CS310 Natural Language Processing
自然语言处理
Lecture 14 - Interesting Papers from Recent NLP Conferences

Instructor: Yang Xu

主讲人：徐炀

xuyang@sustech.edu.cn

Paper list

- Scaling in Cognitive Modeling -- Reading Time
- Reconstruct Language from Cognitive Signal
- Cognitive Reframing of Negative Thoughts
- Shuowen and JieZi

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Game: Guess the Paper from the Chart

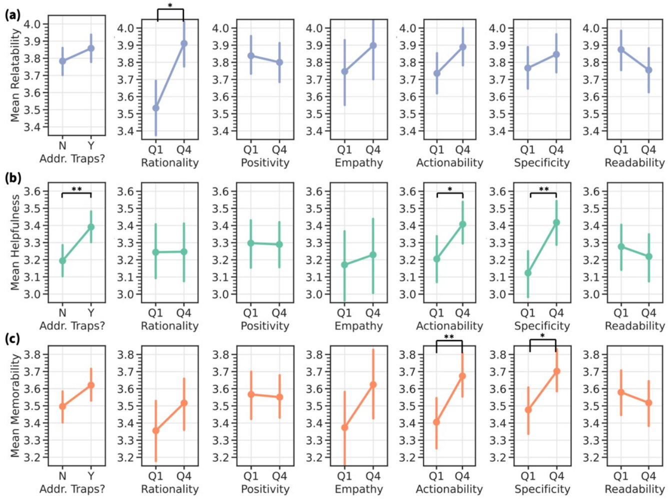
Scaling in Cognitive Modelling: a Multilingual Approach to Human Reading Times

Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction

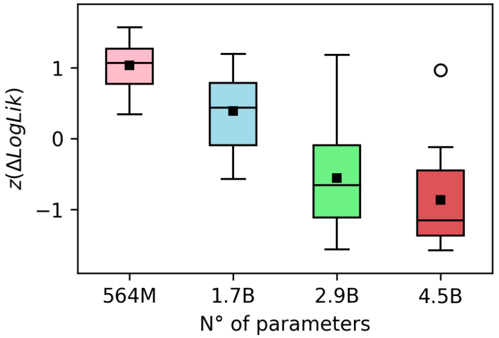
UniCoRN: Unified Cognitive Signal ReconstructioN bridging cognitive signals and human language

Shuō Wén Jiě Zì: Rethinking Dictionaries and Glyphs for Chinese Language Pre-training

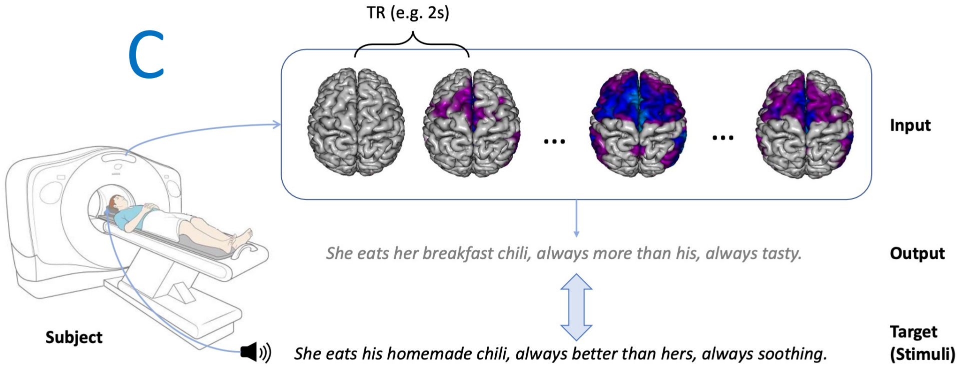
A



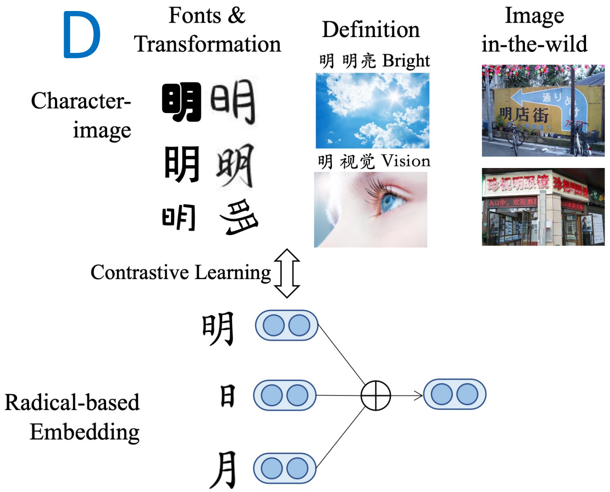
B



C



D



Study Reading Time with LLMs

de Varda and Marelli, 2023

- Processing **difficulty** of a word
- => Divergence (KL-divergence) between the probabilistic state (of the processor/mind) before and seeing a word
- => equivalent to the ***surprisal*** of a word

measured by LLM

$$surprisal(w_i) = -\log P(w_i|w_1, w_2, \ldots, w_{i-1})$$

- Existing theories: There is a logarithmic linking function between surprisal and the **cognitive cost** of word processing
- Cognitive cost: measured by eye-movement and reading time (RT)
- Question: How does the choice of different LLMs affect this relation?

```
Surprisal ~ RT
```

Reading Time

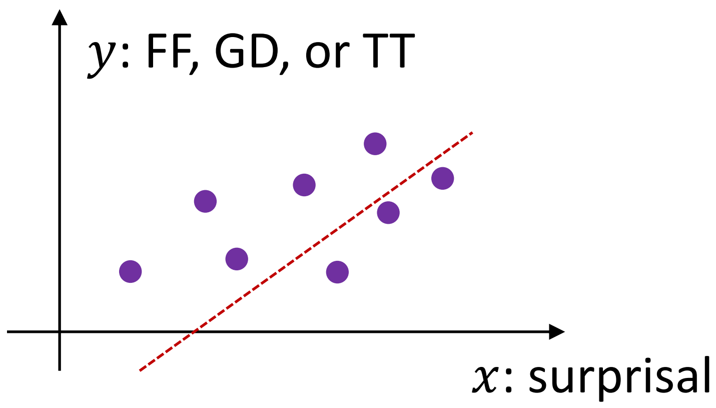
de Varda and Marelli, 2023

- Analyzed three measures of eye movement behavior for each word w_i
- **First fixation (FF)**
  - The time elapsed during the first fixation on word w_i
- **Gaze duration (GD)**
  - the sum of the fixations landing on w_i before the gaze leaves the word for the first time
- **Total reading time (TT)**
  - The total amount of time spent looking at w_i, including fixations returning to the word after having left it.

