CS310 Natural Language Processing
自然语言处理
Lecture 13 - Cognitive Sciences and Language

Instructor: Yang Xu

主讲人：徐炀

xuyang@sustech.edu.cn

Overview

- What is CogSci?
- Why CogSci matters to NLP?
- CogSci 101
- Topics in CogSci in relation to Language -- Case studies

What is CogSci?

- Cognitive science is the study of the human **mind** and **brain**, focusing on how the mind represents and manipulates knowledge and how mental representations and processes are realized in the brain.
  (JHU CogSci Dept. website https://cogsci.jhu.edu/about/)

- **Mind**: intangible, abstract, metaphysics (形而上)
- **Brain**: tangible, concrete, “physics” (形而下)

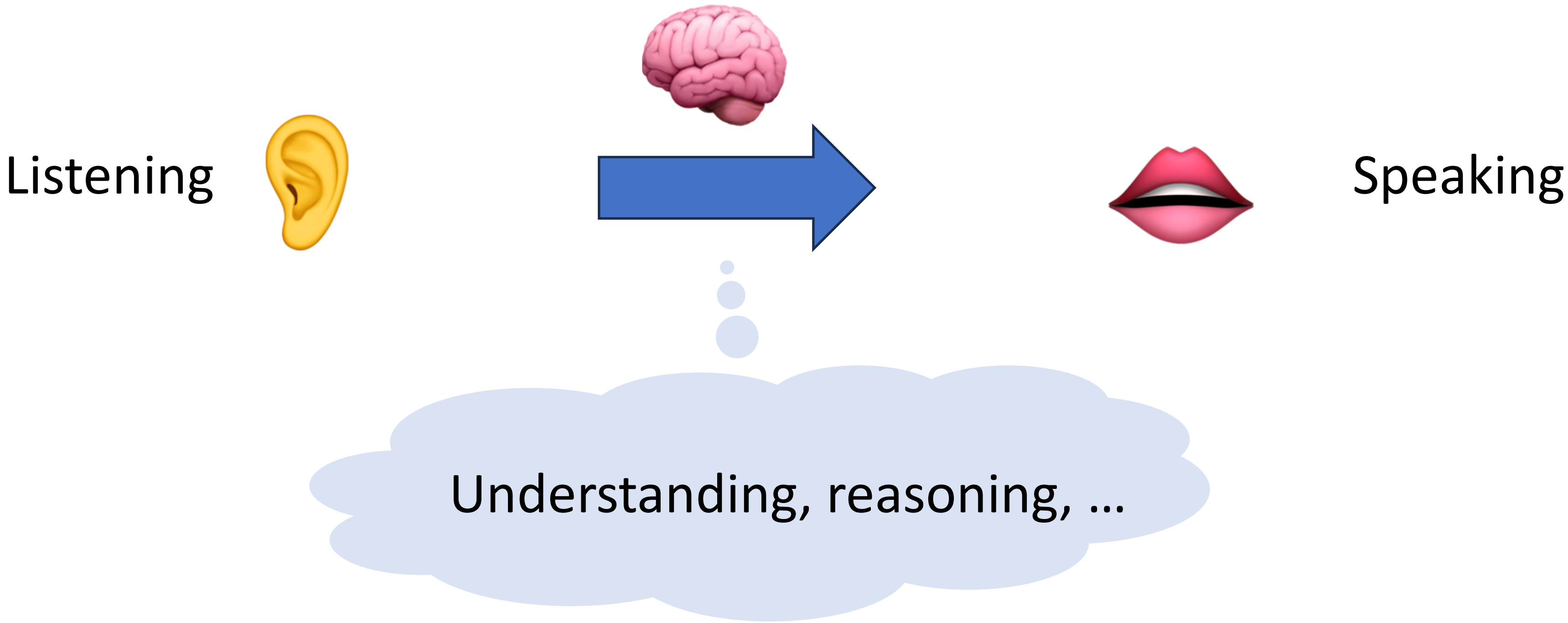
What is CogSci? --Wikipedia

- Cognitive science is the interdisciplinary, scientific study of the mind and its processes … (what)
- … include language, perception, memory, attention, reasoning, and emotion … (what specifically)
- … borrow from fields such as linguistics, psychology, artificial intelligence, philosophy, neuroscience, and anthropology (How)

