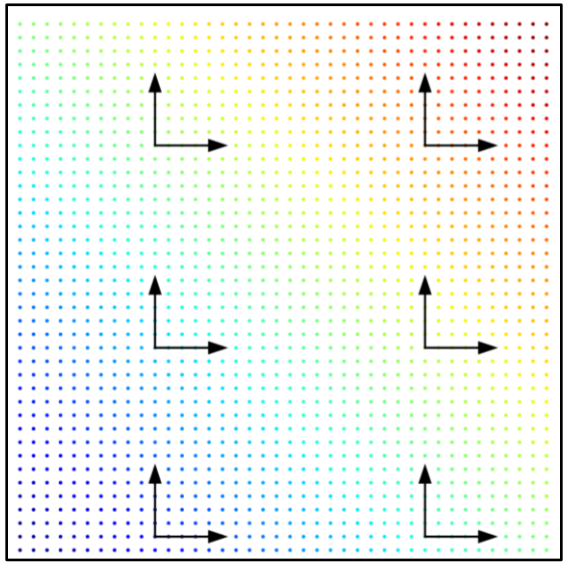
linear combination



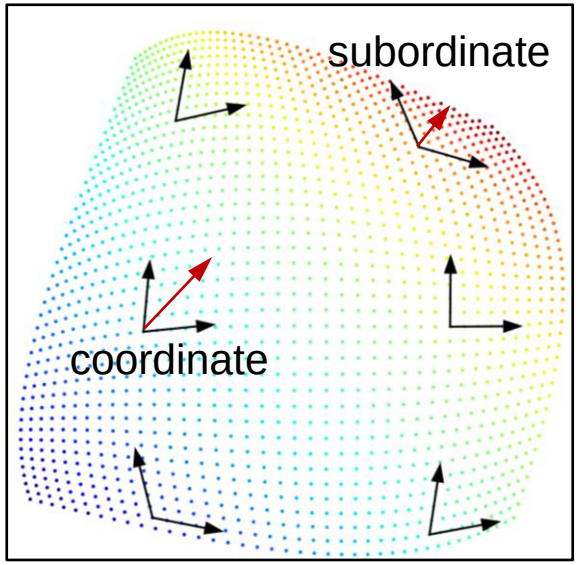
nonlinear combination



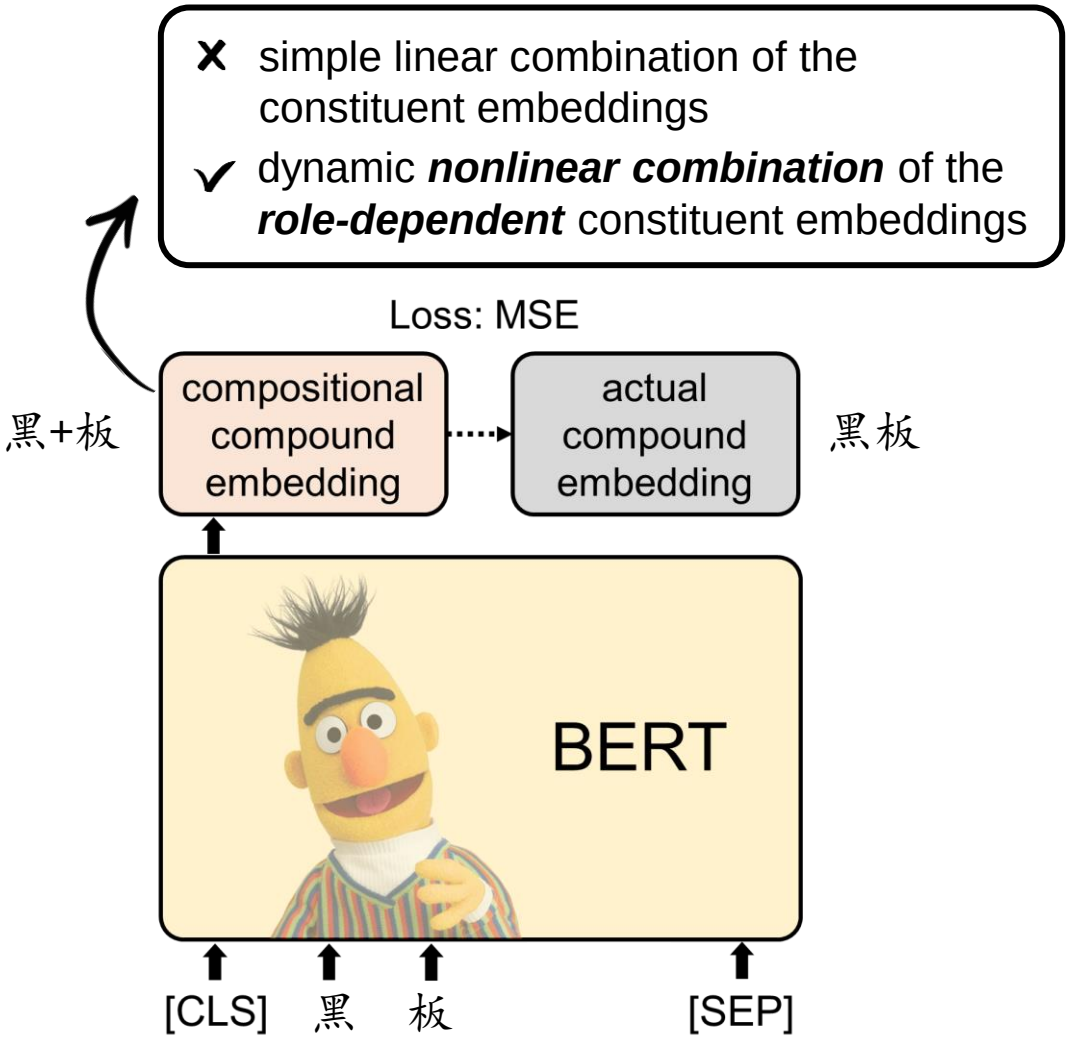
A computational model to derive compositional compound meaning