Reading Time ~ Surprisal

de Varda and Marelli, 2023

- LLM examined: XGLM -- a transformer-based decoder-only LM
- Sizes: 564M, 1.7B, 2.9B, and 4.5B
- Statistical method: fit **linear mixed-effects regression models**, using surprisal as fixed effect predictors, and the three fixation measurement (FF, GD, TT) as dependents, respectively



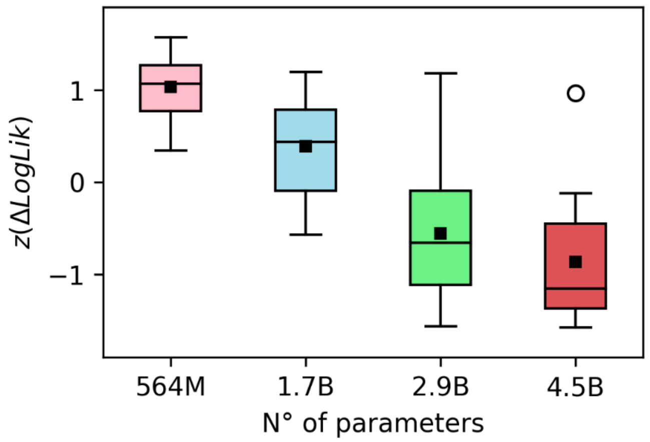
Compare
- Experimental: predictors = others + **surprisal**
- Baseline: predictors = others only

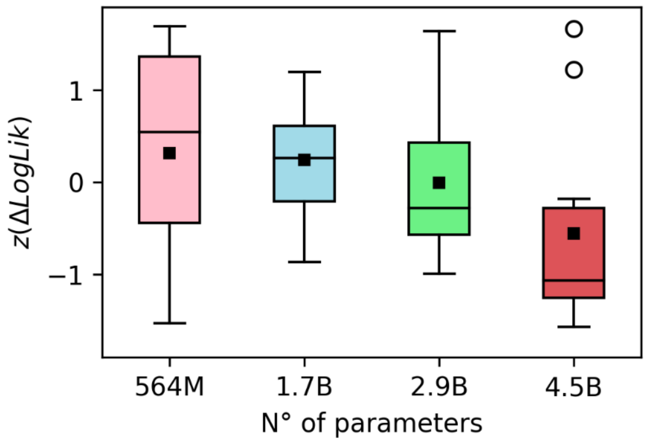
565M
1.7B
2.9B
4.5B

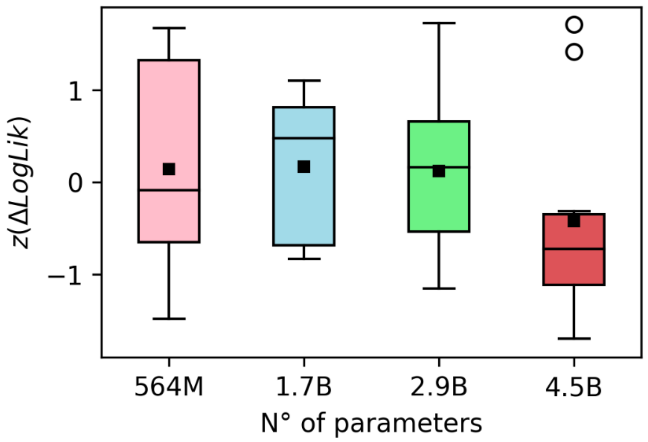
Larger $\Delta\mathrm{LogLik}$ between experimental and baseline means better predicting effect of surprisal

The mad scientist wants to know if the surprisal from larger LLMs indeed provide better $\Delta\mathrm{LogLik}$

ΔLogLik from different LLM Surprisal





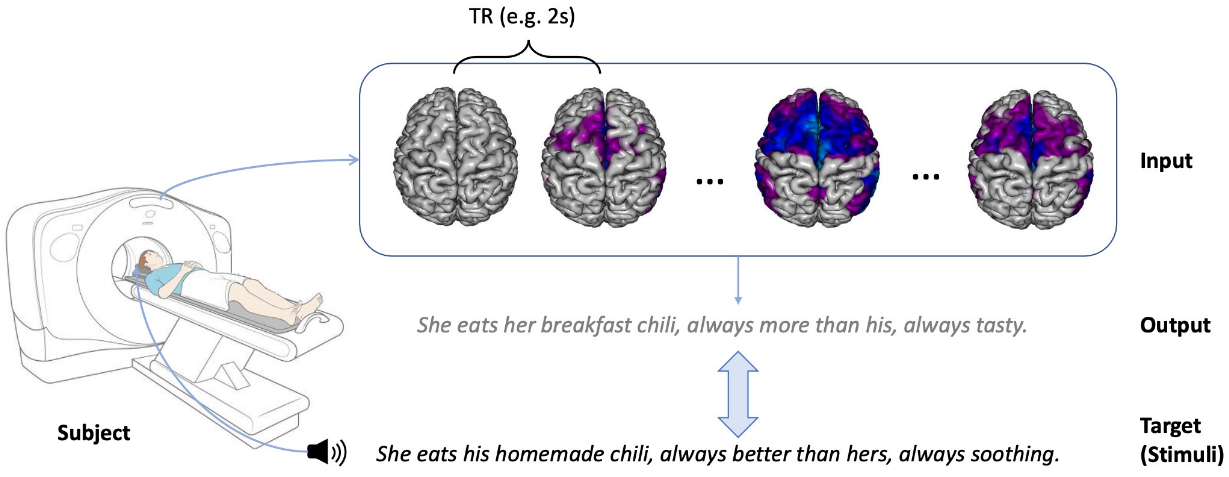


Take-home-message: Large multilingual Transformer-based models were outperformed by their smaller variants in predicting early eye movement measurements of processing difficulty