Source: https://en.wikipedia.org/wiki/Cognitive_science

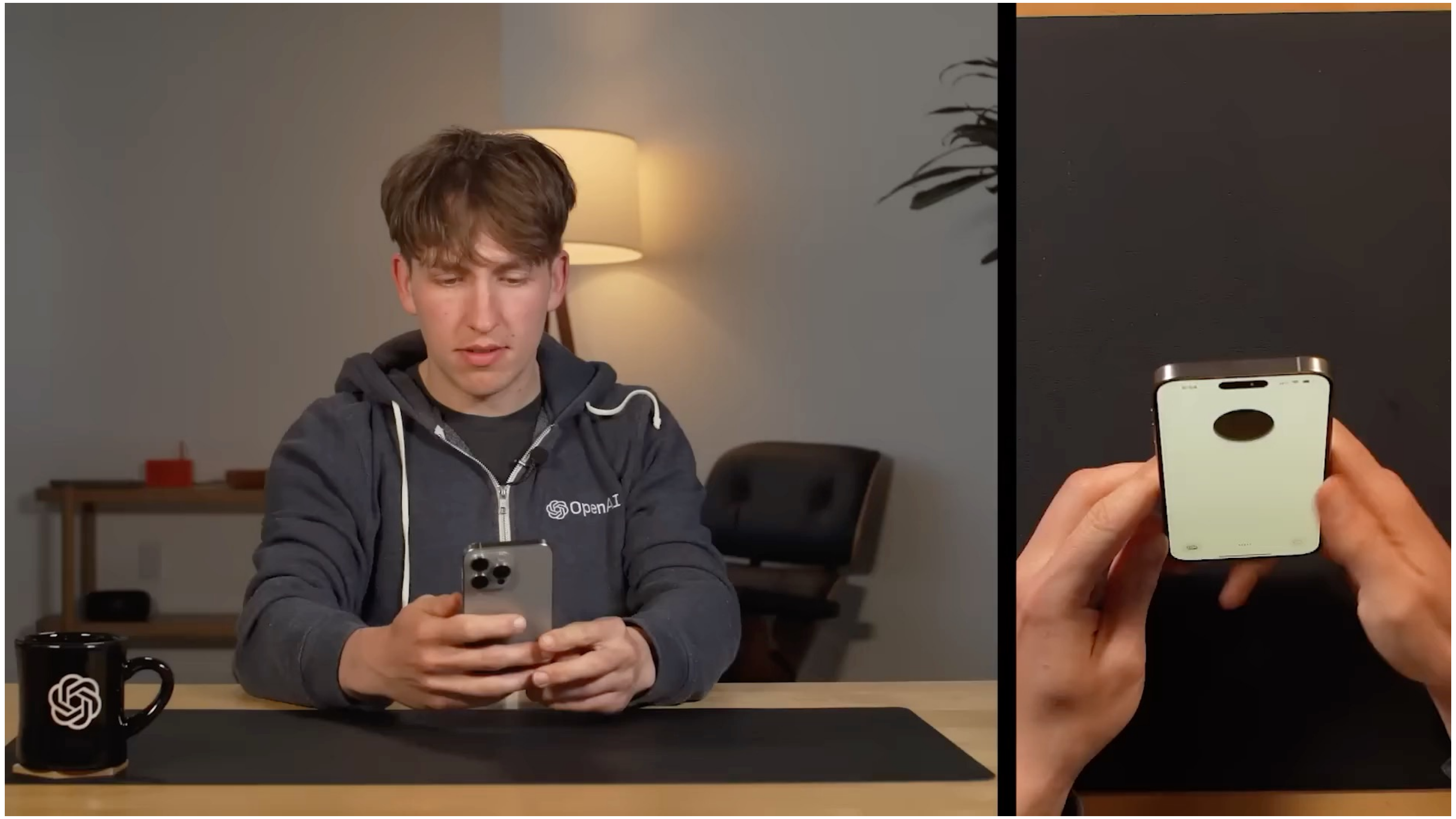
Why CogSci Matters to NLP?

- Language is not just about “listening” and “speaking”



Why CogSci Matters to NLP?

GPT-4o released in May 2024



话轮转换

Turning-taking, which is common in human-to-human conversations

35'': AI stopped speaking to wait for the human

51'': AI stopped to wait (after hearing "*it is …*")

Link: https://youtu.be/vgYi3Wr7v_g

Why CogSci Matters to NLP?

Turning-taking, which is common in human-to-human conversations

GPT-4o released in May 2024

3’15’’: Male AI tried to interrupt but quickly stopped

Link: https://youtu.be/MirzFk_DSiI

Overall Picture of Cognitive Science (Sciences)