linear combination

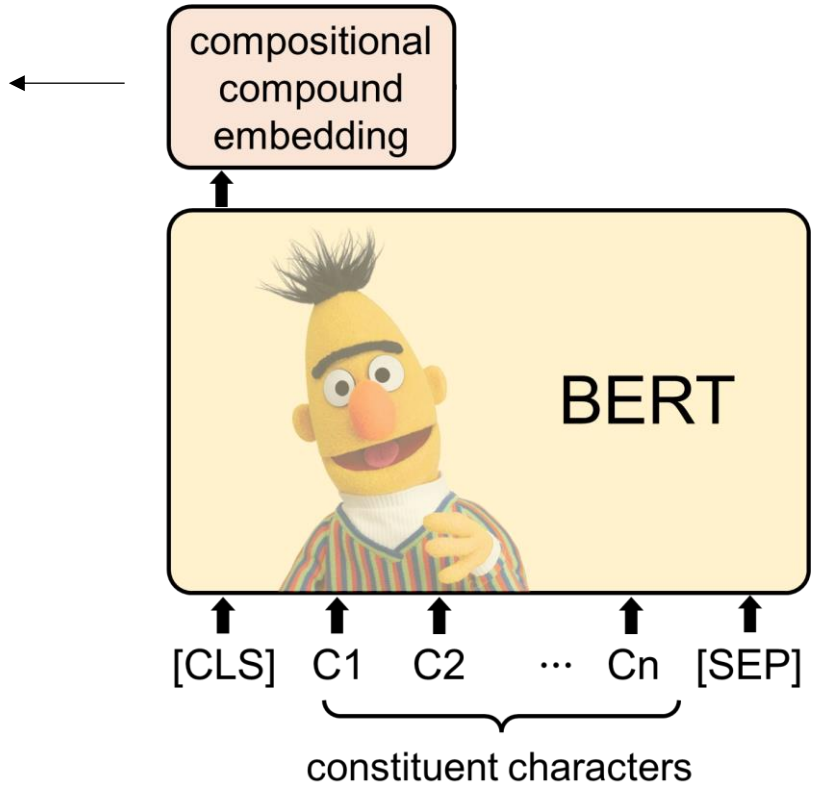


nonlinear combination

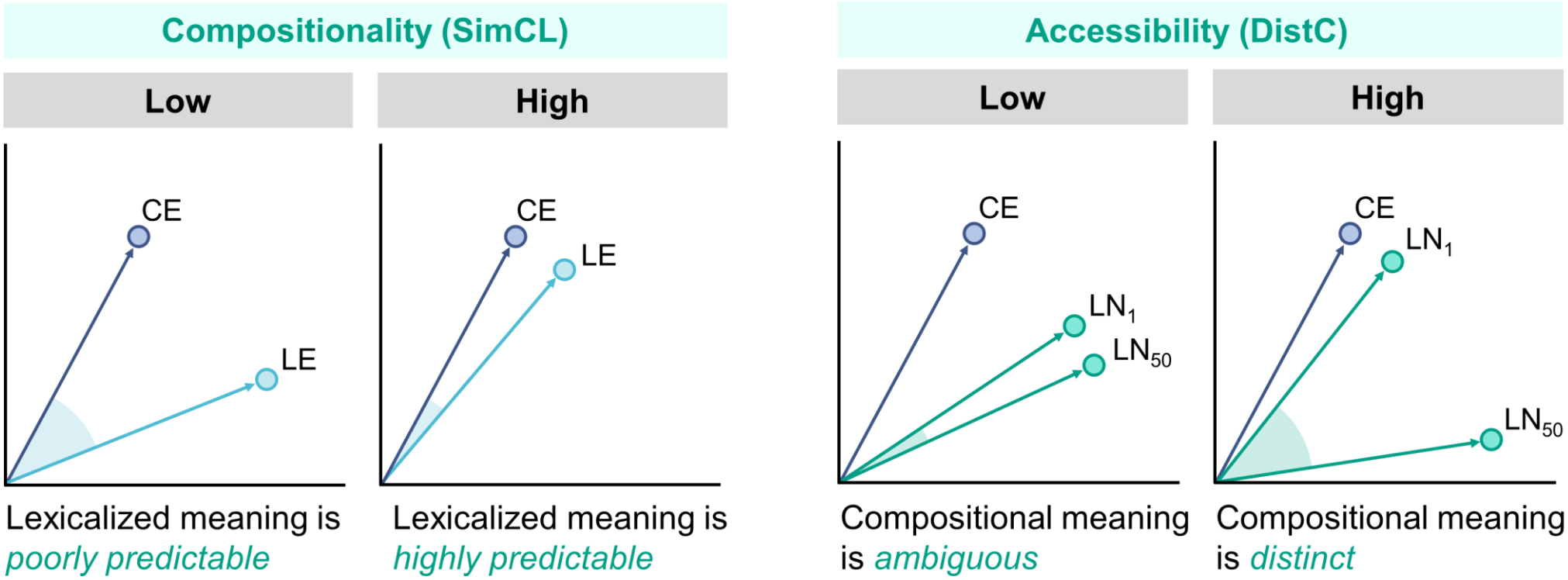


A computational model to derive compositional compound meaning

The attributes of the compositional meaning representation



Computed metrics for attributes of the compositional embedding



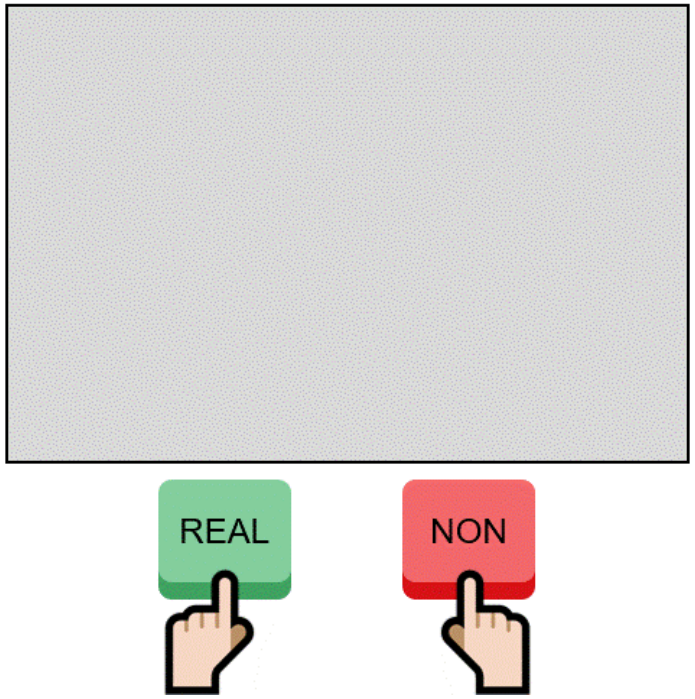
CE = compositional embedding **LE** = lexicalized embedding **LN** = lexicalized neighbor

开方 乐器 拂晓 复杂

Effects of the computed metrics on compound processing

Megastudy of Lexical Decision in Simplified Chinese

10,022 two-character words



Tsang et al. (2018)

Database of eye-movement measures on words in Chinese reading

6,128 two-character words

Zhang et al. (2022)

Effects of the computed metrics on compound processing

- Analytical procedure

Baseline model

Computed metrics

Interactions

Response	~	LogWF		Word frequency
		+ NumStroke		Number of strokes
		+ C1.LogCF	+ C2.LogCF	Character frequency
		+ C1.LogFS	+ C2.LogFS	Family size
		+ C1.LogNoM	+ C2.LogNoM	Number of meanings
		+ C1.LogNoP	+ C2.LogNoP	Number of pronunciations
		+ C1.ST	+ C2.ST	Semantic transparency
		+ SimCL		
		+ DistC		
		+ Interactions		
		+ (1|C1)	+ (1|C2)	

C1 = first character C2 = second character

Effects of the computed metrics on compound processing

- Lexical decision of real words — higher compositionality → faster and more accurate response

	zRT					ERR				
	Estimate	*t*	*p*	% ΔR^2	R^2	Estimate	*t*	*p*	% ΔR^2	R^2
Intercept	-0.33	-101.94	< 0.001			3.29	75.66	< 0.001		
LogWF	-0.20	-65.60	< 0.001			-2.05	-39.28	< 0.001		
NumStroke	0.004	6.31	< 0.001							
C1.LogCF	0.03	5.46	< 0.001			0.35	4.67	< 0.001		
C2.LogCF	0.03	4.58	< 0.001			0.52	6.29	< 0.001		
C1.LogFS	-0.06	-6.65	< 0.001			-0.50	-3.55	< 0.001		
C2.LogFS	-0.06	-6.52	< 0.001			-0.67	-4.82	< 0.001		
C2.LogNoM	0.05	3.72	< 0.001							
C1.LogNoP	0.08	2.46	0.014							
C1.ST	-0.01	-2.47	0.014			-0.37	-4.62	< 0.001		
C2.ST						-0.23	-2.93	0.003		
Baseline					0.425					0.185
SimCL	-0.40	-8.14	< 0.001	1.97	0.434	-4.18	-5.36	< 0.001	0.74	0.186
DistC	-0.95	-4.41	< 0.001	0.33	0.435					
SimCL × LogWF	0.21	4.14	< 0.001	0.30	0.436	8.22	9.26	< 0.001	5.55	0.196
SimCL × C1.LogFS						-6.00	-3.54	< 0.001	0.60	0.198
DistC × C2.LogCF	0.58	2.47	0.014	0.07	0.437					
Composition metrics				2.69	0.437				6.97	0.198

Effects of the computed metrics on compound processing

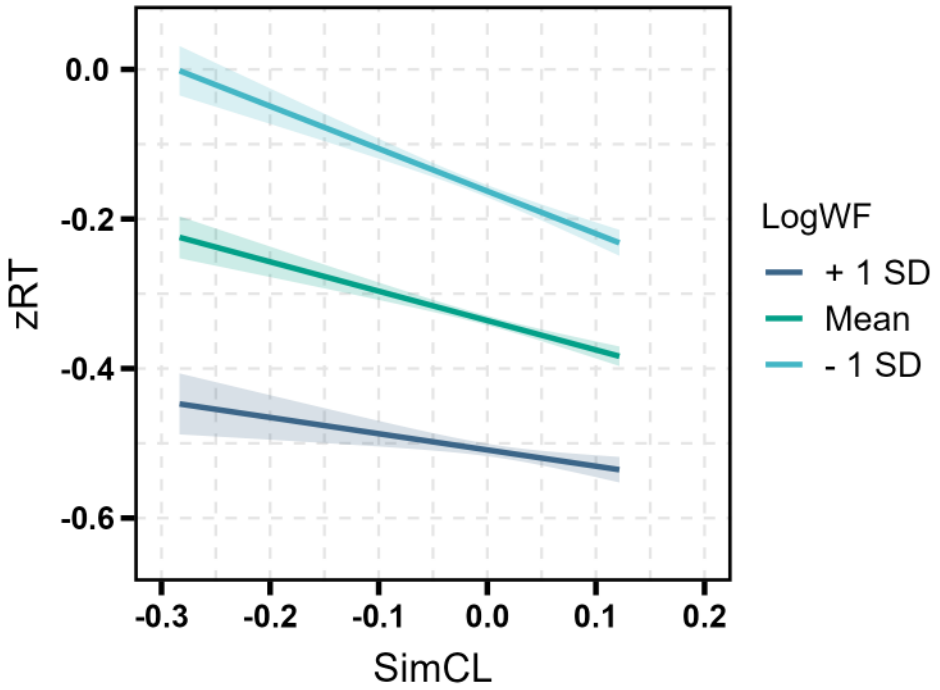
- Lexical decision of real words higher distinctness → faster response

	zRT					ERR				
	Estimate	*t*	*p*	% ΔR^2	R^2	Estimate	*t*	*p*	% ΔR^2	R^2
Intercept	-0.33	-101.94	< 0.001			3.29	75.66	< 0.001		
LogWF	-0.20	-65.60	< 0.001			-2.05	-39.28	< 0.001		
NumStroke	0.004	6.31	< 0.001							
C1.LogCF	0.03	5.46	< 0.001			0.35	4.67	< 0.001		
C2.LogCF	0.03	4.58	< 0.001			0.52	6.29	< 0.001		
C1.LogFS	-0.06	-6.65	< 0.001			-0.50	-3.55	< 0.001		
C2.LogFS	-0.06	-6.52	< 0.001			-0.67	-4.82	< 0.001		
C2.LogNoM	0.05	3.72	< 0.001							
C1.LogNoP	0.08	2.46	0.014							
C1.ST	-0.01	-2.47	0.014			-0.37	-4.62	< 0.001		
C2.ST						-0.23	-2.93	0.003		
Baseline					0.425					0.185
SimCL	-0.40	-8.14	< 0.001	1.97	0.434	-4.18	-5.36	< 0.001	0.74	0.186
DistC	-0.95	-4.41	< 0.001	0.33	0.435					
SimCL × LogWF	0.21	4.14	< 0.001	0.30	0.436	8.22	9.26	< 0.001	5.55	0.196
SimCL × C1.LogFS						-6.00	-3.54	< 0.001	0.60	0.198
DistC × C2.LogCF	0.58	2.47	0.014	0.07	0.437					
Composition metrics				2.69	0.437				6.97	0.198