Cognitive Signal Reconstruction

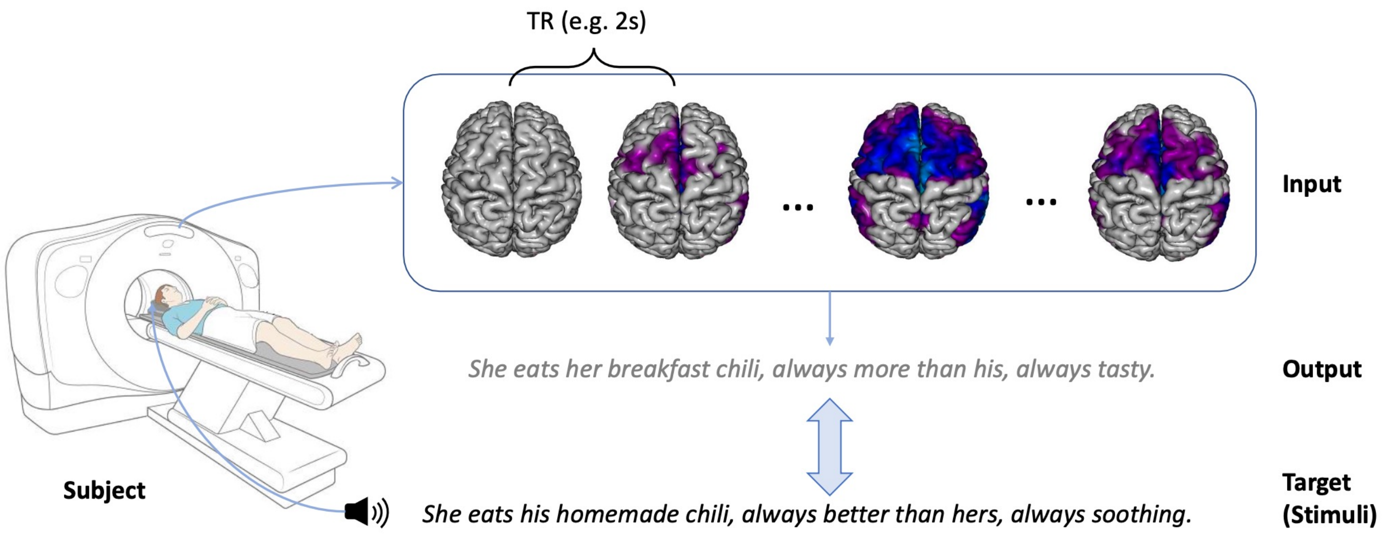
Xi et al., 2023

- Decoding text stimuli from cognitive signals (e.g. fMRI) paves the way for building versatile Brain-Computer Interface
- Task -- **fMRI2text**: the first open-vocabulary for bridging fMRI time series and human language
- Solution -- **UniCoRN**: Unified Cognitive Signal ReconstructioN for Brain Decoding -- a baseline solution leveraging a pretrained language model as decoder



fMRI2text Task Definition

- A human subject is instructed to **read** or **listen** to the text stimuli, while an fMRI volume is acquired every fixed repetition time (TR)
- Given an fMRI time series of length T, $\mathcal{F} \coloneqq \{f_1, f_2, \ldots, f_T\}$
- The task is to decode the corresponding text tokens $W \coloneqq \{w_1, w_2, \ldots, w_n\}$ of the stimuli used during the acquisition of fMRI volumes

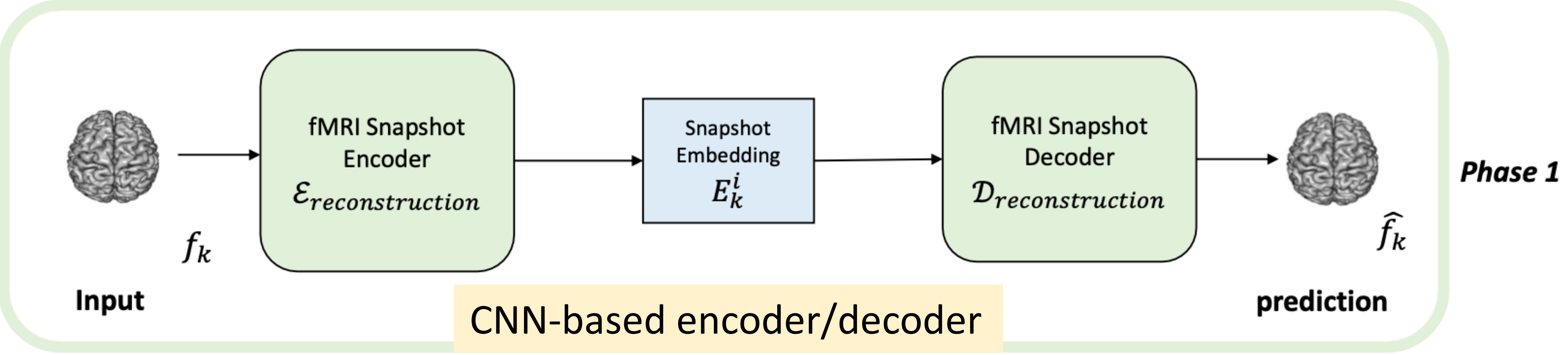


Each fMRI frame corresponds to a specific timeframe, and is aligned with an undetermined number of tokens rather than a fixed one

UniCoRN

- Two stages (three phases):
- 1. Cognitive signal reconstruction to train the cognitive signal encoder
  - Phase 1: Snapshot reconstruction
  - Phase 2: Series reconstruction
- 2. Cog2text decoding to convert the embeddings of the cognitive signals to human language
  - Phase 3: Generate text tokens