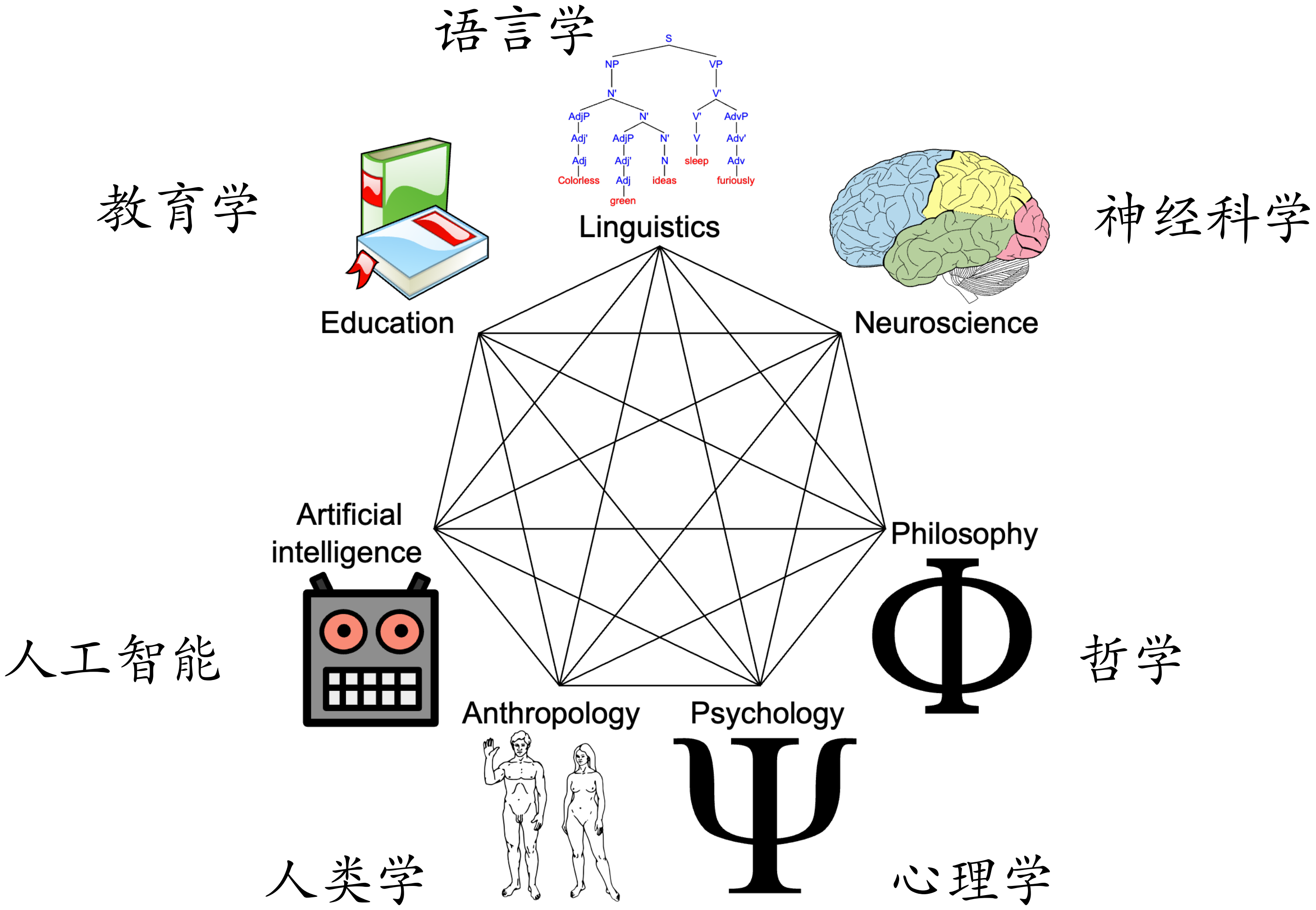
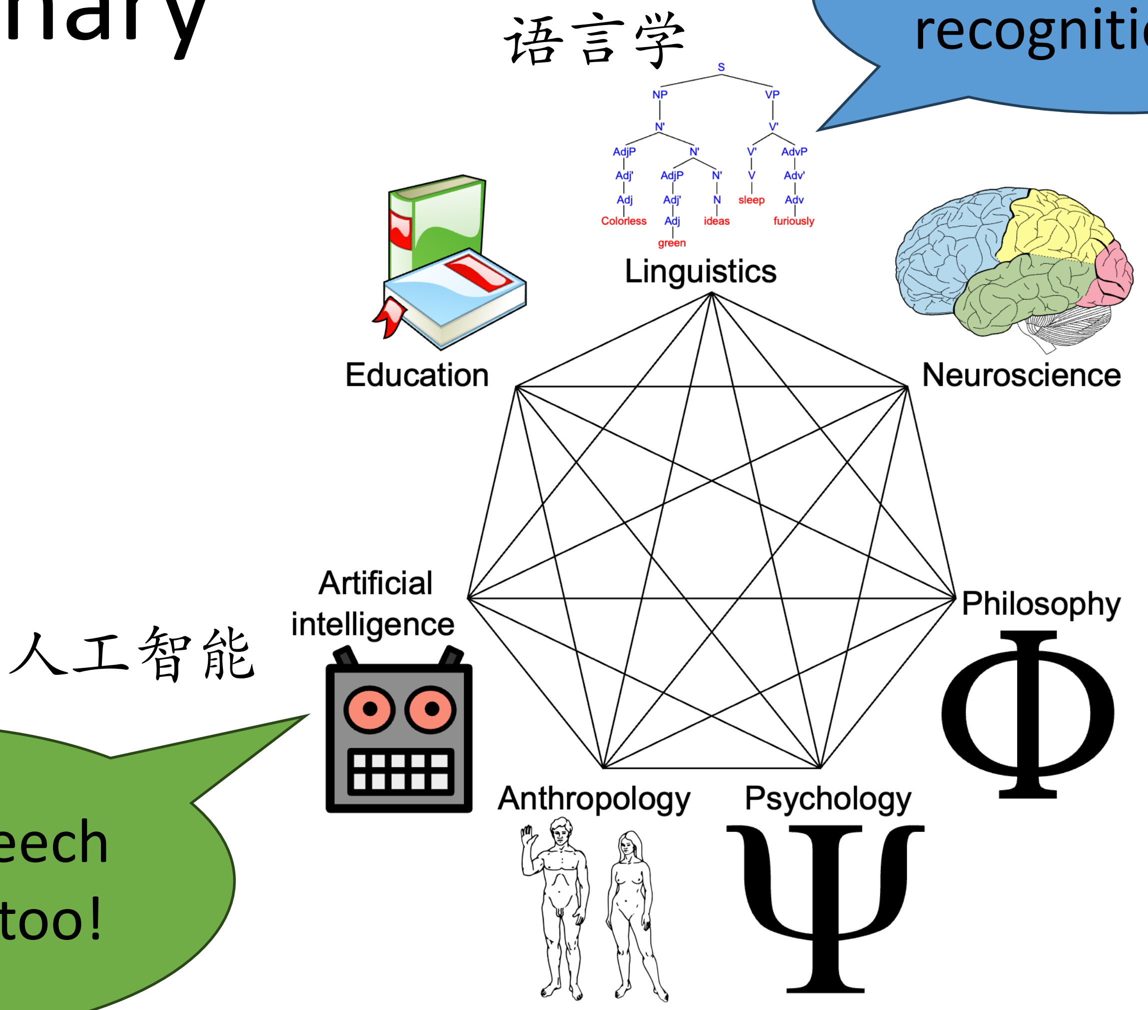


