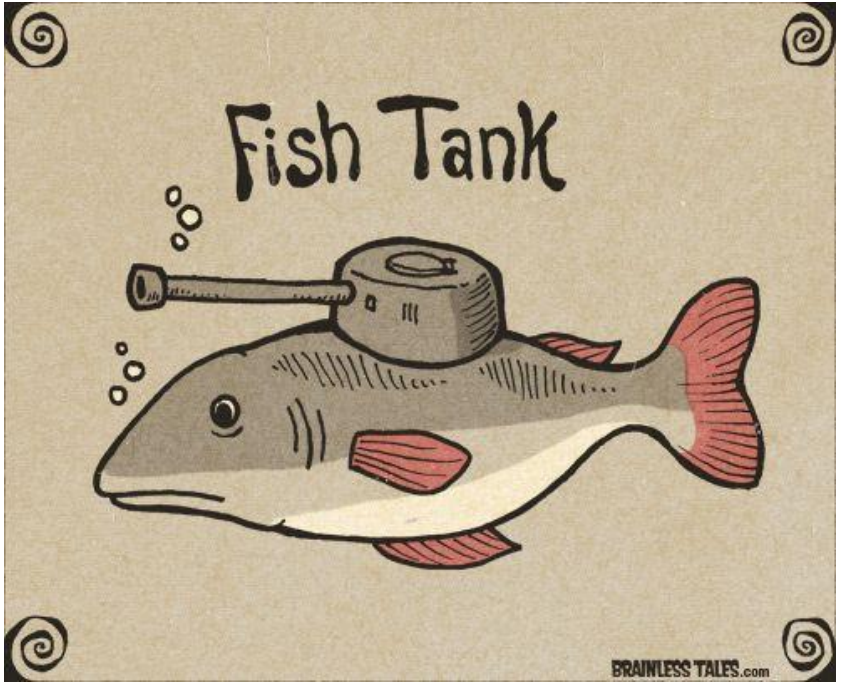
Making Sense from the Parts: What Chinese Compounds Tell Us About Reading and Learning

Cheng-Yu Hsieh

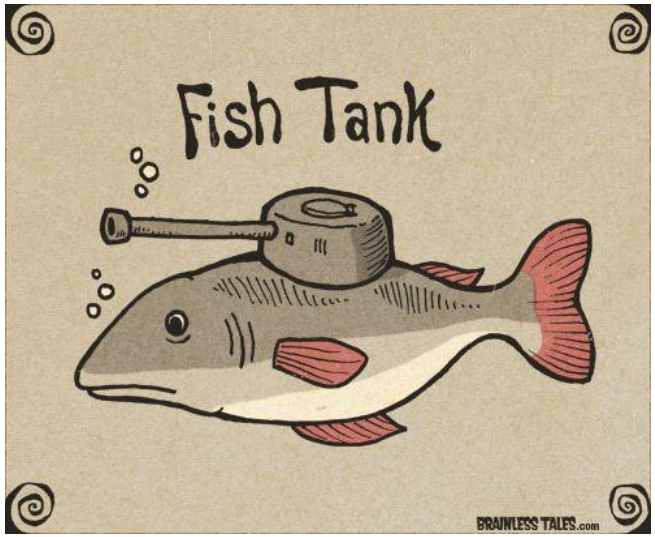
27th October 2025 @University of Milano-Bicocca

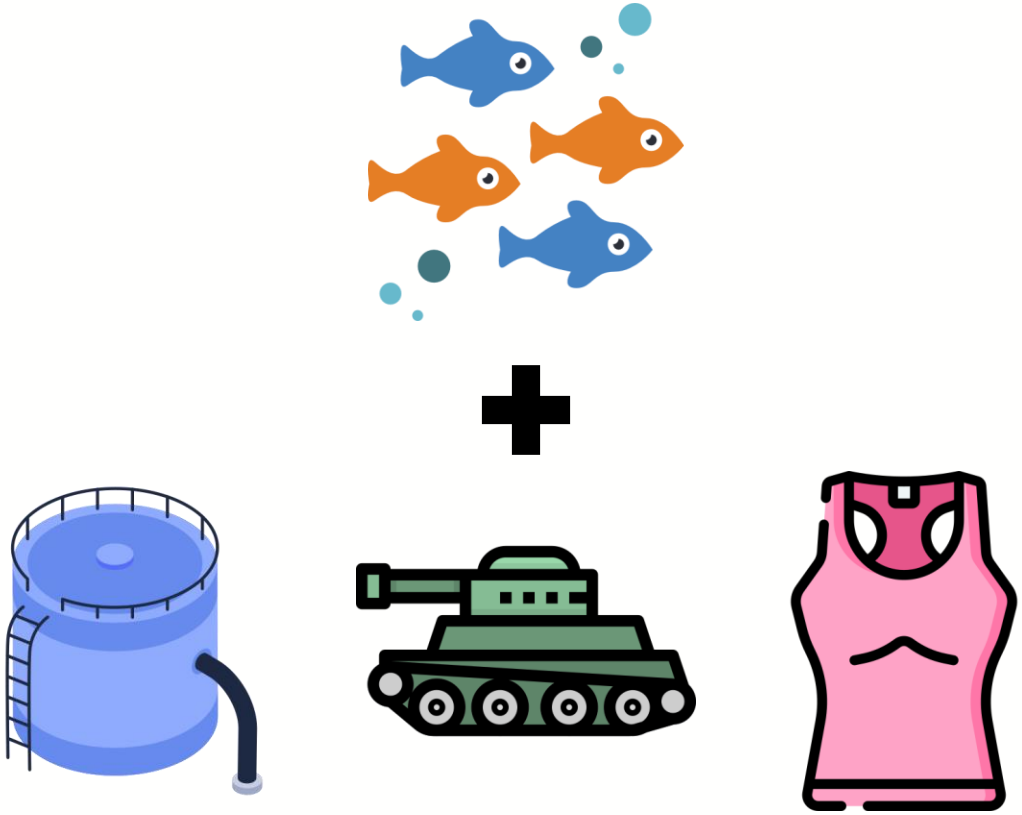
ROYAL HOLLOWAY UNIVERSITY OF LONDON

Fish tank



Fish tank





Chinese characters

這花瓶真漂亮。

Chinese characters

A basic unit of meaning

這花瓶真漂亮。

‘this’

Chinese characters

A basic unit of meaning

這**花**瓶真漂亮。

Chinese characters

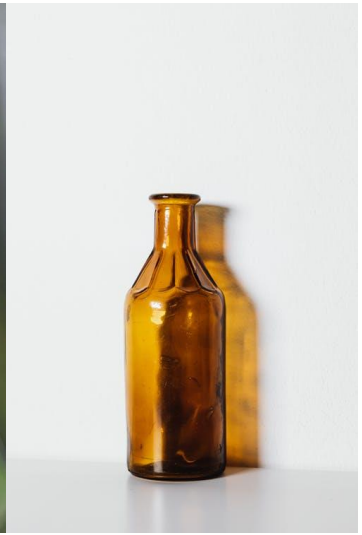
A basic unit of meaning

這花**瓶**真漂亮。

Chinese characters do not always stand alone

2-character compound word (about 75%)