Phase 1 & 2 -- Learning encoders



Learn snapshot encoder through reconstruction task -- to reconstruct the original fMRI frame f_k

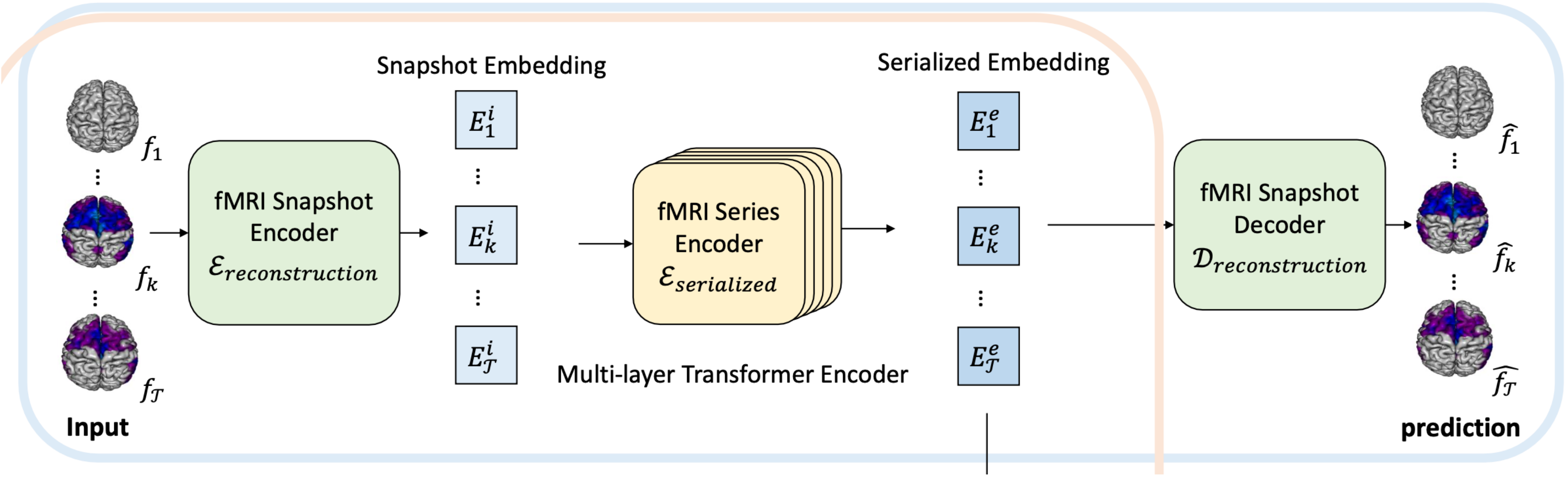
$$E_k^i = \mathcal{E}_r(f_k)$$

$$\mathcal{E}_r = \arg\min_{\mathcal{E}} \mathrm{MAE}(\mathcal{D}_r(\mathcal{E}_r(f_k)), f_k)$$

using mean average error (MAE) loss

Phase 1 & 2 -- Learning encoders

- Phase 2 is similar to phase 1, but adding a multi-layer transformer encoder $\mathcal{E}_s$ to further encode the series of fRMI embeddings E_k^i from phase 1
- Trained with same reconstruction task

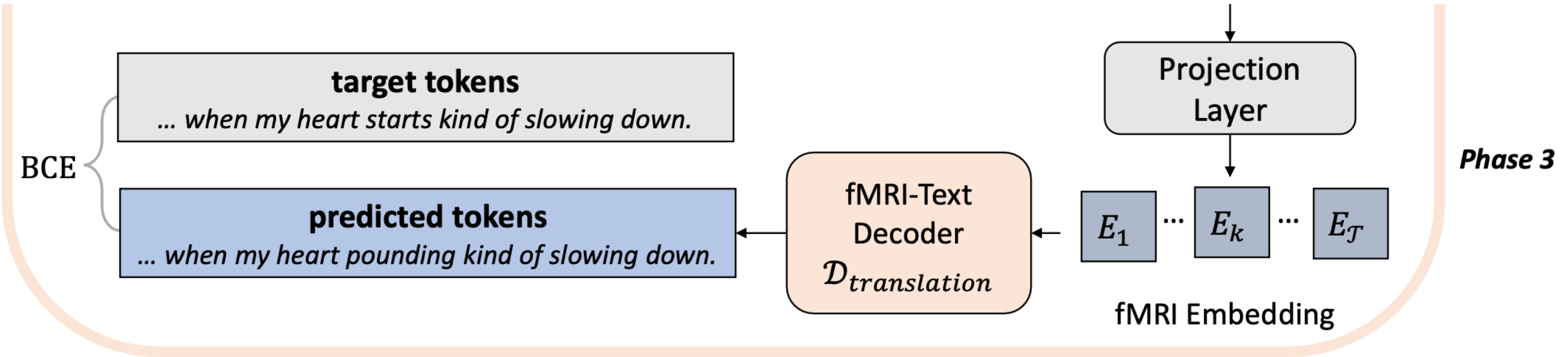


$$E^e_{k\sim\mathcal{T}} = \mathcal{E}_s(E^i_{k\sim\mathcal{T}})$$

$$\mathcal{E}_s = \arg\min_{\mathcal{E}} \mathrm{MAE}(\mathcal{D}_r(\mathcal{E}_s(E^i_{k\sim\mathcal{T}})), E^i_{k\sim\mathcal{T}})$$

Phase 3 -- Cog2text decoding

- High-level idea: consider each original frame of fMRI as a word-level representation of “the foreign language spoken by the human brain”
- Similar to machine translation task



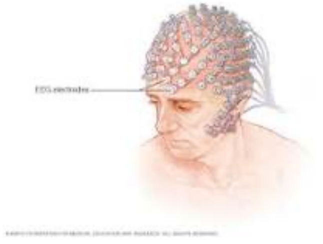
Decoding Effect

Split Method	$\mathcal{T}$	Results
consecutive time	10	T: the policeman, um, he **doesn't** even **say** anything to **Sherlock**... P: and first, the, she just **doesn't** *talk* though **Sherlock**...
by stimuli	10	T: I think it's **some sort of** mass hyp**nosis** or something... P: and you *a sort of* the You**nosis** session something...
random time	1	T: He woke up early the **next morning** P: I's up and **morning** *other day*
random time	3	T: she put her **arm** through mine and squeezed it a little bit. P: I says her *shoulder* through mine and I it a little bit
random time	5	T: Um, it was an **extremely Darwinian moment** for me, uh, **because**... P: I and, like *best* **Darwinian moment** for me, and, *for*...