Figure source: https://en.wikipedia.org/wiki/Cognitive_science

CogSci is Hyper Interdisciplinary

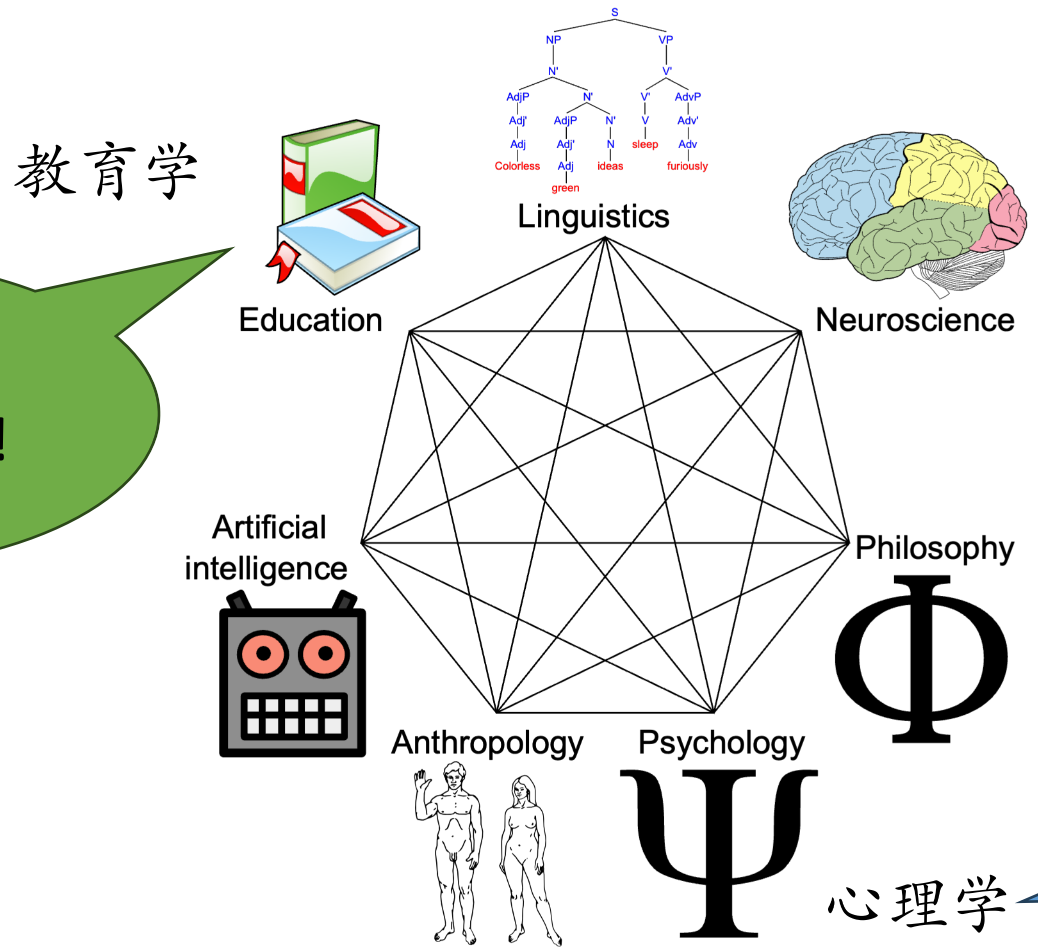
We study speech recognition!



We study speech recognition, too!

CogSci is Hyper Interdisciplinary

We study how humans study!