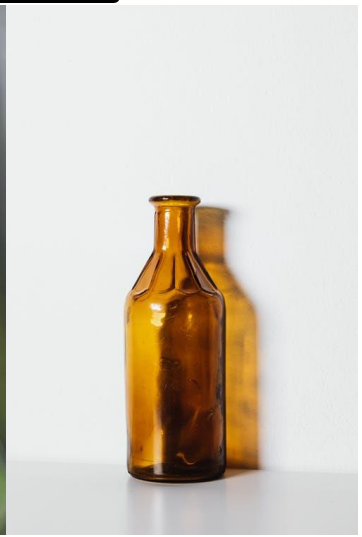
這**花瓶**真漂亮。



Chinese characters do not always stand alone

‘huā’

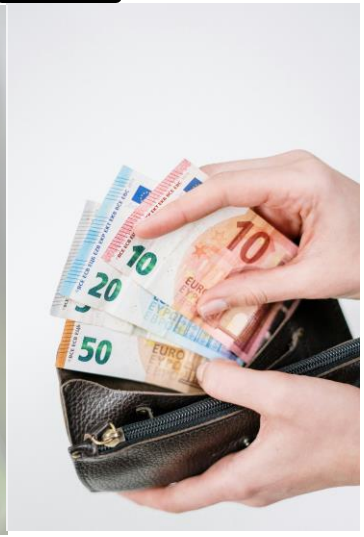
花瓶

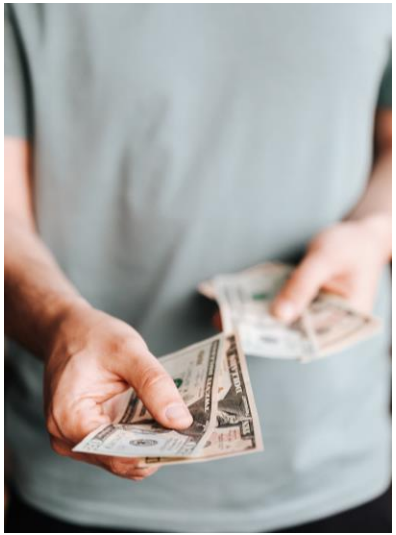


Chinese characters may bear multiple meanings

‘huā’

花錢





Chinese characters may bear multiple meanings

‘huā’

花枝

- flower
- to spend

Inconsistent meanings of characters

‘huā’

花枝

Character meanings …

- depend on the word contexts
- can behave unreliably across compounds
- can be hard to generalise

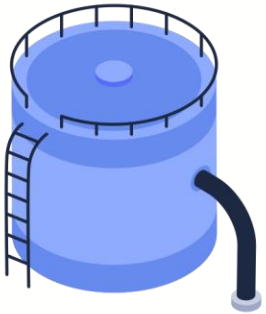
Novel compound: New combination of characters

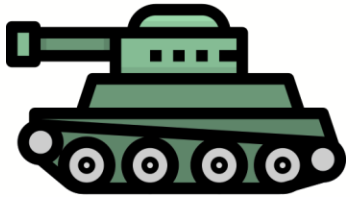
‘huā’

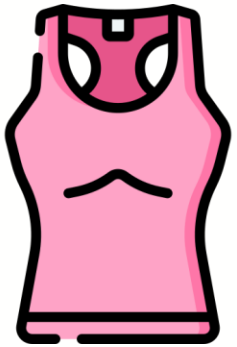
花課

- Novel compound processing requires character knowledge.
- Novel compounds are commonly found in daily life.

ice *tank*







Research question

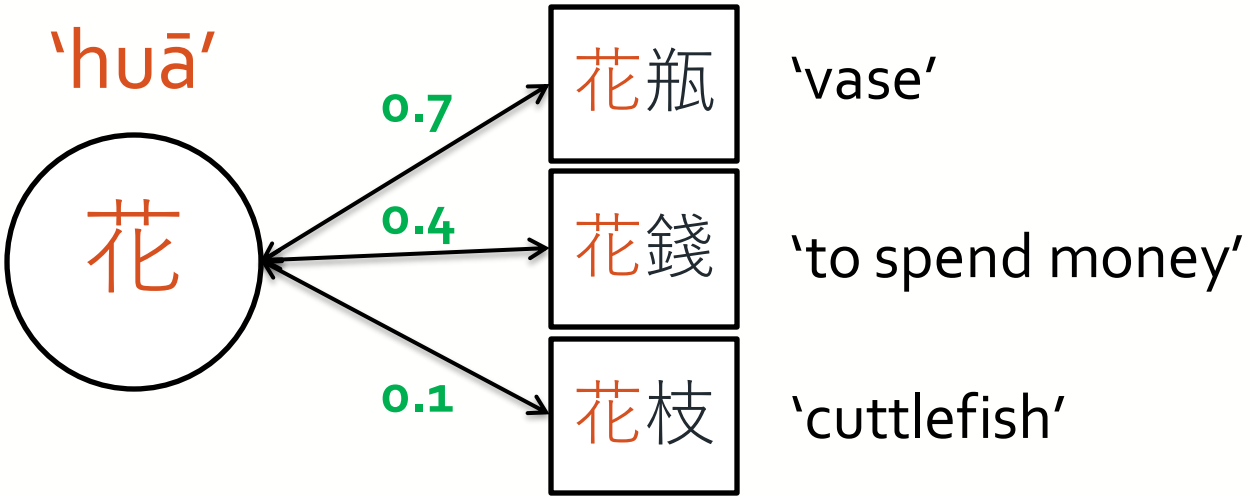
- How do Chinese readers combine the meanings of parts into a coherent whole when the consistency of the parts varies?

Research question

- How do Chinese readers combine the meanings of parts into a coherent whole when the **consistency** of the parts varies?

Semantic consistency — Reliability of meaning

- distributional semantic model: vectors of characters & compounds
- **semantic similarity**: proximity between two vectors