fMRI2text task

(1)	T: Stephen Rea, **Aidan Quinn**, and Alan Bates play Desmond's legal **eagles**... P: He Hara, **Aidan Quinn**, and Alan Bates play Desmond's legal **eagles**... B: He Baldwina, *Longan* shows, and Alan Lloyd play Hannibal's legal eternally...
(2)	T: the **sight** of this grandiloquent quartet **lolling** in pretty Irish settings is a **pleasant** enough **thing** P: the **sight** of this grandiloquent Shet **lolling** in pretty Irish American is a lot enough **thing** B: the real of this this asquent Shet filmolling's grand much American is a talented enough *film*

EEG2text task

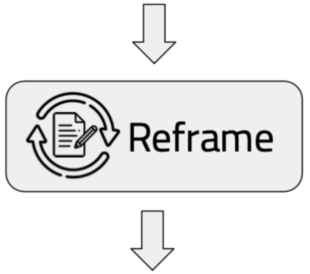


Cognitive Reframing of Negative Thoughts

- Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction, Sharma et al., 2023
- **Background**: Negative thoughts are a natural part of human cognition. However, for people experiencing mental health challenges, such thoughts are often entrenched, automatic and emotionally triggering, making it difficult to overcome them in-the-moment
- A therapeutic intervention to overcome negative thoughts is **Cognitive Reframing**, in which a negative thought is replaced with a more hopeful "reframed thought", which offers an alternative perspective on one's situation (Beck 1976)

Situation
I participated in a hackathon and I lost

Thought
I'll never become a successful programmer



"I may not become the most successful programmer, but I will keep trying"

Task Definition

- How to develop cognitive reframing methods that automatically generate relatable, helpful and memorable reframed thoughts
- **Task** of *Cognitive Reframing*: Given a situation S_i and a negative thought T_i, the task is to generate a reframed thought R_i

- Start by curating (准备) data for situations and negative thoughts ($\boldsymbol{S}_i$ and $\boldsymbol{T}_i$)
  - 180 pairs of diverse situations with negative thoughts from Thought Records Dataset (Burger et al., 2021)
  - Manually curate 120 pairs of self-reported situations and thoughts from Mental Health America (MHA)
- Recruit 15 current mental health practitioners and clinical psychology graduate students to **write reframed thoughts ($\boldsymbol{R}_i$)**

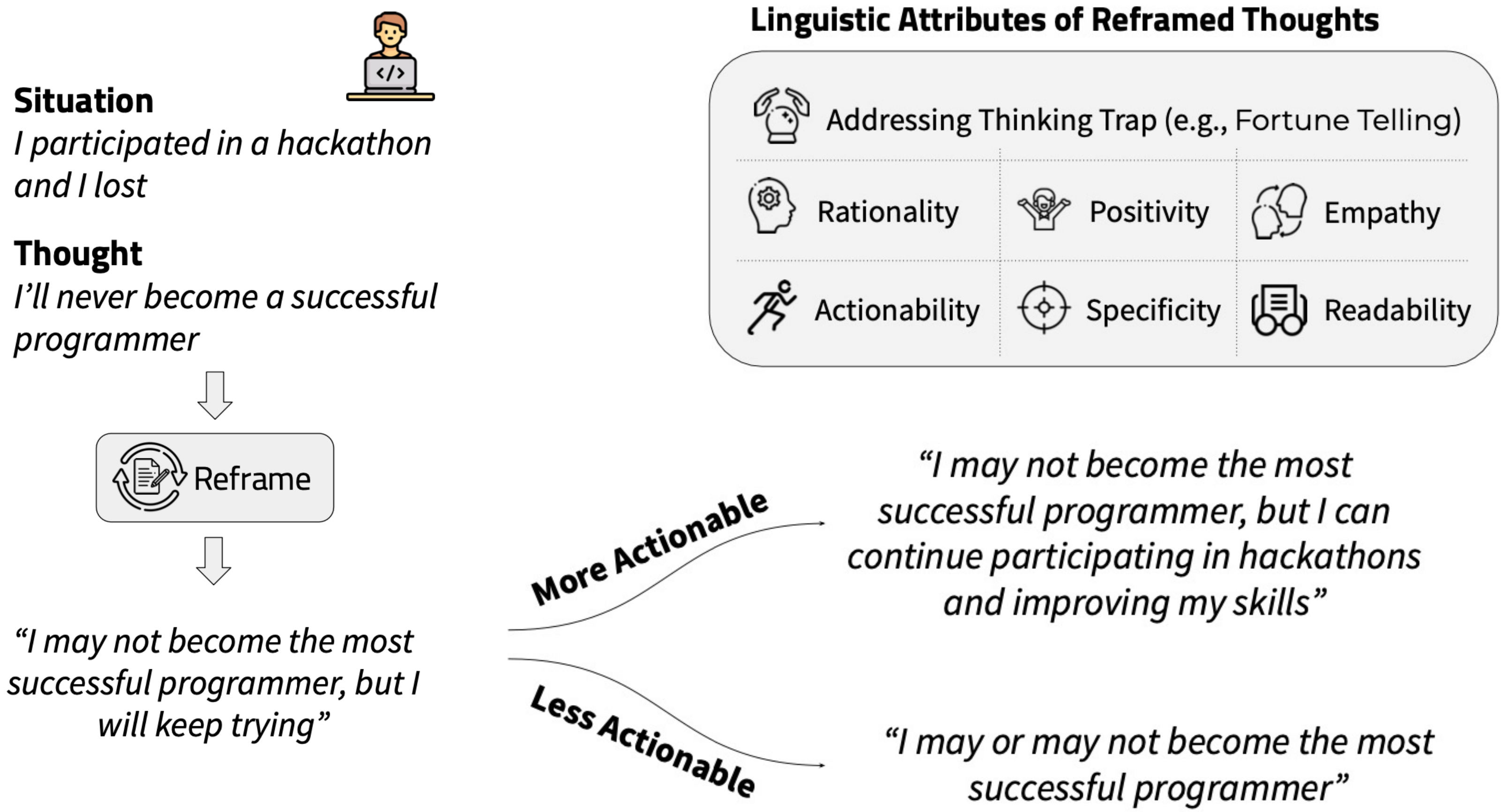
Measuring Reframing Attributes

- Addressing Thinking Traps 避免思维陷阱
- Rationality 理性
- Positivity 正向
- Empathy 共情
- Actionability 可执行
- Specificity 具体
- Readability 可读