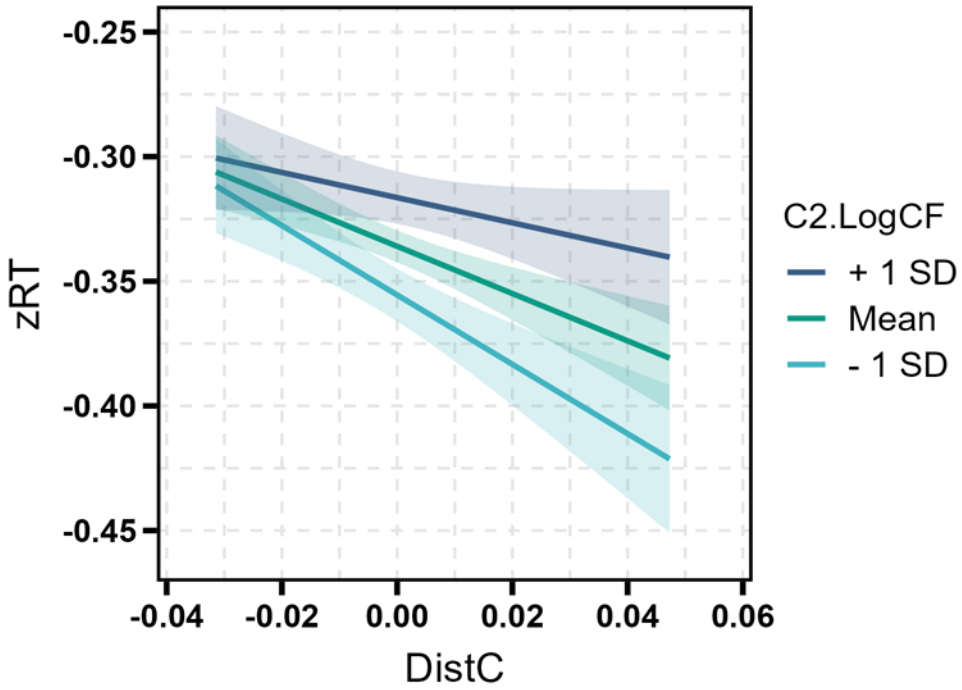
Effects of the computed metrics on compound processing

- Lexical decision of real words

Interaction between word frequency and compositionality



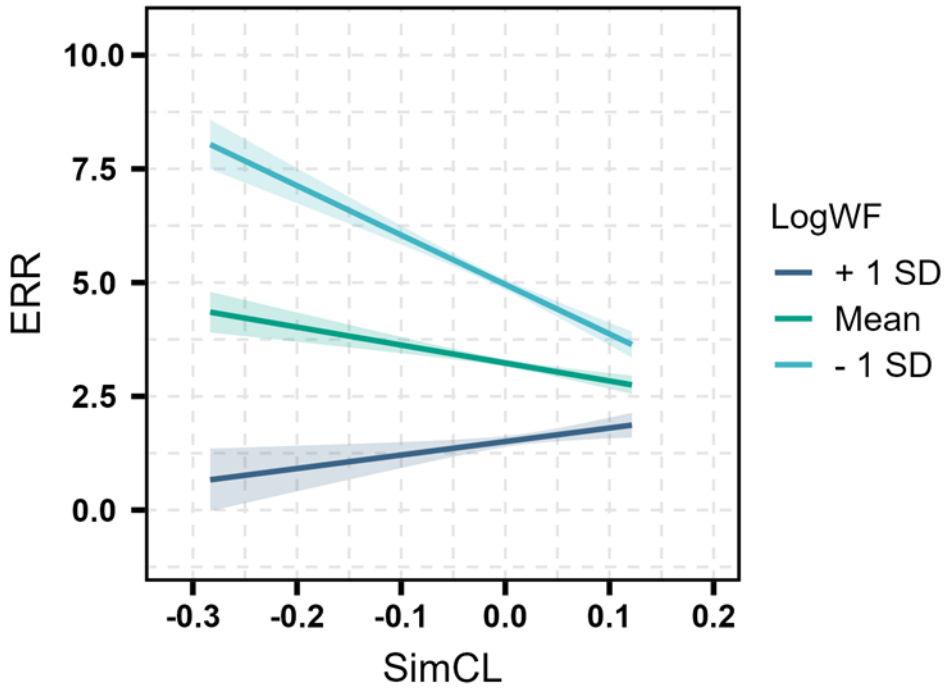
Interaction between character frequency and distinctness



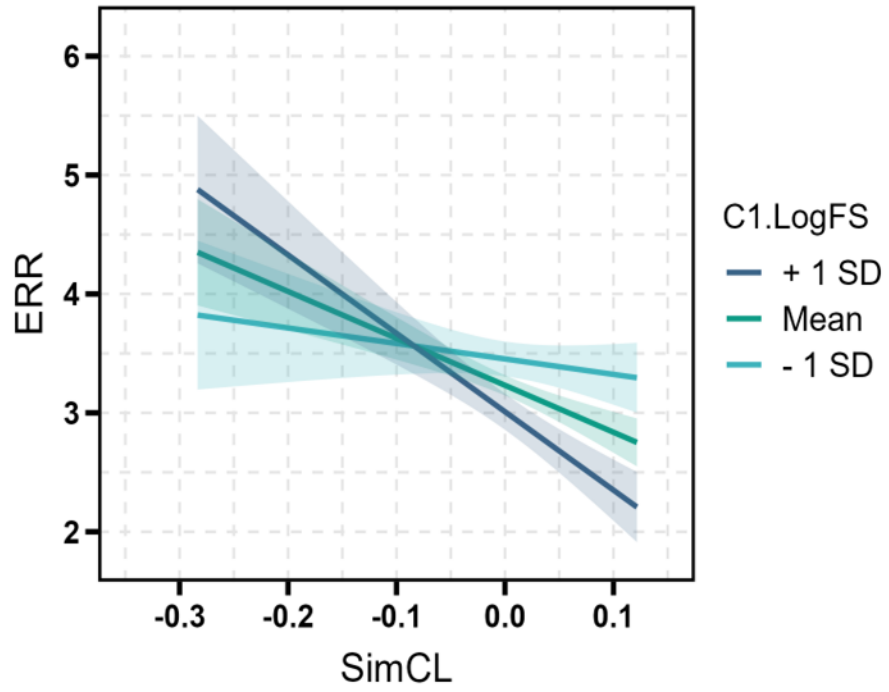
Effects of the computed metrics on compound processing

- Lexical decision of real words

Interaction between word frequency and compositionality



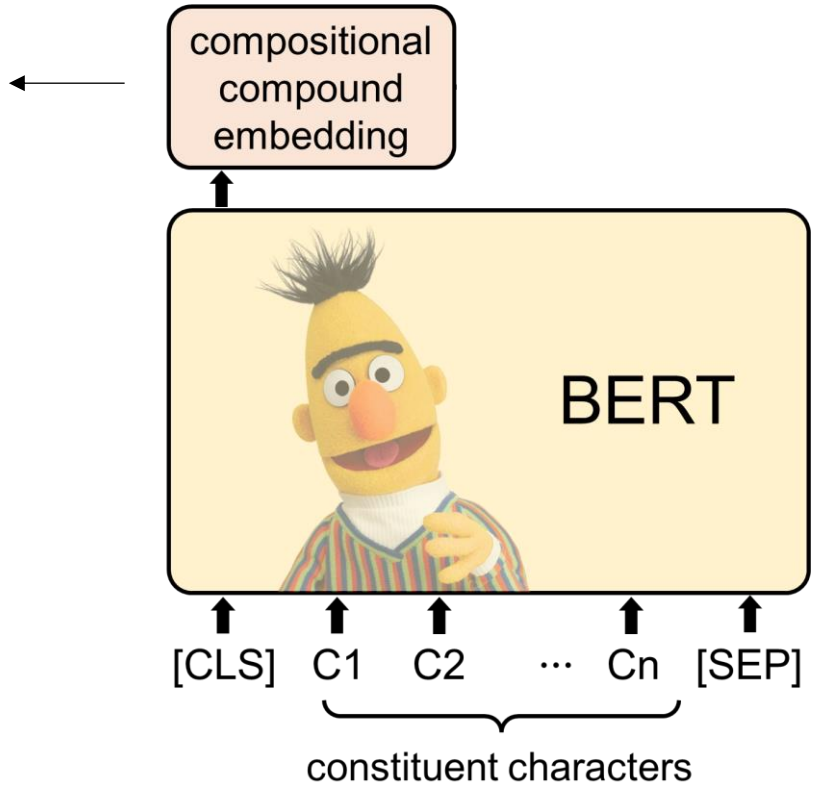
Interaction between character frequency and compositionality