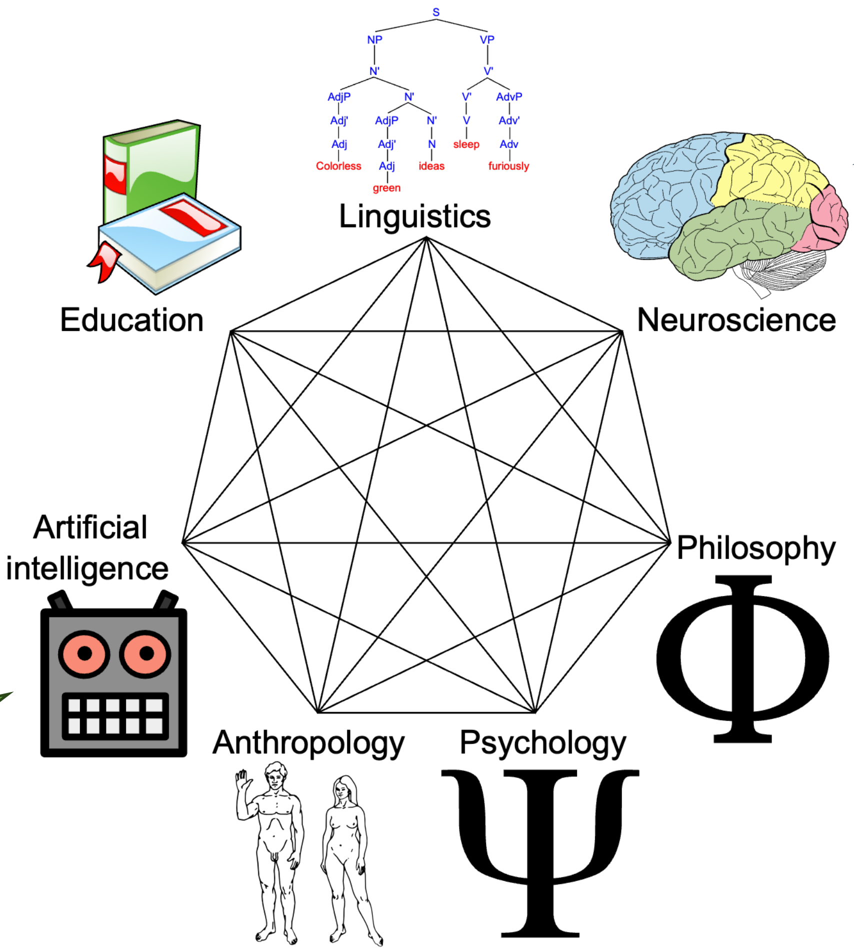
We study how humans study, too!