Reframe Generation

- **Challenge**: a simple in-context learning (few-shot) method with a fixed set of examples often failed to appropriately reframe situations and thoughts, for which no relevant in- context examples were provided
- **Solution**: Retrieval-based in-context learning
- For each situation $\boldsymbol{S}_i$ and negative thought $\boldsymbol{T}_i$, we retrieve k-similar examples from the annotated dataset
  - using the cosine similarity scores from RoBERTa embeddings
- Then use those examples to prompt GPT-3 to generate new reframes $\boldsymbol{R}_i$
- Generation with controllable linguistic attributes, e.g., rationality, empathy, etc.

Controlling Reframing Attributes



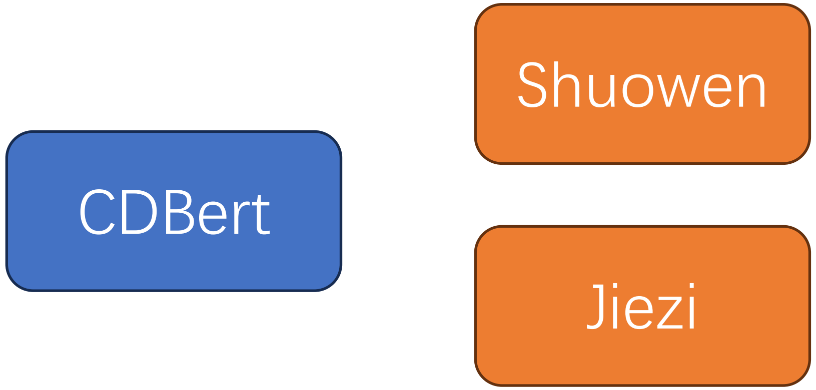
Shuowen and Jiezi

Wang et al., 2023

- Rethinking ***Dictionaries*** and ***Glyphs*** for Chinese Language Pre-training

n. 字形；图象字符；

- **Summary**: CDBert -- A new learning paradigm that enhances the semantics understanding ability of the Chinese Pretrained Language Models with dictionary knowledge and structure of Chinese characters.