semantic consistency = 0.4

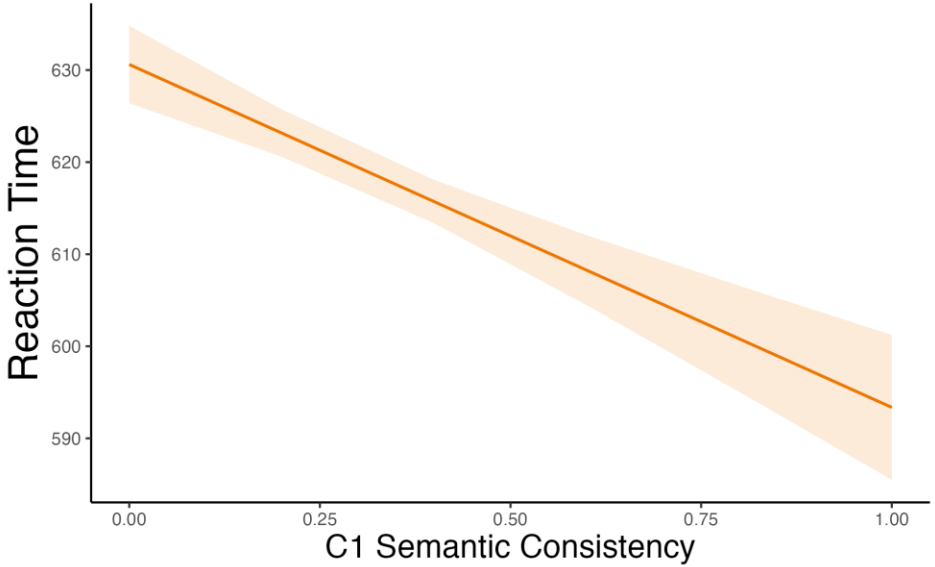
Megastudy of lexical decision data

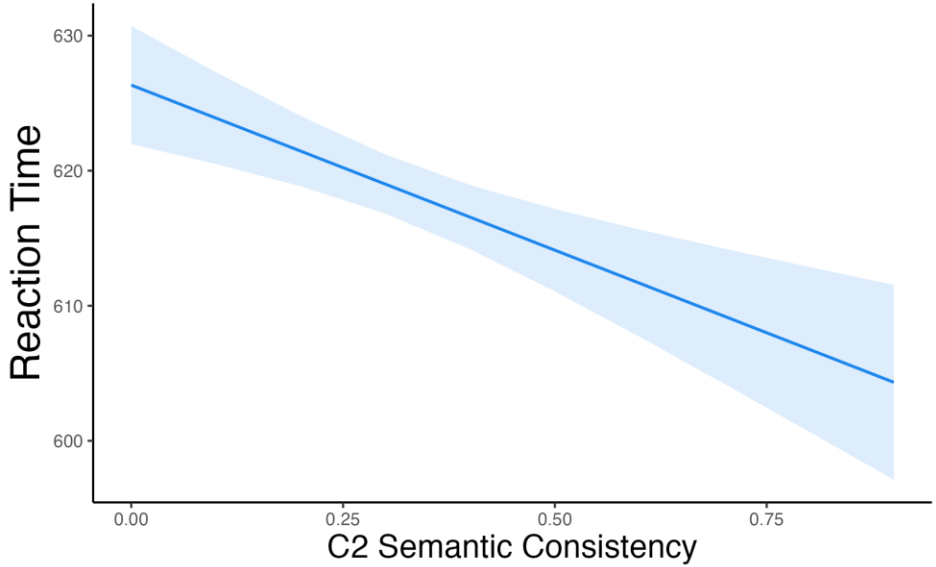
- Does the semantic inconsistency of characters impact the lexical decision of compounds?
  - 22,000 existing compounds and novel compounds (Tse et al., 2017, *BRM*).

'vase'

'huā–class'

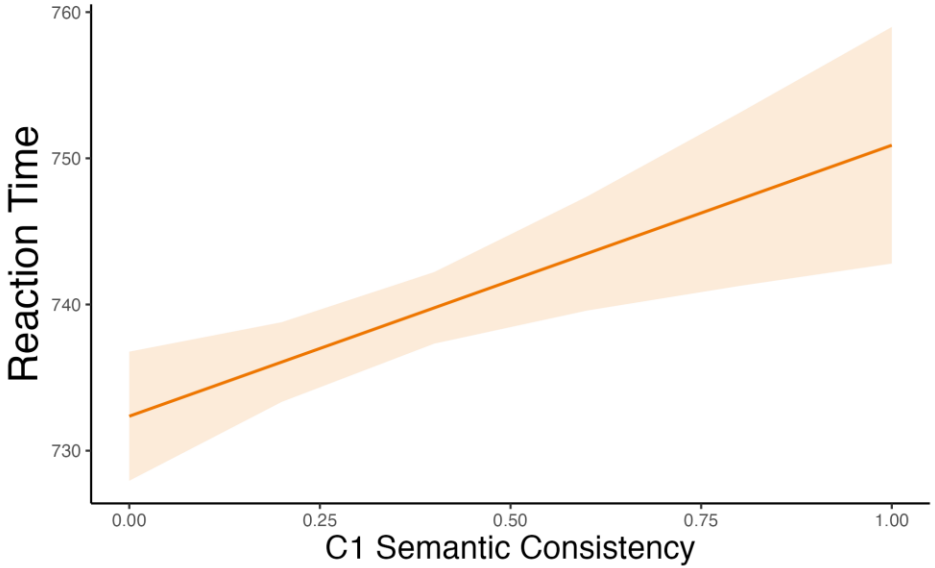
High semantic consistency speeds recognition of compounds.

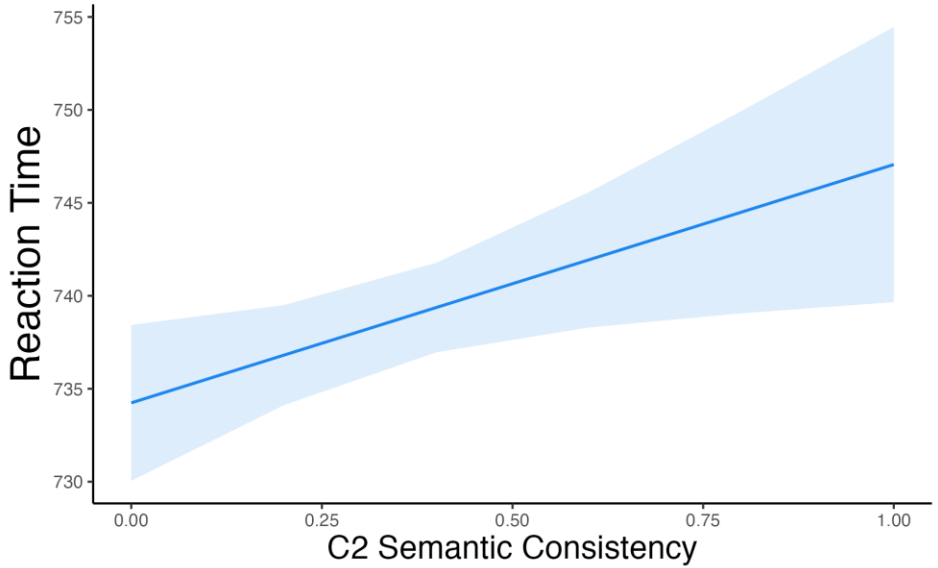




- Unreliable characters slow down the recognition of the compound.

High semantic consistency slows rejection of novel words.





- Novel compounds containing reliable characters look less like real compounds and are harder to reject.

Summary

- The reliability of communicating meanings impacts the processing of both existing and novel compounds.

→ Character knowledge is applied during compound processing.

- Readers are sensitive to the meaningful information that they acquire through their experience with compound words.

Learning the concept of *dog*

Inconsistent characters in novel compounds

ROYAL HOLLOWAY UNIVERSITY OF LONDON

‘huā’ 花 + 課 → ‘huā’ 花課

- How do readers make sense of novel compounds, especially those containing “misbehaving” characters?