CogSci is Hyper Interdisciplinary

We study the structure and functions of human brain

神经科学

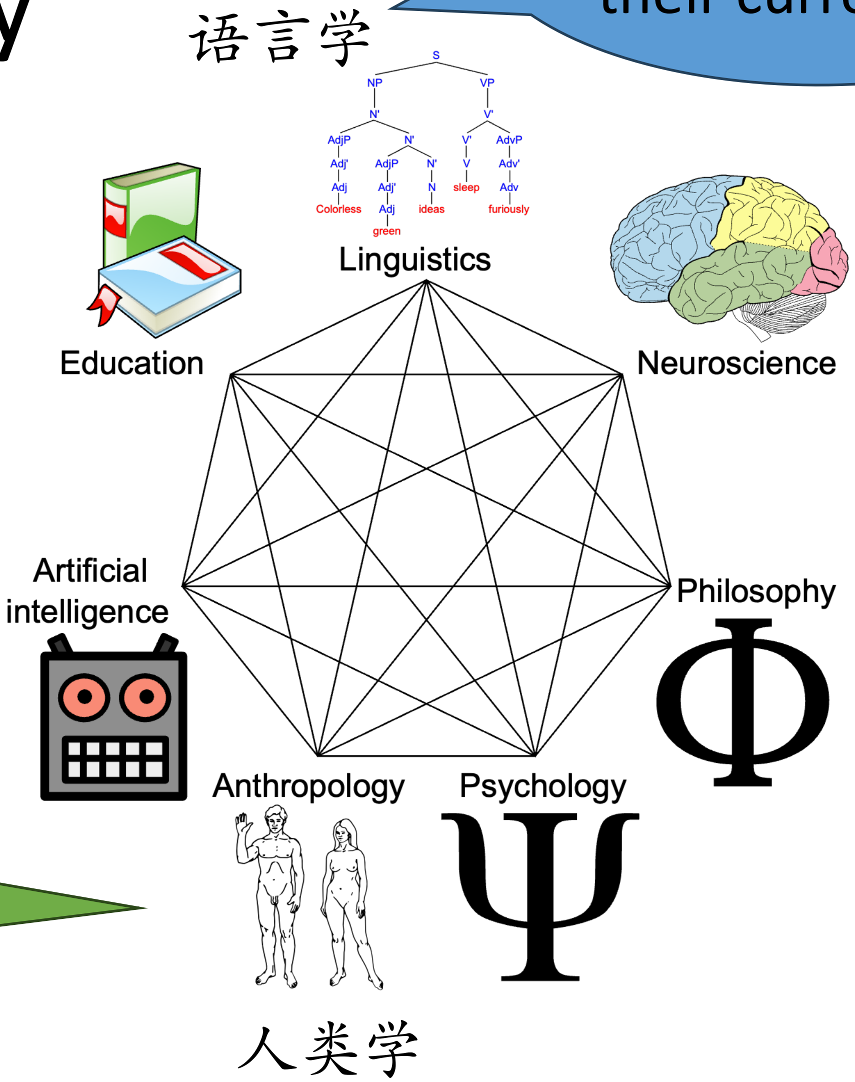
人工智能



Our tools are inspired by neural cells in brain

CogSci is Hyper Interdisciplinary

We study how languages become their current forms



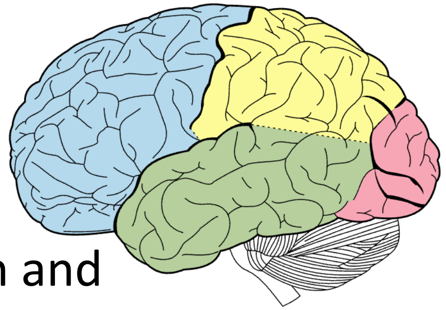
We study the symbolic systems of mankind as well

CogSci in Textbooks

- 认知基础，史忠植，2021，机械工业出版社

第1章绪论 Neural basis
第2章脑认知的神经基础
第3章心理表征 Representation
第4章视觉和注意 Visual perception and attention
第5章听觉和言语
第6章认知语言学 Audio perception and speech/language
第7章学习

Learning Cognitive linguistics



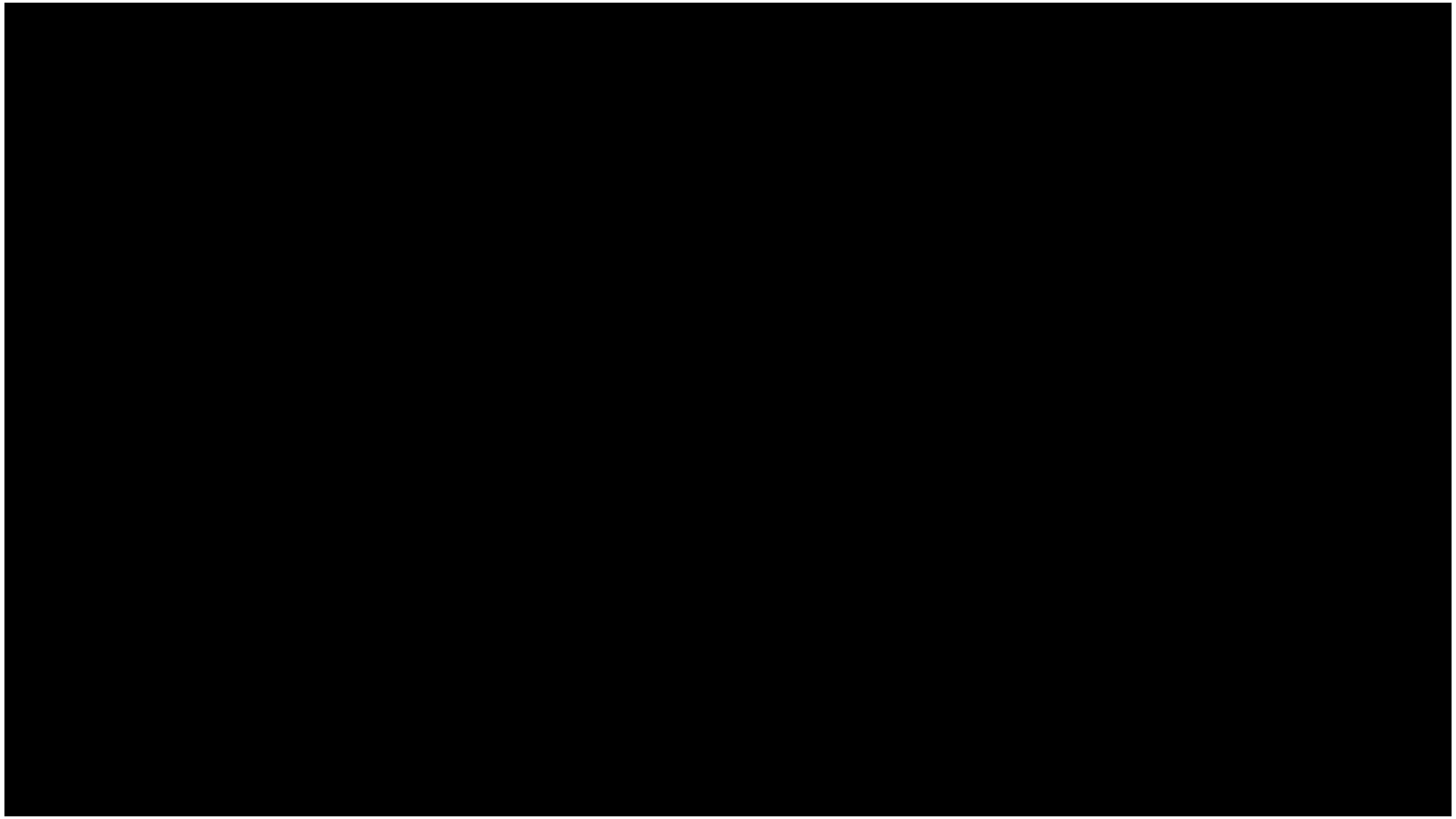
Memory
第8章记忆机理 Though, reasoning, decision making
第9章思维、推理和决策。
第10章-智力发展 Development
第11章-情绪问题 Emotion
第12章-意识 Consciousness
第13章-认知模型 Cognitive modeling
第14章-认知模拟