Effects of the computed metrics on compound processing

distinctness

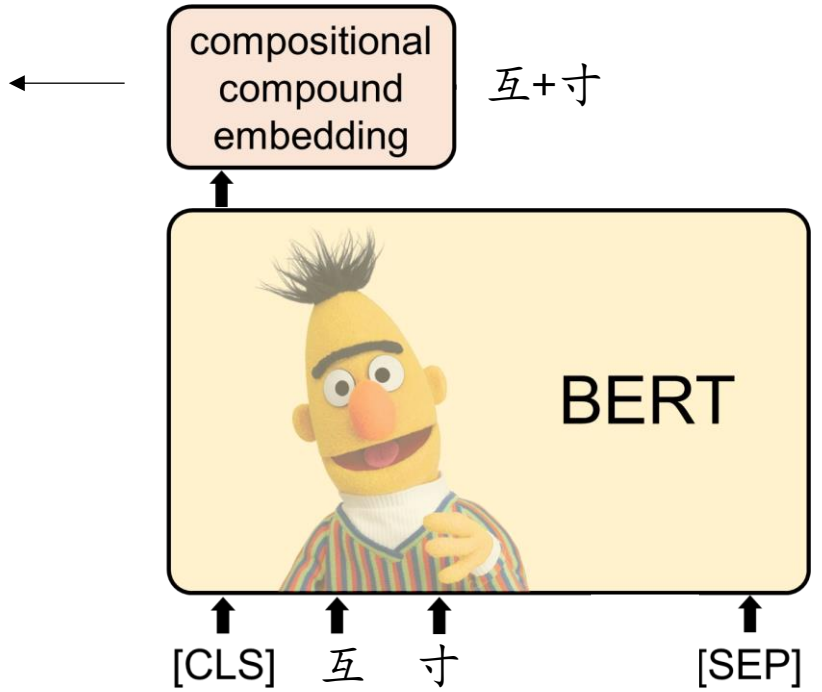
compositional meaning representation of nonwords



Effects of the computed metrics on compound processing

distinctness

compositional meaning representation of nonwords



Effects of the computed metrics on compound processing

- Lexical decision of nonwords — higher distinctness → slower and less accurate response

	zRT					ERR				
	Estimate	*t*	*p*	% ΔR^2	R^2	Estimate	*t*	*p*	% ΔR^2	R^2
Intercept	0.28	71.25	< 0.001			9.52	101.21	< 0.001		
NumStroke	0.01	12.12	< 0.001			0.08	4.14	< 0.001		
C1.LogCF	-0.05	-9.23	< 0.001			-1.22	-8.93	< 0.001		
C2.LogCF	-0.02	-4.07	< 0.001			-0.77	-5.17	< 0.001		
C1.LogFS	0.11	10.33	< 0.001			3.18	11.01	< 0.001		
C2.LogFS	0.13	12.56	< 0.001			4.26	15.75	< 0.001		
C1.LogNoM	0.05	3.11	0.002			2.27	5.24	< 0.001		
Baseline					0.074					0.091
DistC	6.90	12.26	< 0.001	17.51	0.087	295.38	18.72	< 0.001	36.05	0.124
DistC × C2.LogCF	-2.08	-3.17	0.002	0.80	0.088					
Composition metrics				18.45	0.088				36.05	0.124

寸互 vs 微哭

Effects of the computed metrics on compound processing

- Eye-tracking data

higher compositionality → shorter first fixation duration

	First fixation duration					Total fixation duration				
	Estimate	*t*	*p*	% ΔR^2	R^2	Estimate	*t*	*p*	% ΔR^2	R^2
Intercept	261.39	595.50	< 0.001			416.35	242.40	< 0.001		
LogWF	-8.07	-17.36	< 0.001			-29.09	-16.55	< 0.001		
NumStroke	0.71	7.60	< 0.001			1.93	5.63	< 0.001		
C2.LogCF	2.02	3.33	< 0.001			11.60	4.43	< 0.001		
C2.LogNoM						-18.85	-2.34	0.019		
Baseline					0.086					0.082
SimCL	-28.48	-3.60	< 0.001	2.76	0.088	-113.43	-3.80	< 0.001	4.63	0.086
DistC						-250.69	-2.10	0.036	0.80	0.086
DistC × LogWF						301.43	2.50	0.012	1.52	0.088
Composition metrics				2.76	0.088				7.07	0.088

Effects of the computed metrics on compound processing

- Eye-tracking data

higher compositionality → shorter total fixation duration

	First fixation duration					Total fixation duration				
	Estimate	*t*	*p*	% ΔR^2	R^2	Estimate	*t*	*p*	% ΔR^2	R^2
Intercept	261.39	595.50	< 0.001			416.35	242.40	< 0.001		
LogWF	-8.07	-17.36	< 0.001			-29.09	-16.55	< 0.001		
NumStroke	0.71	7.60	< 0.001			1.93	5.63	< 0.001		
C2.LogCF	2.02	3.33	< 0.001			11.60	4.43	< 0.001		
C2.LogNoM						-18.85	-2.34	0.019		
Baseline					0.086					0.082
SimCL	-28.48	-3.60	< 0.001	2.76	0.088	-113.43	-3.80	< 0.001	4.63	0.086
DistC						-250.69	-2.10	0.036	0.80	0.086
DistC × LogWF						301.43	2.50	0.012	1.52	0.088
Composition metrics				2.76	0.088				7.07	0.088

Effects of the computed metrics on compound processing

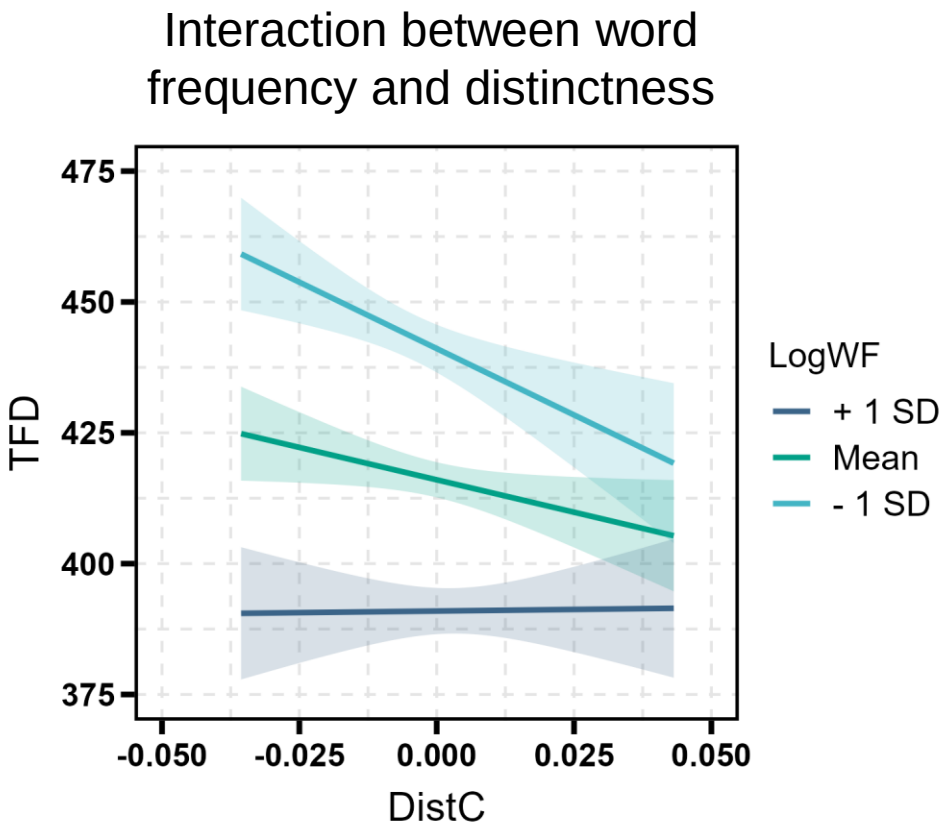
- Eye-tracking data

higher distinctness → shorter total fixation duration

	First fixation duration					Total fixation duration				
	Estimate	*t*	*p*	% ΔR^2	R^2	Estimate	*t*	*p*	% ΔR^2	R^2
Intercept	261.39	595.50	< 0.001			416.35	242.40	< 0.001		
LogWF	-8.07	-17.36	< 0.001			-29.09	-16.55	< 0.001		
NumStroke	0.71	7.60	< 0.001			1.93	5.63	< 0.001		
C2.LogCF	2.02	3.33	< 0.001			11.60	4.43	< 0.001		
C2.LogNoM						-18.85	-2.34	0.019		
Baseline					0.086					0.082
SimCL	-28.48	-3.60	< 0.001	2.76	0.088	-113.43	-3.80	< 0.001	4.63	0.086
DistC						-250.69	-2.10	0.036	0.80	0.086
DistC × LogWF						301.43	2.50	0.012	1.52	0.088
Composition metrics				2.76	0.088				7.07	0.088