Building *compositional* representations (CAOSS)

lexical representation

compositional representation

difference minimised

Building *compositional* representations (CAOSS)

lexical representation

compositional representation

difference minimised

Trained on 20K compounds

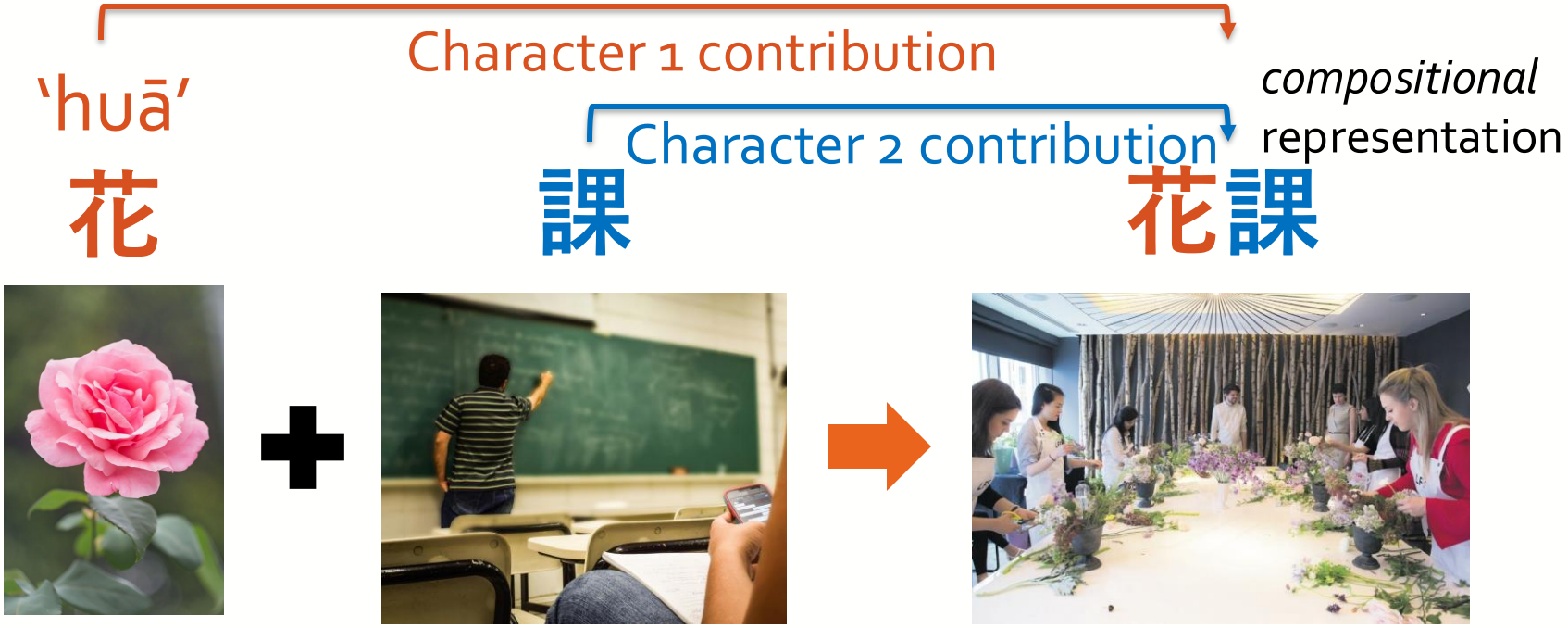
Building *compositional* representations (CAOSS)

compositional representation

lexical representation

approximation of novel word meaning

Building *compositional* representations (CAOSS)

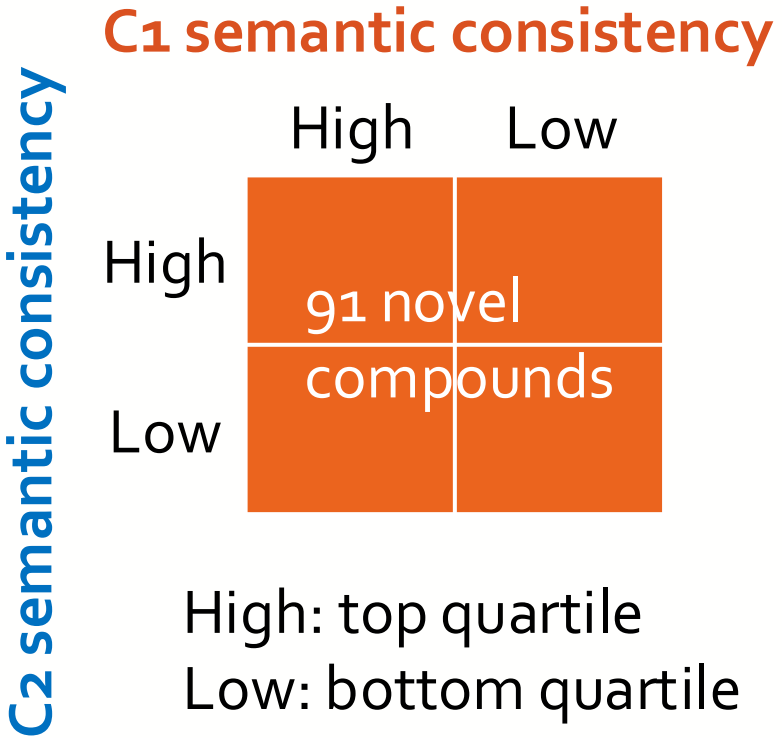


- Do unreliable characters contribute less to the meaning of the novel word?

Manipulating semantic consistency of characters

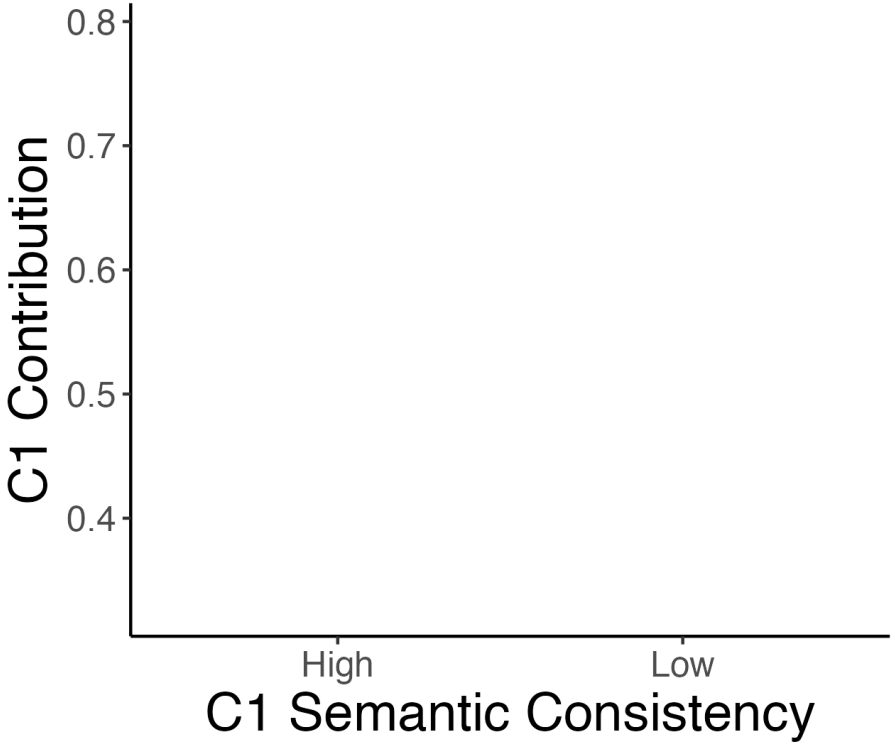
meaningfulness rating ≥ 3
(on a scale of 1-5)