Cognitive simulation

Example: Attention

Daniel J. Simons

Link: https://youtu.be/IGQmdoK_ZfY

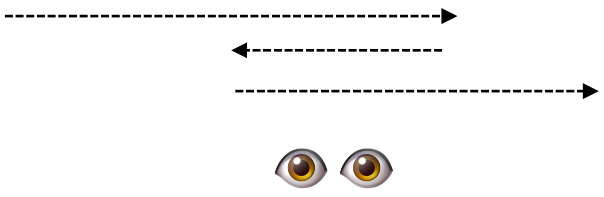


What if we ask AI models to watch this video and ask the same questions?

Example: Garden-path sentences

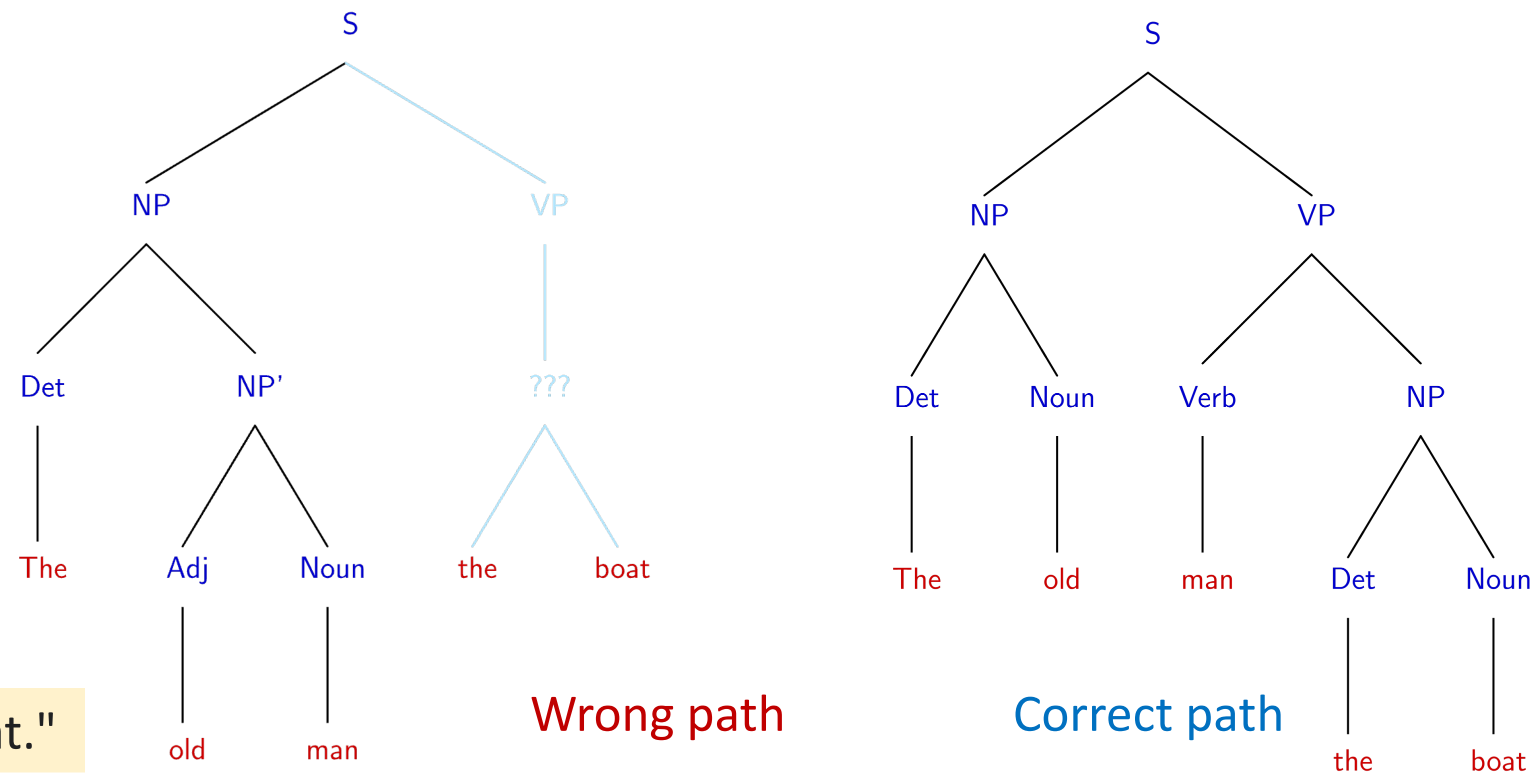
- A **garden-path** sentence is a grammatically correct sentence that starts in such a way that a reader's most likely interpretation will be incorrect

"The old man the boat."



Eye movement patterns

"The old are those who man the boat."



Example: Garden-path sentences

- More examples

"The complex houses married and single soldiers and their families."

"The complex provides housing for the soldiers, married or single, as well as their families."

"The horse raced past the barn fell."

"The horse that was raced past the barn fell."