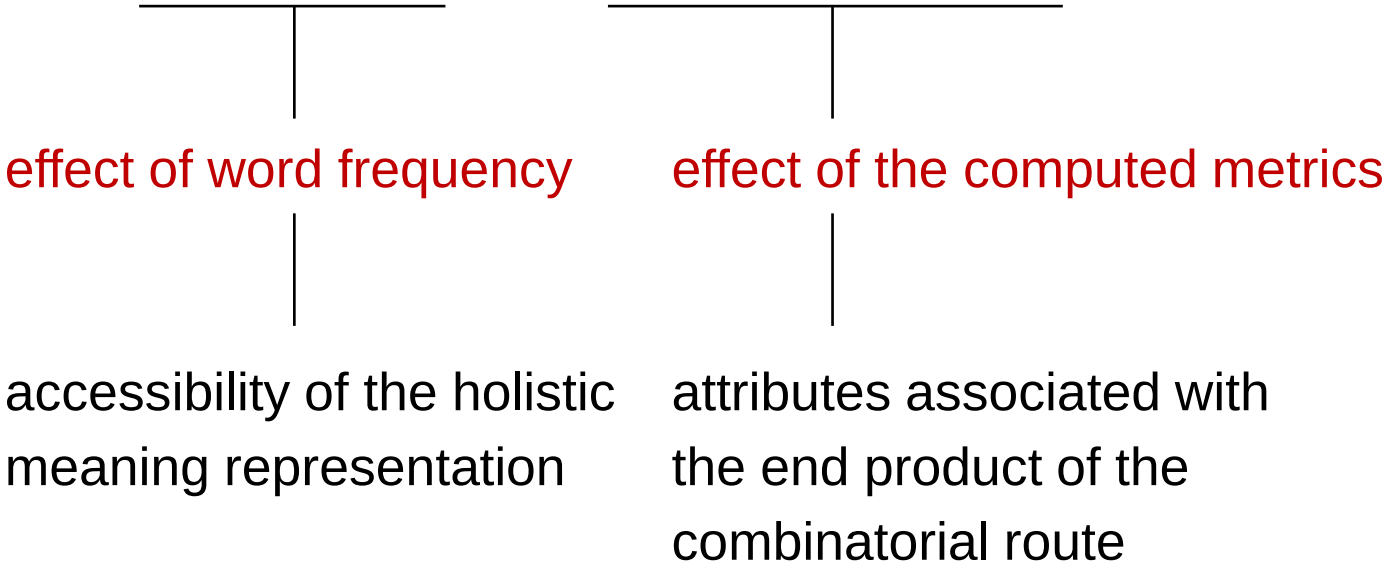
Effects of the computed metrics on compound processing

- Eye-tracking data



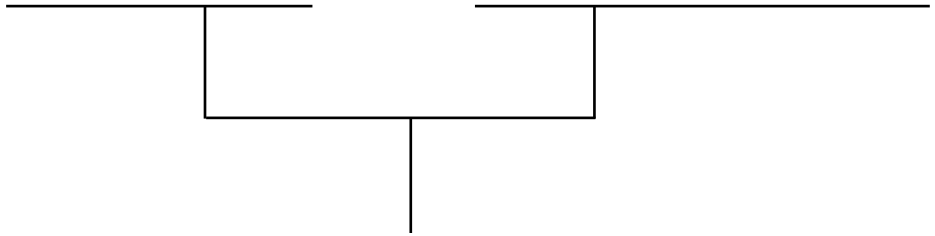
A computational characterization of the dual-route framework

- Compound meaning is accessed via a *holistic route* and a *combinatorial route*



A computational characterization of the dual-route framework

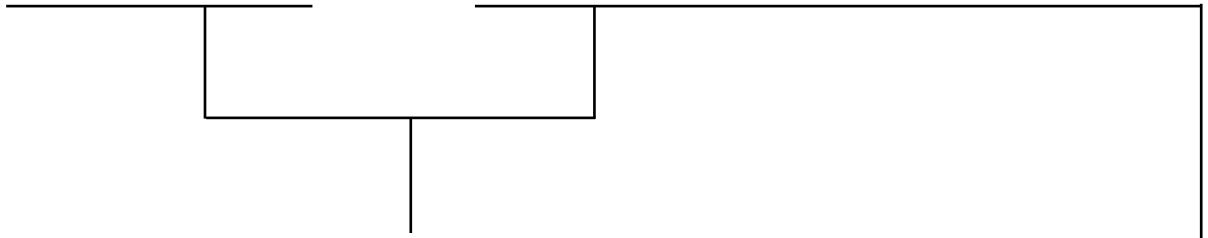
- Compound meaning is accessed via a *holistic route* and a *combinatorial route*



compositionality	relationship of the two routes
↑	convergence of the two routes ⇒ facilitate processing
↓	divergence of the two routes ⇒ hinder processing

A computational characterization of the dual-route framework

- Compound meaning is accessed via a *holistic route* and a *combinatorial route*



compositionality relationship of the two routes

↑ convergence of the two routes ⇒ facilitate processing

↓ divergence of the two routes ⇒ hinder processing

accessibility efficacy of the combinatorial route

↑ high distinctness ⇒ compositional meaning is easily accessible

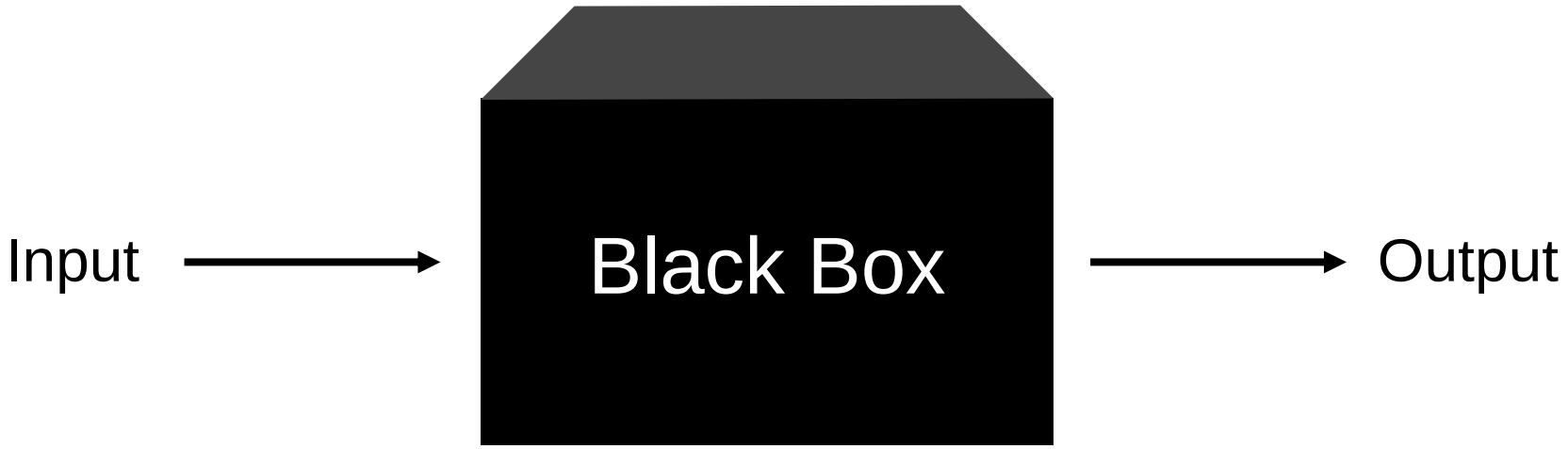
↓ low distinctness ⇒ interference from other competitors

A computational characterization of the dual-route framework

- The meaning-composition process is executed *by default* [El-Bialy et al., 2013; Libben, 2014]
  - *Low frequency words:* holistic representation of is not always available
  - *High frequency words:* combinatorial route as a backup route