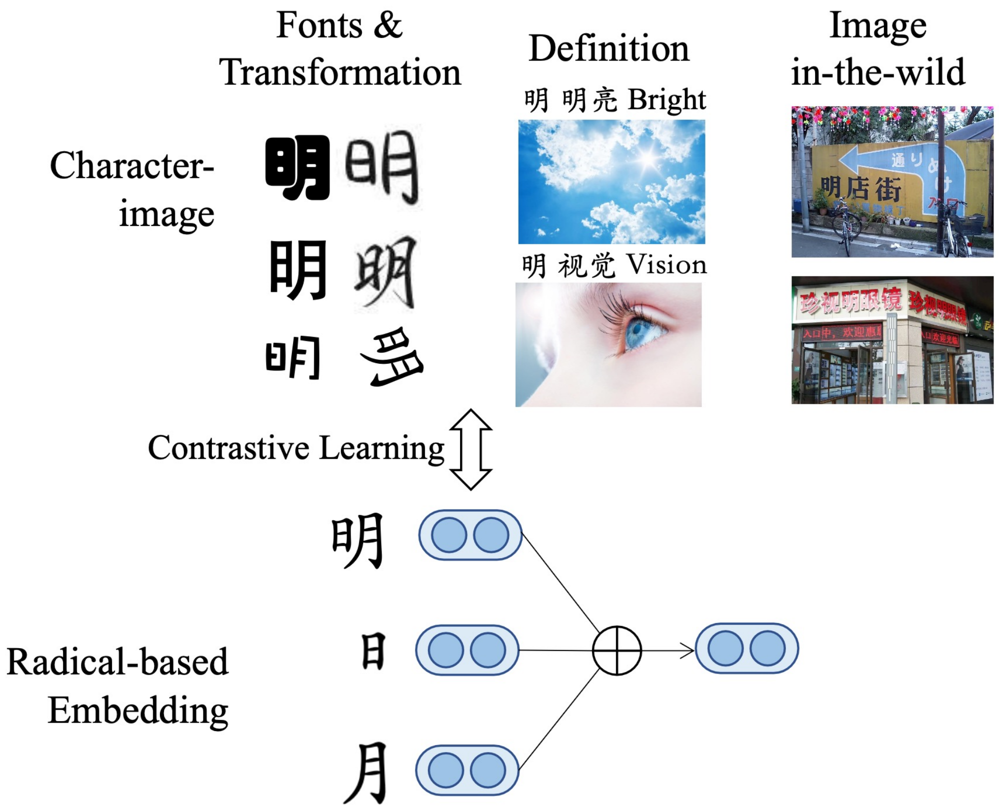


retrieving the most appropriate meaning from Chinese dictionary

enhancing the characters' glyph representations with structure understanding

Jiezi (解字) part

Wang et al., 2023



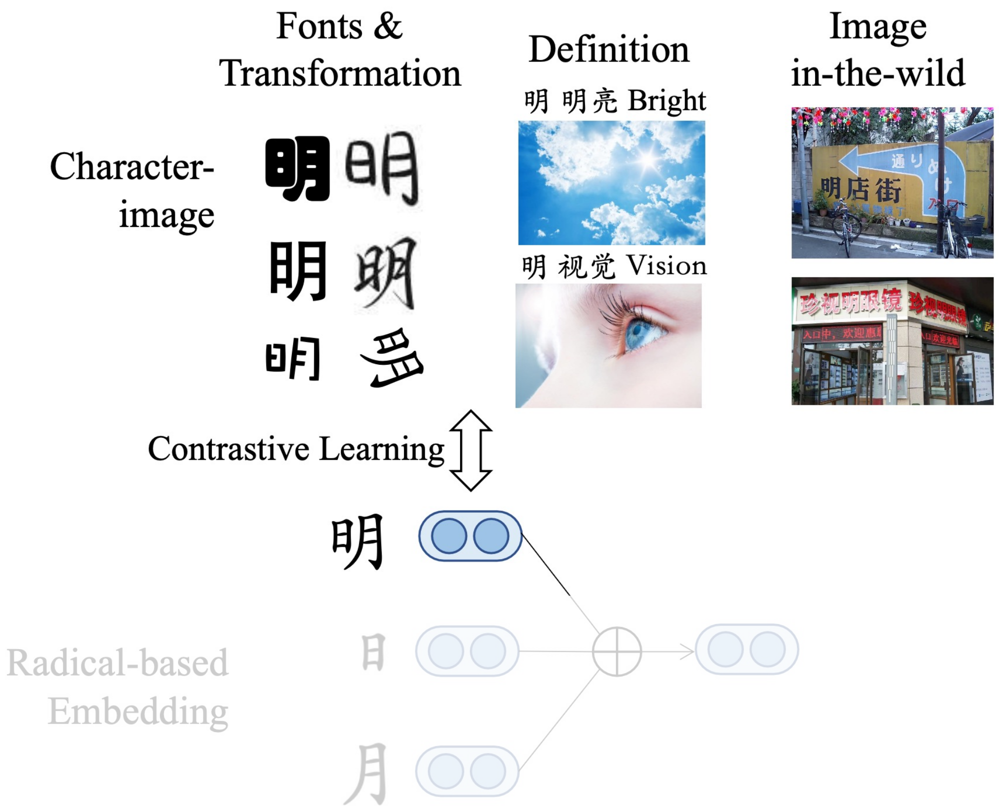
Two structured learning strategies to capture the semantics of Chinese characters:

- **CLIP** enhanced glyph representation
- **Radical-based** character embedding (偏旁部首)

CLIP enhanced glyph representation

Wang et al., 2023

- Learn glyph representations through contrastive learning

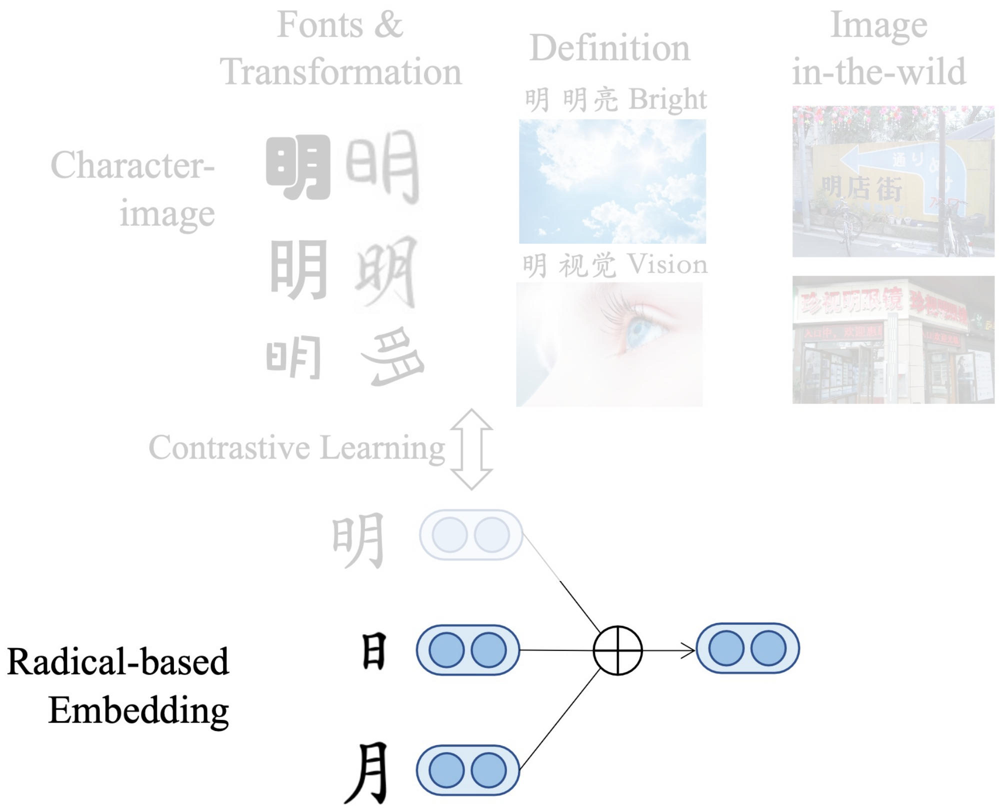


- Concatenate character *c* with its definition *def* as text input and generate a **picture of the character as visual input**
  (Model initialized with a pretrained Chinese-CLIP model, Yang et al., 2022)

- Generate **images of characters** by transformation: font, size, direction etc.
  -- to alleviate pixel-level noise

- Introduce some Chinese character images in wild
  -- improve model robustness

Radical-based character embedding



- Radical-based embedding for end-to-end pretraining
- Add a radical vocabulary
- Add the radical embedding for each character
- E.g. $e(\text{明}) = e(\text{日}) + e(\text{月})$

Shuowen (说文) part

Wang et al., 2023

- Dictionary as a pretrained knowledge
- **Three steps** for looking up the dictionary:
- 1) Masked Entry Modeling (**MEM**)
  -- To clarify the meaning of entry
- 2) Contrastive Learning for Synonym and Antonym (**CL4SA**)
  -- Deal with ambiguous meanings
- 3) Example Learning (**EL**)
  -- learn more accurate meanings through classical examples