1500 novel compounds → 378 novel compounds →

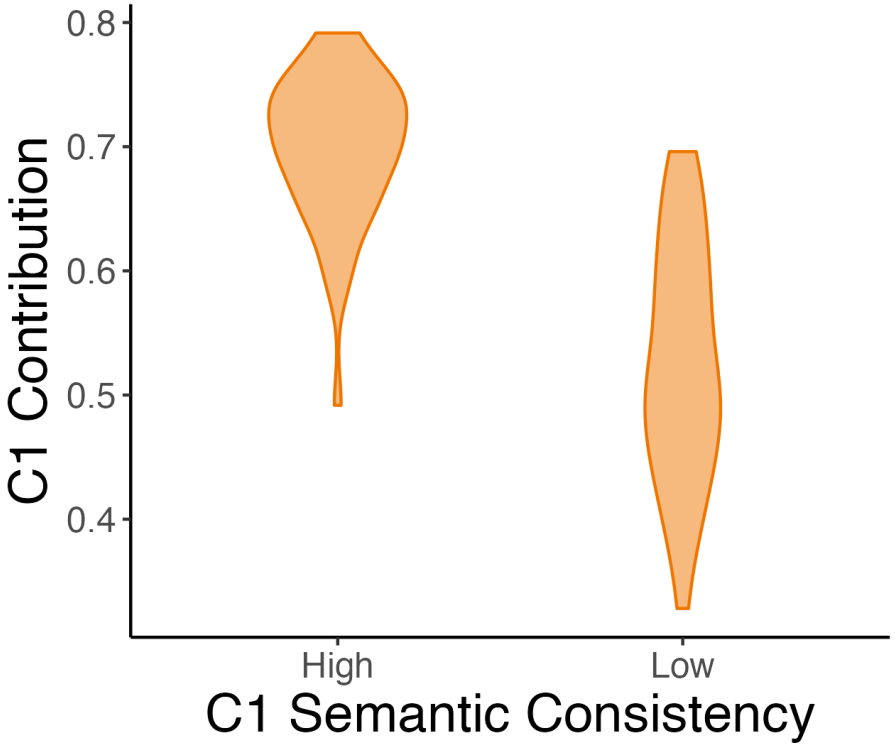


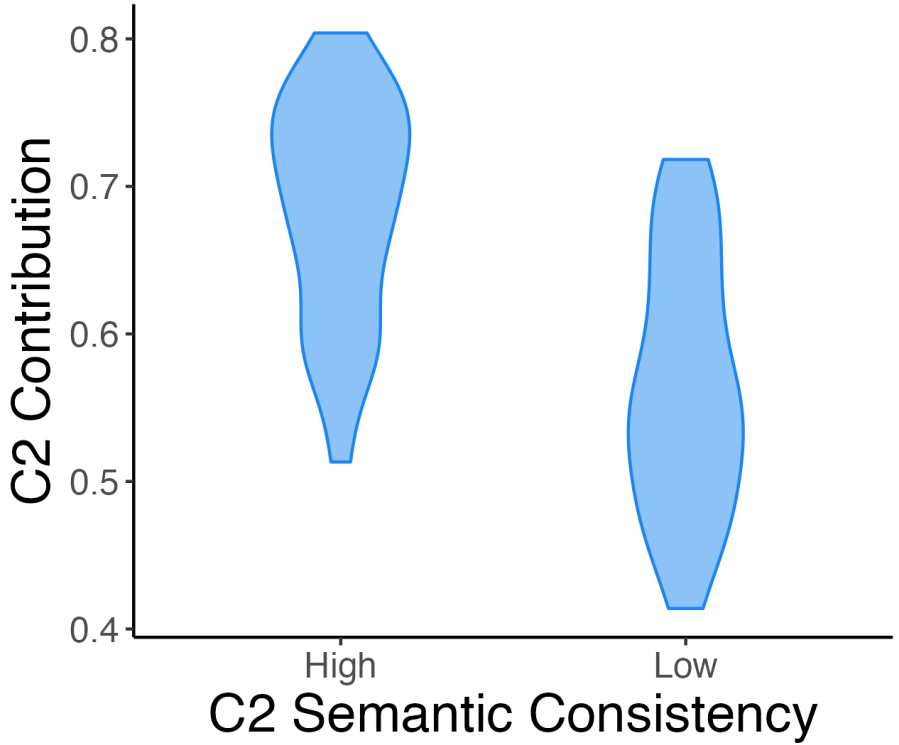
High: top quartile
Low: bottom quartile

Consistent characters tend to contribute more



Consistent characters tend to contribute more





- The model learns to weight reliable characters more.

Behavioural study: Free association task

‘human’ 人譯 ‘translation’ (*n* = 91)