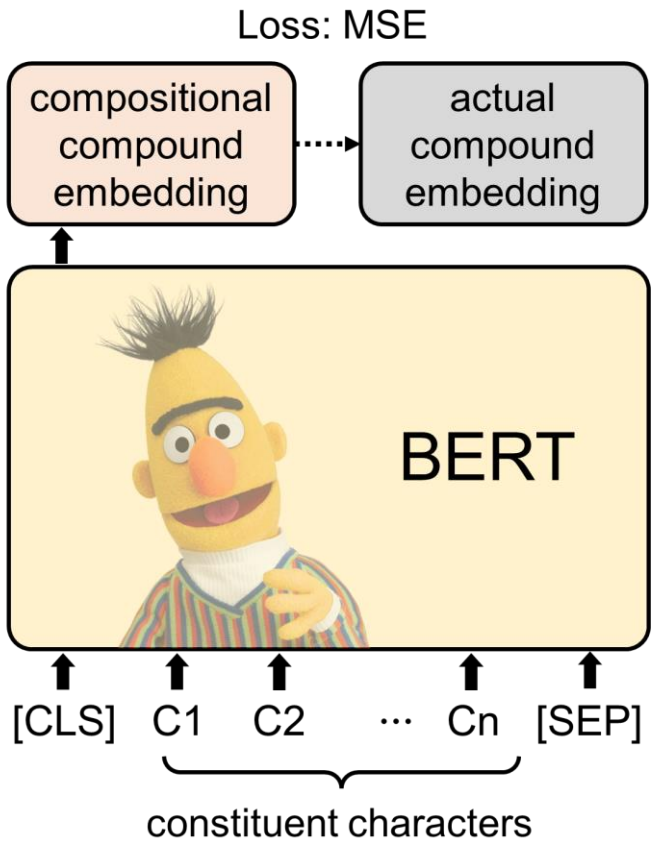
  Why?
  - Maximize the opportunity to understand compound words
  - More economical storage of lexical knowledge in the mental lexicon

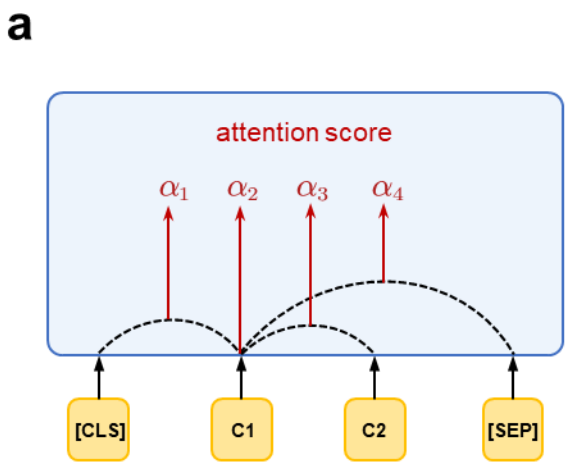
What the computational model has learnt during training?

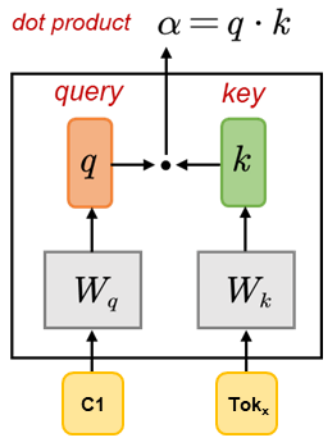


Input is converted
into output

What the computational model has learnt during training?



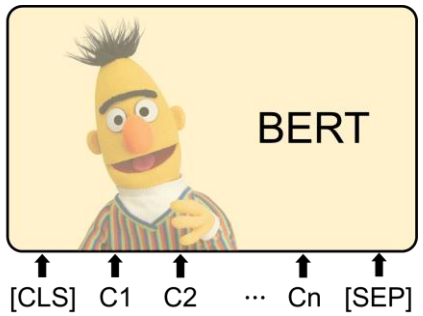


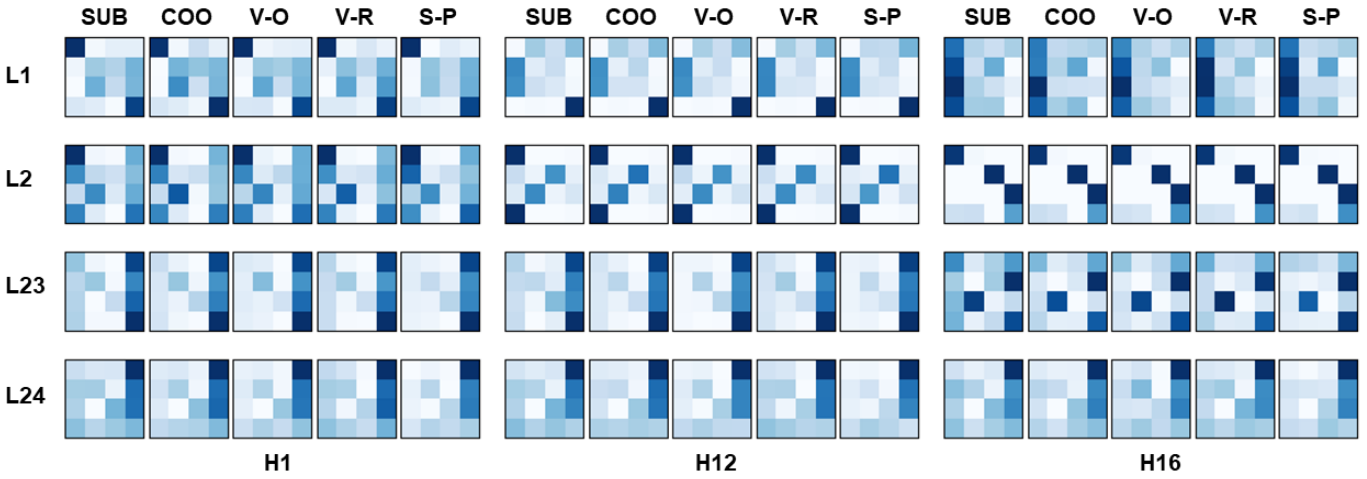


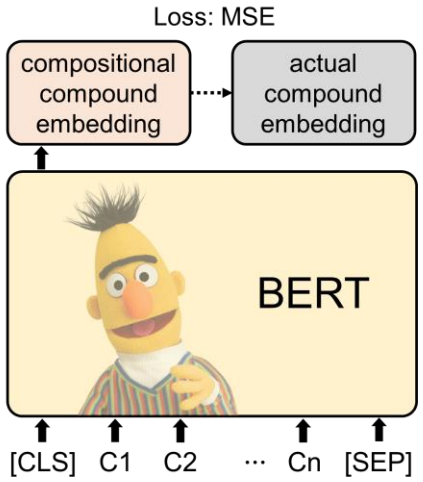
c

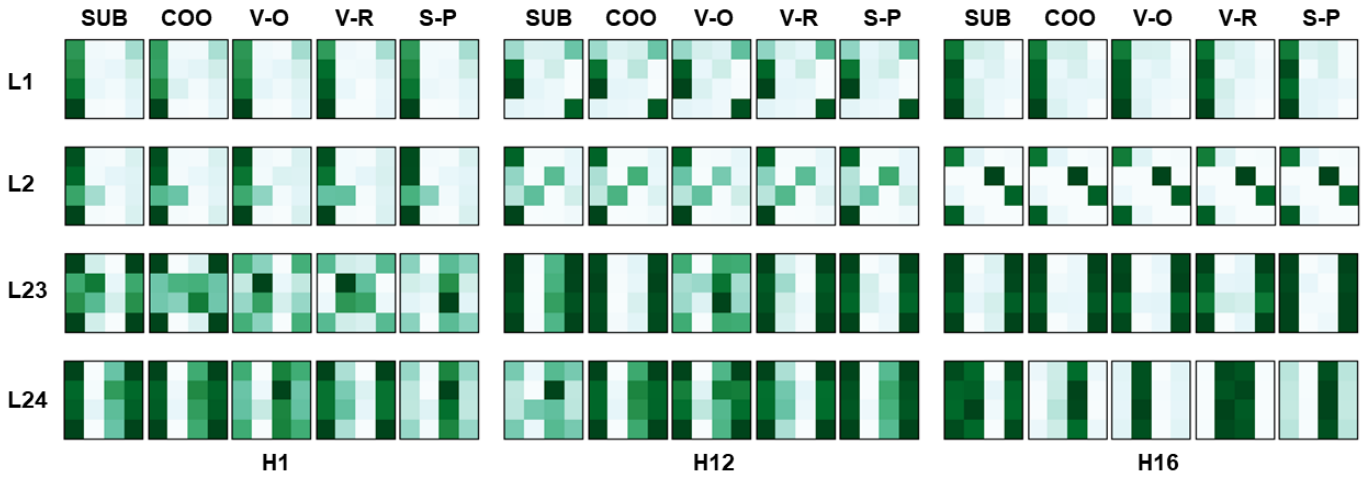
	[CLS]	C1	C2	[SEP]
[CLS]				
C1	α_1	α_2	α_3	α_4
C2				
[SEP]				

What the computational model has learnt during training?