Masked Entry Modeling (MEM)

- Similar to Masked Language Modeling (MLM)
- Concatenate the entry `<ent>` to its meaning or definition `<def>`
  - **Input**: `{[CLS] <ent> [SEP] <def> [SEP]}`

- Then replace `<ent>` with `[MASK]` and attempts to predict it
- $\mathcal{L}_{\mathrm{MEM}}$ = the cross entropy between the predicted entry and ground truth

Contrastive Learning for Synonym and Antonym (CL4SA)

- Create positive pairs `<ent, syno>`, with synonyms from dictionary and negative pairs `<ent, anto>` with antonyms
- **Goal**: to push positive pairs closer and pushing negative pairs further

$$\mathcal{L}_{cl4sa} = -\log \frac{e^{h_{ent} \cdot h_{syno}}}{e^{h_{ent} \cdot h_{syno}} + e^{h_{ent} \cdot h_{anto}}}$$

Tips: Since the antonyms in the dictionary are much less than synonyms, we randomly sampled entries from the vocabulary for compensation.

Example Learning (EL)

- Deal with polysemy (一词多义): most characters/words have more than one meanings/definitions
- Use **Example Learning (EL)** to learn the weight of different definitions for a certain example
- Entry <ent> with K definitions: $def_1, \ldots, def_K$; each meaning def_i has an exempler exa_i
- Use h_{exa} the hidden state of the [CLS] token in exemplar as query Q, and $X = \{h_m^i\}_{i=1}^{K}$ the hidden states of the [CLS] token in the definitions as key K
- Compute the attention score

$$Attn_{def} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

Final loss:

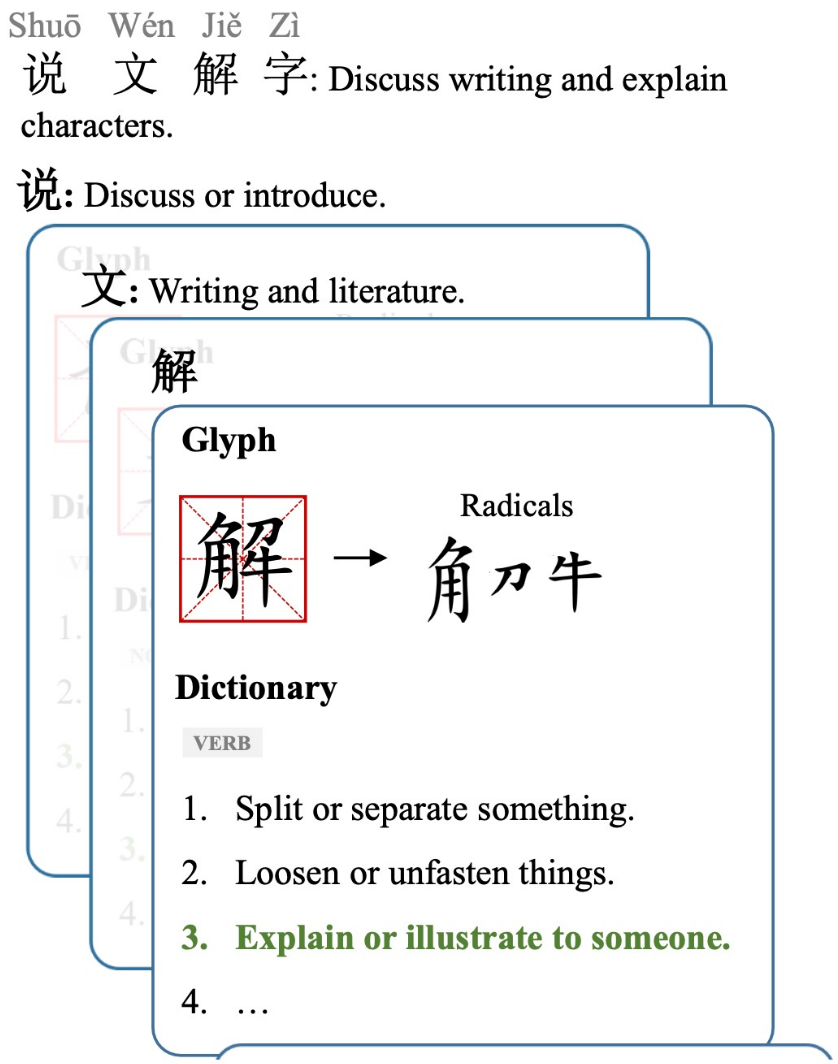
$$\mathcal{L} = \lambda_1 \mathcal{L}_{mem} + \lambda_2 \mathcal{L}_{cl4sa} + \lambda_3 \mathcal{L}_{el}$$

learn the correct meaning through loss:

$$\mathcal{L}_{el} = CrossEntropy(\text{one-hot}(def), Attn_{def})$$

Shuowen + Jiezi

Wang et al., 2023



The expression in green refers to the selected definition of current character

Dataset

- Dictionary data: https://github.com/mapull/chinese-dictionary

```json
[
    {
        "char": "车",
        "pronunciations": [
            {
                "pinyin": "chē",
                "explanations": [
                    {
                        "content": "(象形。甲骨文有多种写法。象车形。本义:车子,陆地上有轮子的运输工具)。"
                    },
                    {
                        "content": "同本义。",
                        "detail": [
                            {
                                "text": "車,舆论之总名。夏后时奚仲所造。象形。",
                                "book": "《说文》。按,横视之肖,或云车少昊时驾牛,奚仲始驾马。"
                            },
                            {
                                "text": "为车。大车、柏车、羊车,皆两辕,驾牛;田车、兵车、乘车,皆一辀,驾马。
                                "book": "《考工记·舆人》"
                            },
                            {
                                "text": "车从马。",
                                "book": "《左传·闵公元年》"
                            }
                        ],
```

Recap

- Scaling in Cognitive Modeling -- Reading Time
- Reconstruct Language from Cognitive Signal
- Cognitive Reframing of Negative Thoughts
- Shuowen and JieZi

- The next idea is from you!