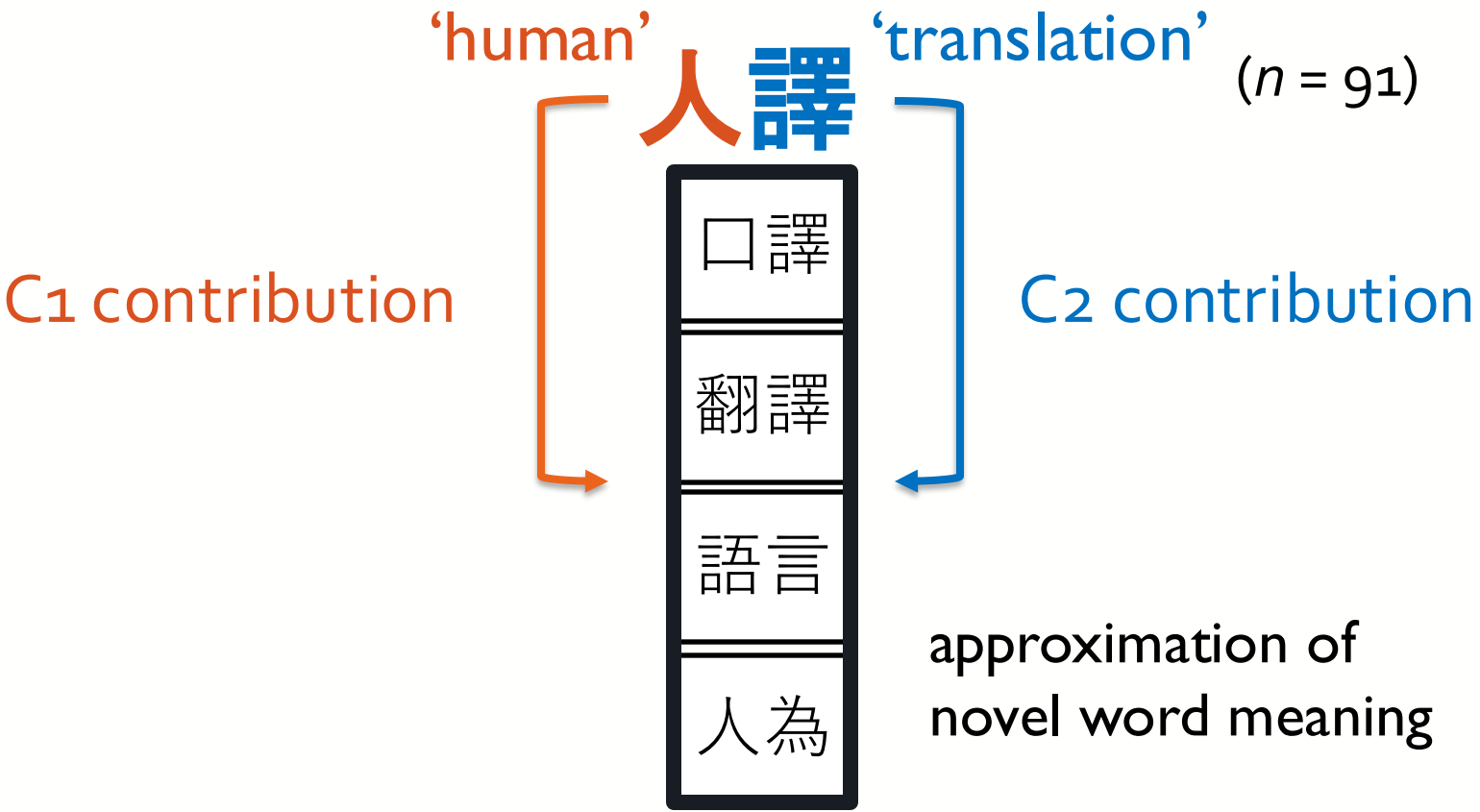
Behavioural study: Free association task

‘human’ 人譯 ‘translation’ (*n* = 91)

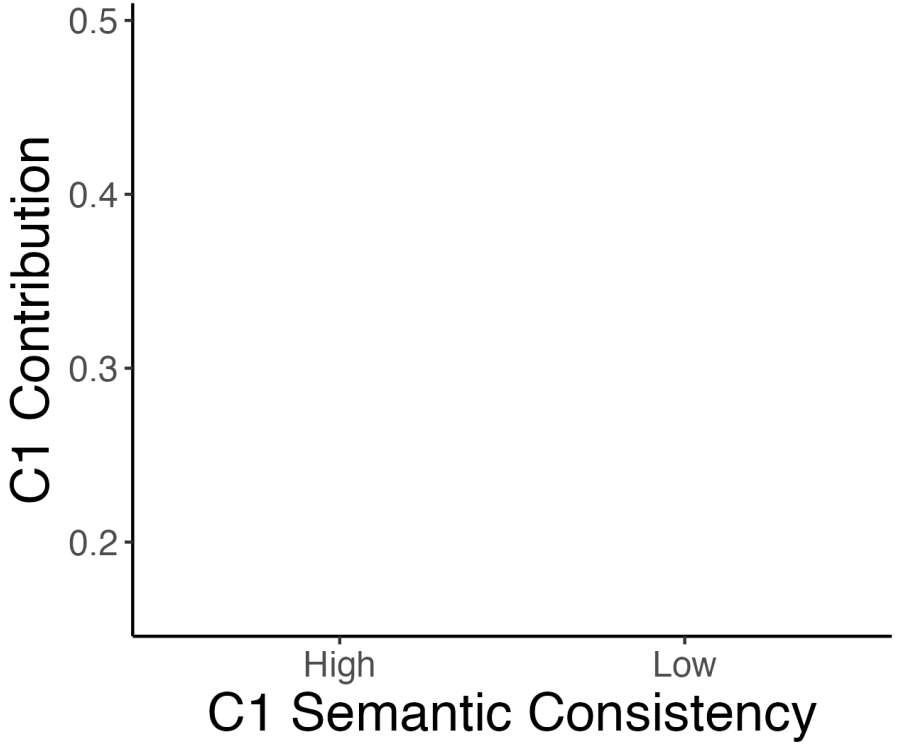
participants’ responses (*n* = 24):
an existing compound word as a synonym of the target novel word

口譯	‘interpreting (between languages)’
翻譯	‘translation’
語言	‘language’
人為	‘man-made’

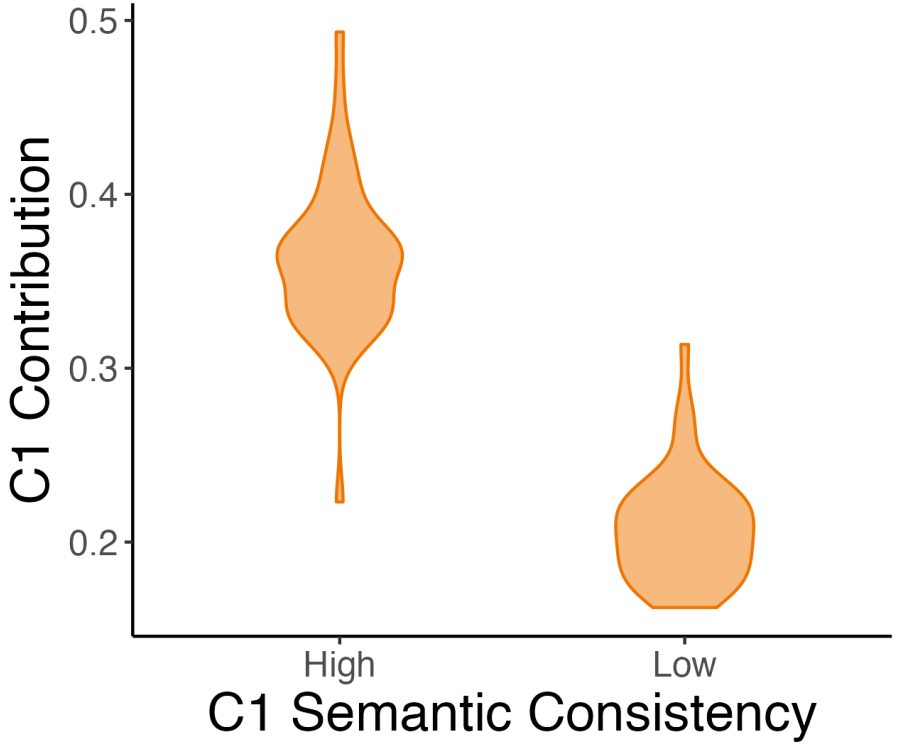
Behavioural study: Free association task