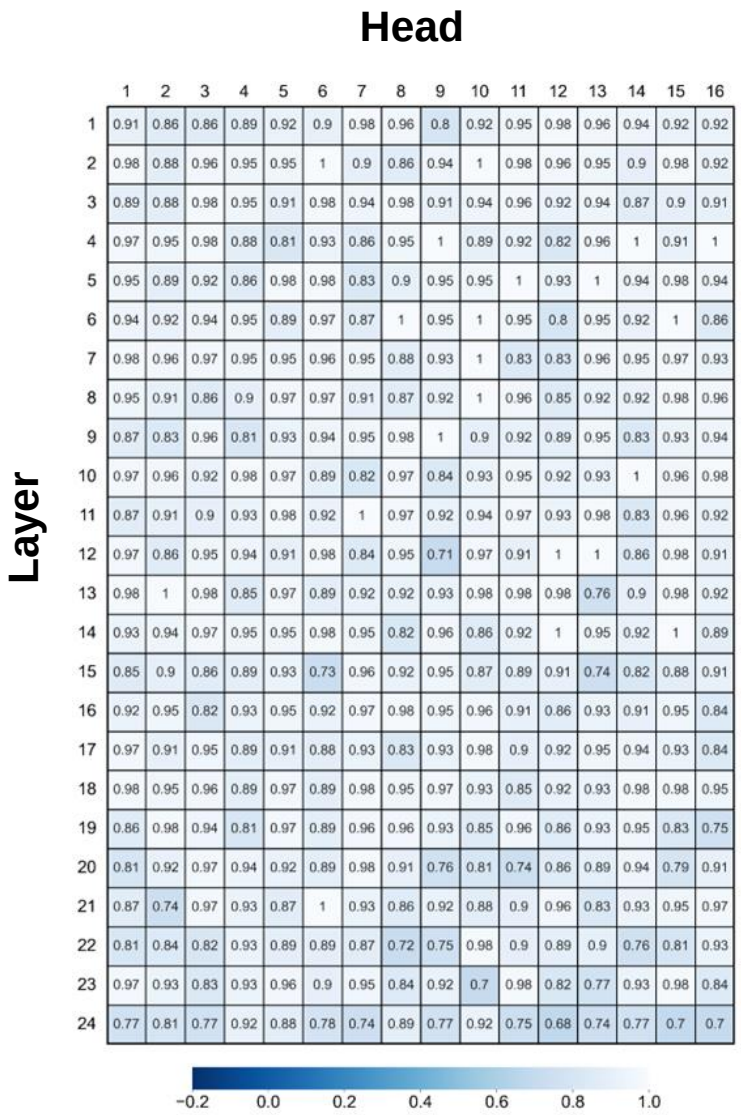


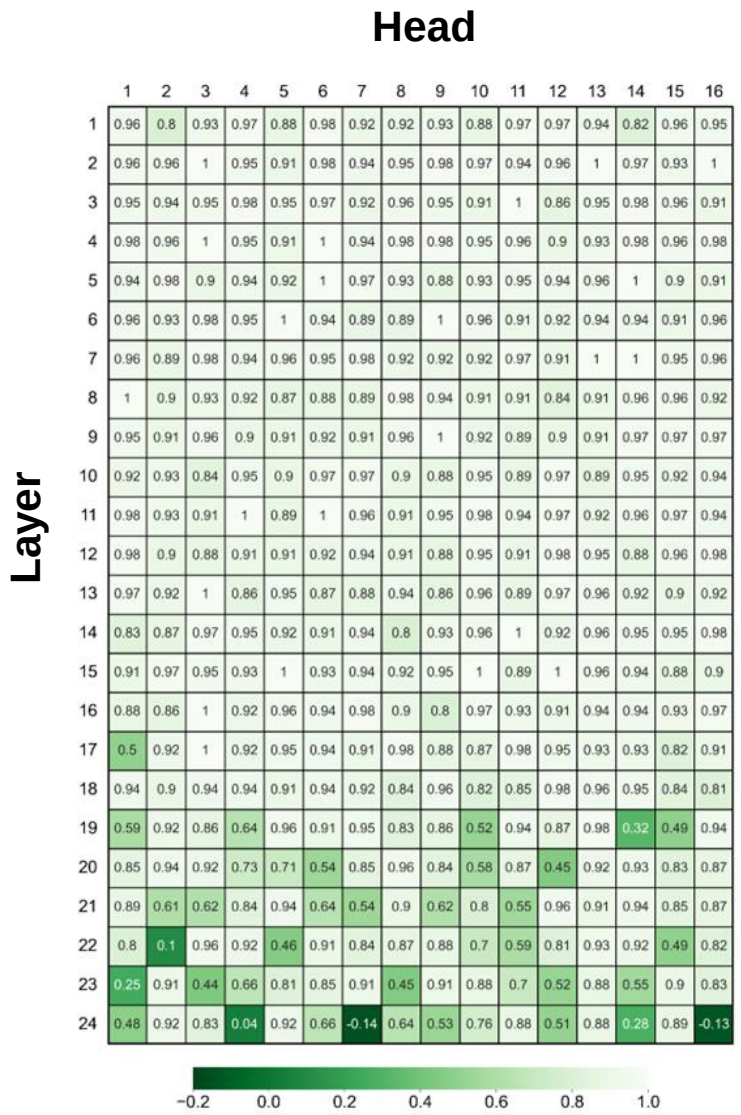






What the computational model has learnt during training?

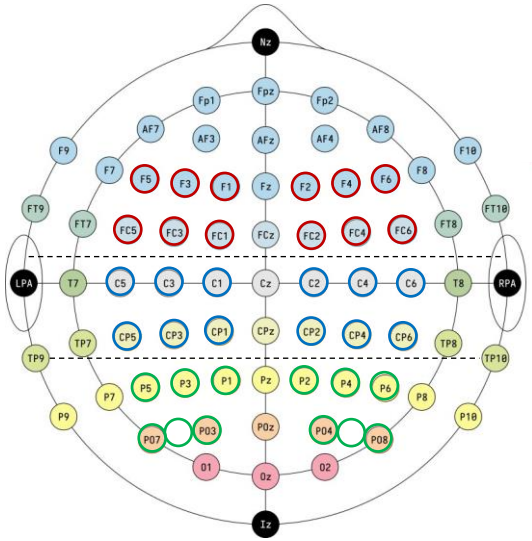




How does the compositional process unfold over time?

- ERP megastudy of Chinese word recognition [Tsang and Zou, 2022]

Trial-level ERP	~	Hemisphere:Right		Dummy variable: Hemisphere
	+	Anteriority:Central		Dummy variable: Anteriority
	+	Anteriority:Posterior		
	+	NumStroke		Number of strokes
	+	LogCD		Word frequency
	+	C1.LogCD	+ C2.LogCD	Character frequency
	+	C1.LogNoH	+ C2.LogNoH	Number of homophones
	+	C1.ST	+ C2.ST	Semantic transparency
	+	SimCL		
	+	DistC		
	+	(1|Subject)	+ (1|Item)	



Left-frontal Right-frontal
Left-central Right-central
Left-posterior Right-posterior

↓

Each time window FDR correction

How does the compositional process unfold over time?