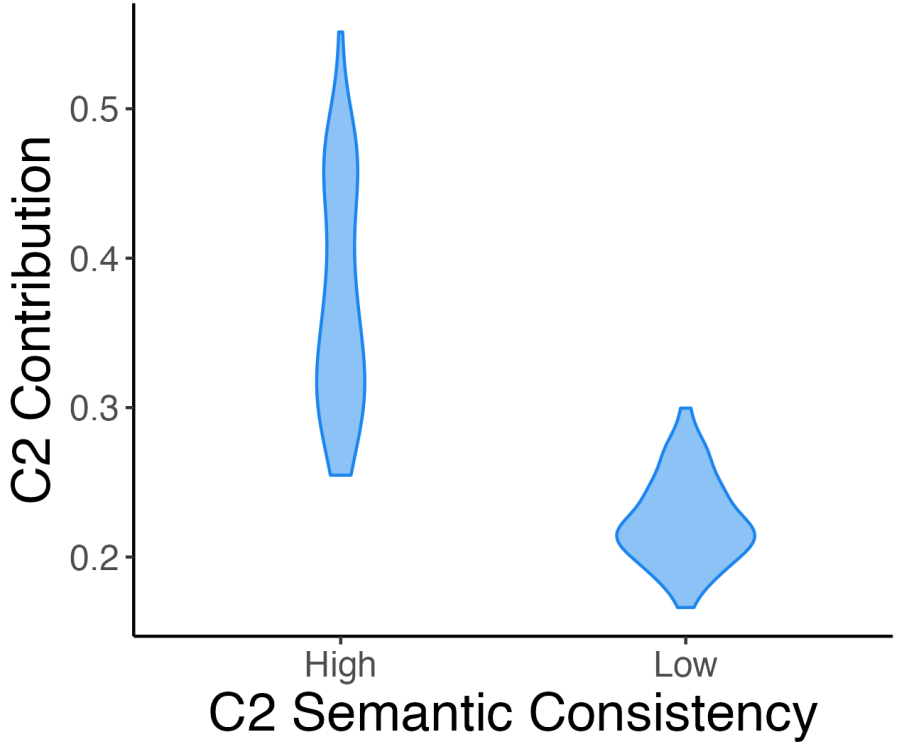


Consistent characters tend to contribute more



Consistent characters tend to contribute more



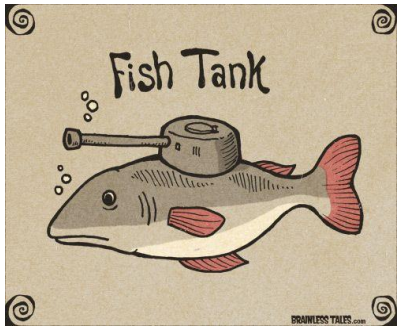


- Readers tend to weight reliable characters more.

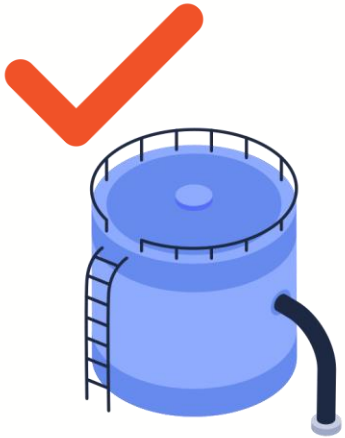
Summary

- Both the model and humans rely more on characters that communicate reliable meanings when they make sense of novel compounds.

Readers look for reliable meaningful information



Consistent constituent may serve as an ***anchor***?



Readers look for reliable meaningful information

‘huā’ 花

+

‘huā’ 課

→

‘huā’ 花課

↓

Consistent characters may serve as an ***anchor***?

Research question

- Do consistent characters serve as an anchor during the comprehension of novel words?
- Do readers tend to agree on the meaning of the inconsistent character if they co-occur with a consistent character?

Consistent character as an ***anchor***?

Target Character		Context Character		Novel Compound
勞 *labour* *fatigue* *credit*	✚	High Consistency 遣 *dispatch*	➡	遣勞
		Low Consistency 大 *big, extremely, old, important*	➡	大勞

Behavioural study: Meaning identification task

target character (*n* = 59)

‘dispatch’ 遣勞

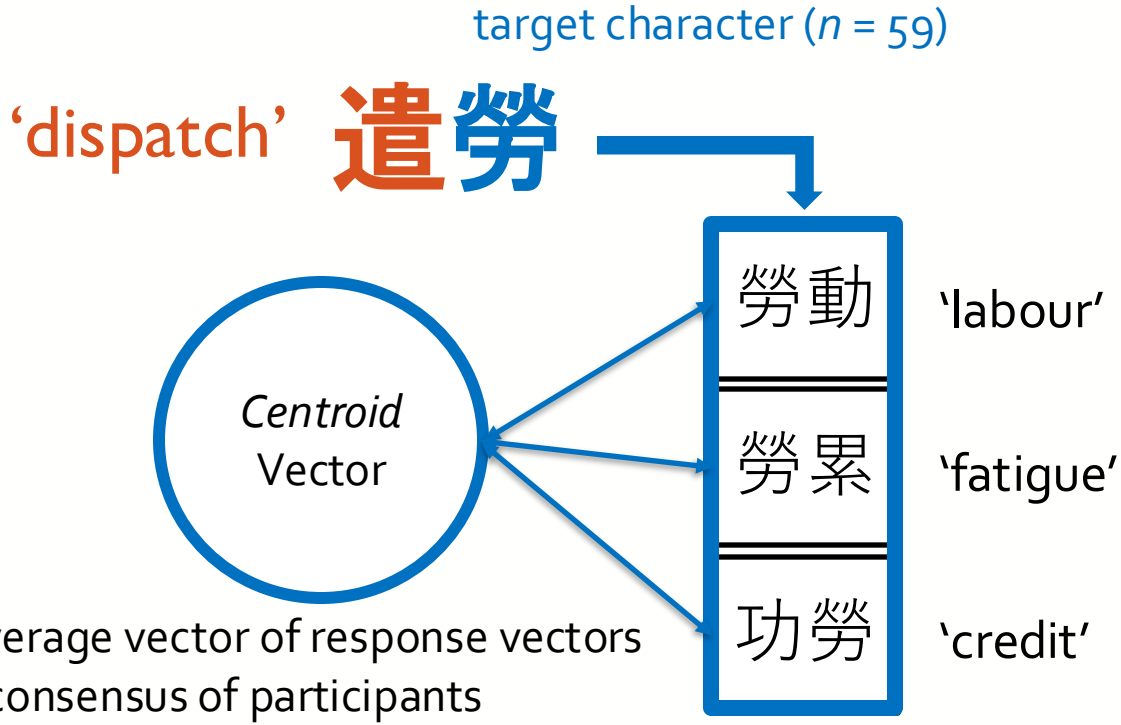
participants’ responses (*n* = 44): an existing compound word as a synonym of the target character

‘labour’

‘fatigue’

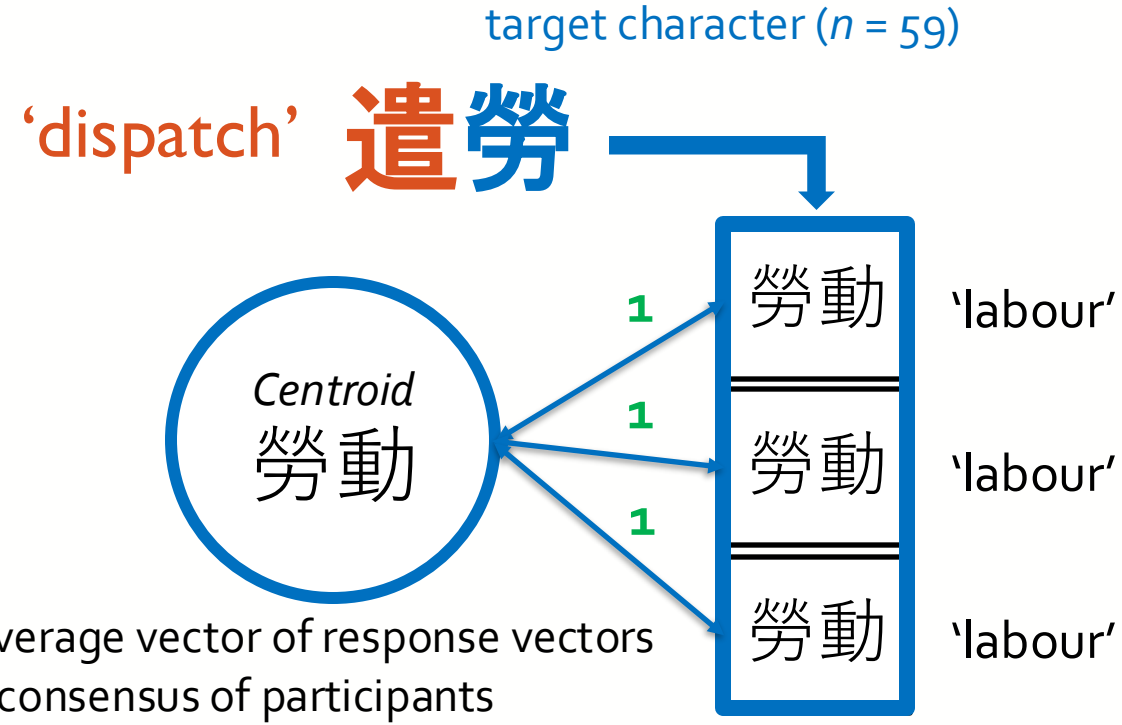
‘credit’

Behavioural study: Meaning identification task



- ***semantic agreement*: the semantic similarity between the *centroid* and each of participants' responses; the value represents the alignment with others' response.**

Behavioural study: Meaning identification task



- ***semantic agreement*: the semantic similarity between the *centroid* and each of participants' responses; the value represents the alignment with others' response.**

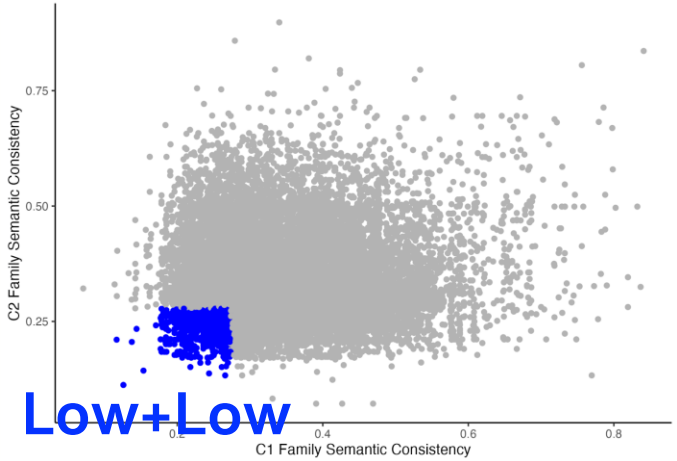
Example of participants' responses

Target: 勞	High consistency 遣 *dispatch*	Low consistency 大 *big, old …*
功勞 *credit*	x 1	x 1
勞動 *labour*	x 19	x 11
勞累 *fatigue*	x 2	x 10
Semantic Agreement	0.78	0.64

Character meaning aligns in consistent contexts.

Summary

- Skilled readers use meaningful cues from consistent characters as an *anchor* to make sense of the ambiguous character.
- Despite ambiguity in characters, meanings of Chinese compounds is more learnable than it appears.



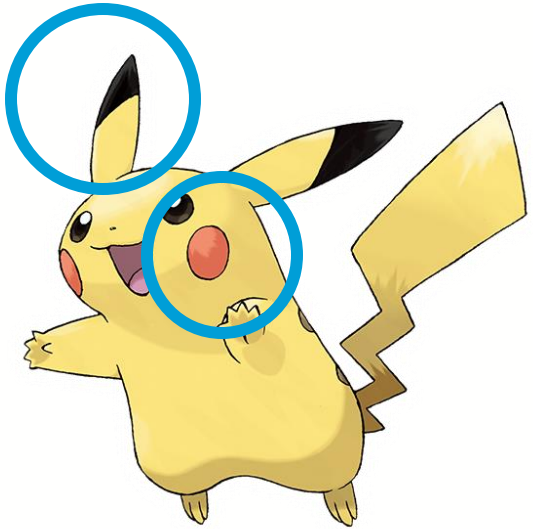
$$\frac{\mathbf{659}}{12{,}242} = 5.4\%$$

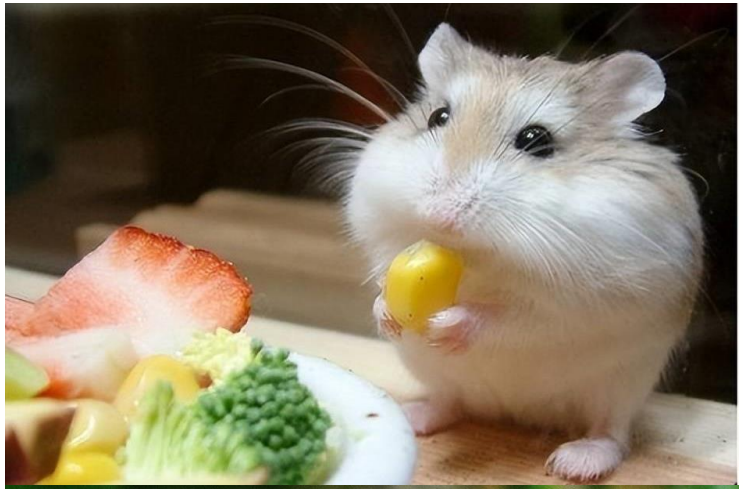
How readers make sense of Chinese compounds?

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Readers pick up knowledge of Chinese characters through exposure to compound words.
- Readers draw on their understanding of a character's behaviour when encountering both familiar and unfamiliar compounds.
- Readers look for reliable, meaningful information to reduce uncertainty when constructing meanings of novel compounds.

Reliable cues for concept learning





Thank you for your attention!

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Semantic Consistency

JEP:LMC

Computational Model

PBR

Cheng-Yu.Hsieh.2021@live.rhul.ac.uk