- Lexical decision of real words

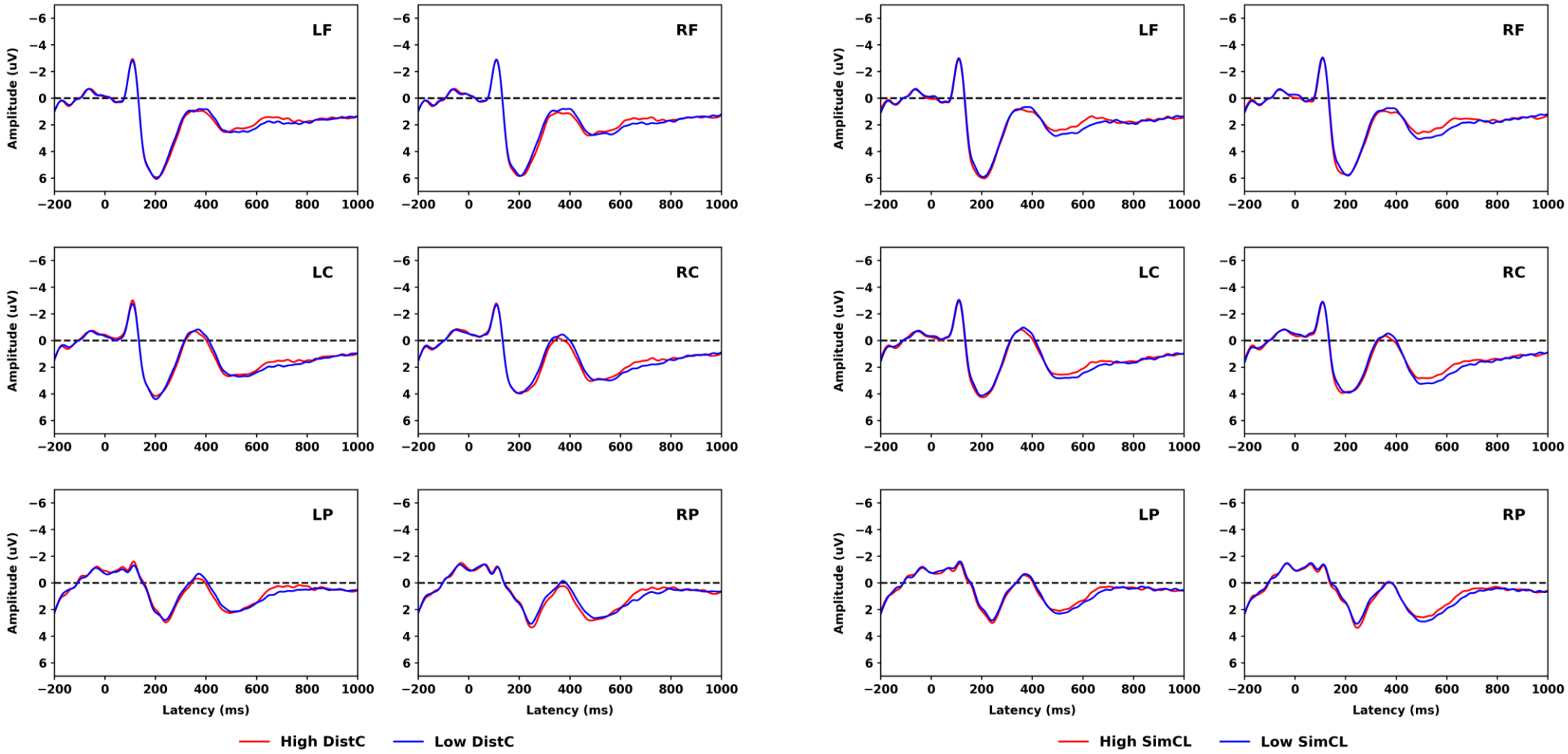
	TW1 0-100	TW2 100-200	TW3 200-300	TW4 300-400	TW5 400-500	TW6 500-600	TW7 600-700	TW8 700-800	TW9 800-900	TW10 900-1000
Intercept	-1.19	0.84	1.83	-1.27	3.15	5.71	4.43	2.57	1.74	1.03
Hemisphere:Right	-0.13	0.01	0.10	0.38	0.38	0.31	0.16	-0.01	-0.05	-0.01
Anteriority:Central	0.01	0.01	-0.01	-0.06	0.00	0.03	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

higher distinctness → smaller N400: Fewer lexicalized distractors and relative ease of semantic access

higher distinctness and compositionality → smaller LPC/P600: Ease of generating a holistic representation

How does the compositional process unfold over time?

- Lexical decision of real words



How does the compositional process unfold over time?

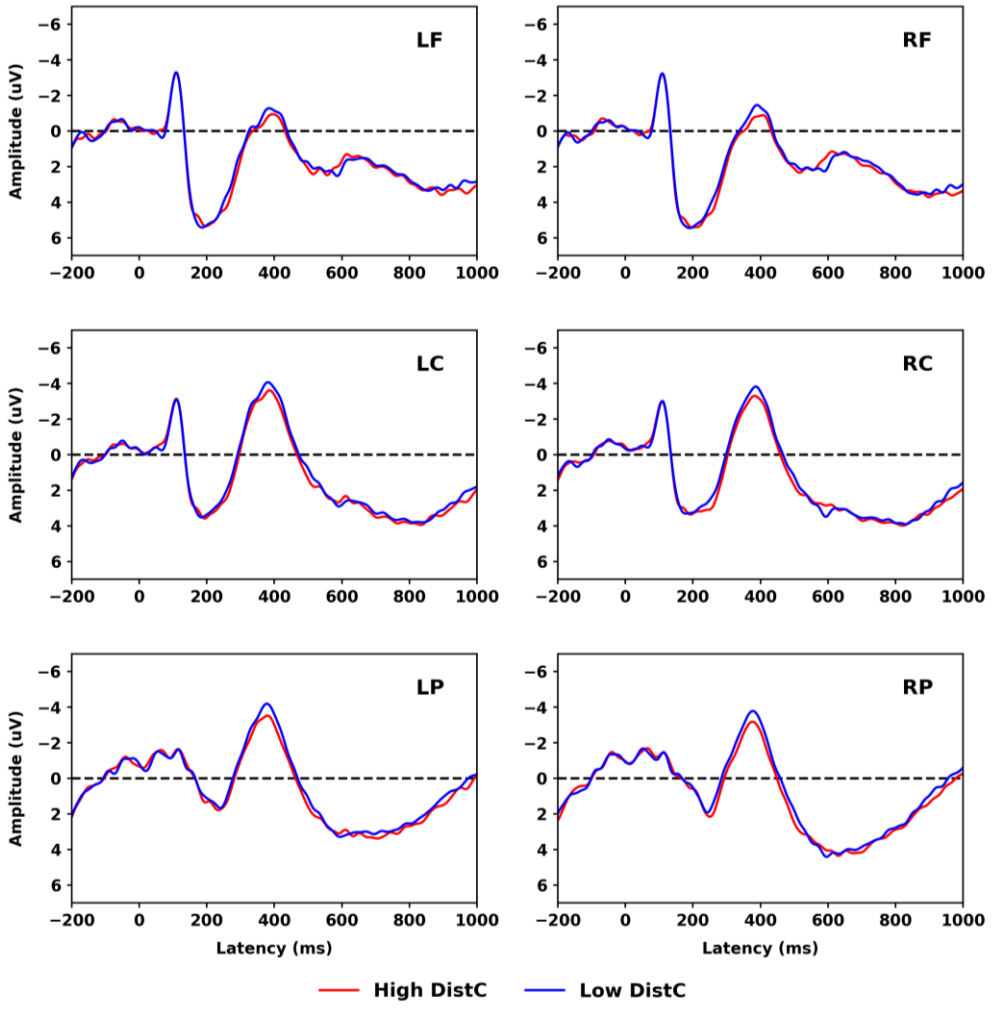
- Lexical decision of nonwords

	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	-1.37	1.39	2.18	-4.14	0.22	0.11	1.66	4.64	3.23	1.34
Hemisphere:Right	-0.12	0.03	0.20	0.34	0.37	0.43	0.38	0.13	0.05	-0.08
Anteriority:Central	0.02	0.00	-0.03	-0.07	-0.01	0.00	0.01	0.04	0.05	0.07
Anteriority:Posterior	-0.78	-1.90	-2.17	-1.25	-0.31	0.59	1.15	0.03	-1.34	-2.25
NumStroke	0.01	0.00	0.00	0.02	-0.02	0.01	0.02	-0.01	0.01	0.02
C1.LogCD	0.06	-0.15	-0.01	0.18	-0.01	0.15	0.03	-0.35	0.06	0.07
C2.LogCD	0.19	0.11	0.11	0.24	-0.24	0.03	-0.09	-0.21	-0.22	0.05
C1.NoH	0.25	0.15	0.09	-0.19	0.19	0.37	0.06	0.28	0.23	-0.10
C2.NoH	0.02	0.14	-0.05	0.33	-0.12	0.05	-0.13	-0.25	-0.20	0.22
DistC	-5.51	-0.84	20.17	35.81	25.49	13.43	2.78	5.09	9.64	18.70

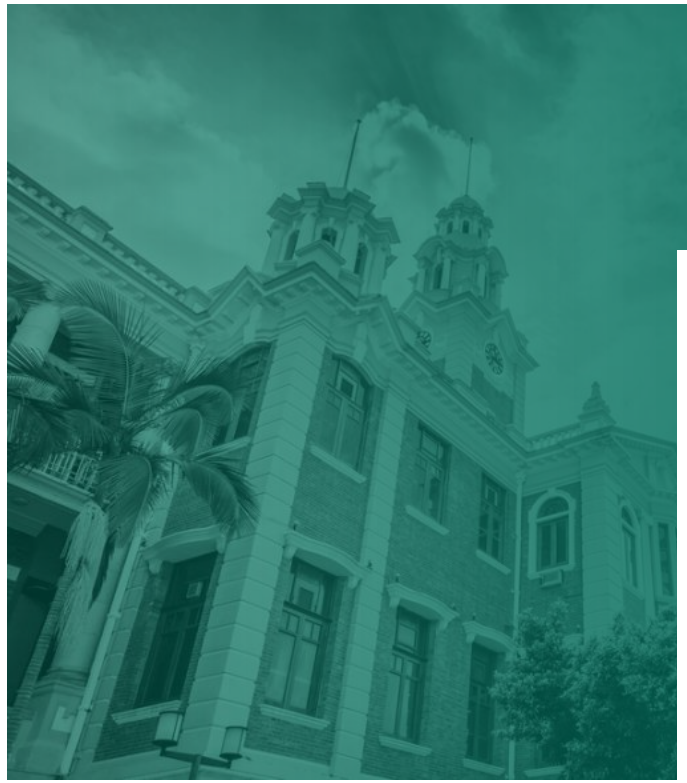
higher distinctness → smaller N400

How does the compositional process unfold over time?

- Lexical decision of nonwords



RESEARCH TALK



THANKS

tianqi93@connect.hku.hk
https://tianqi93.github.io/