Example: Garden-path sentences

- Chinese examples

(1) 咬死猎人的狗逃跑了

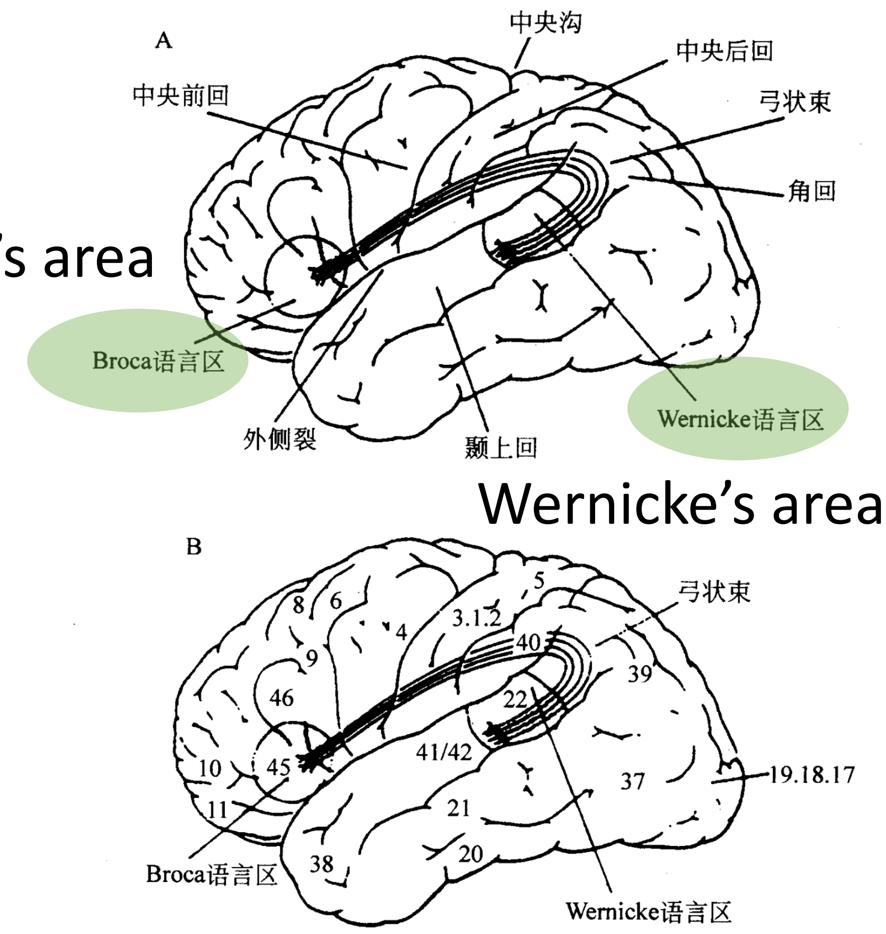
(2) 咬死猎人的狗是熊逃跑的唯一出路

Who was killed?

Example: Language Comprehension and Brain

Parietotemporal occipital

Broca's area



Wernicke's area

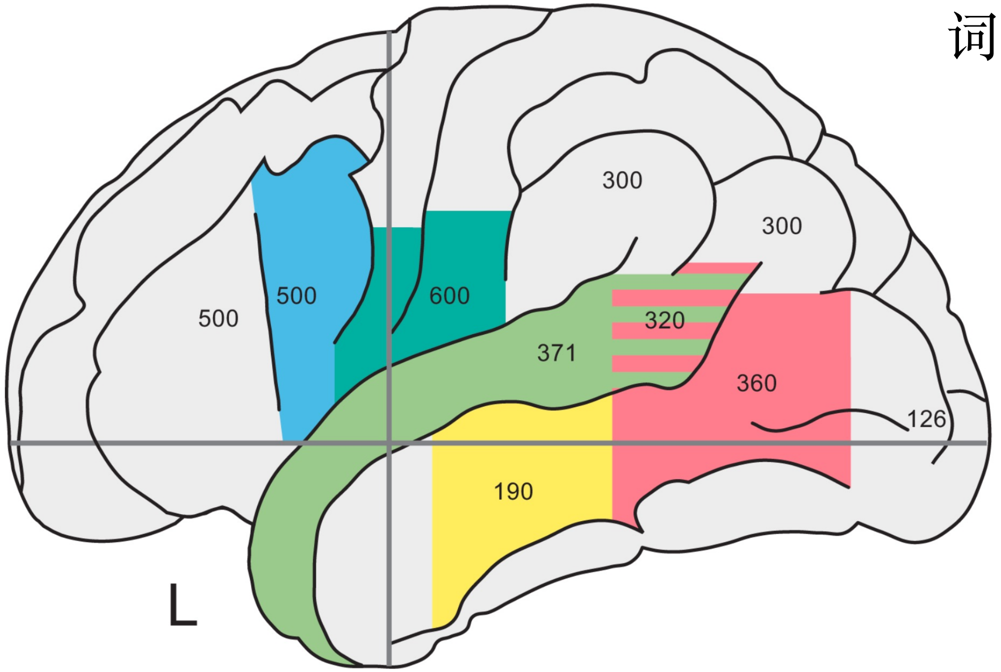
- 声音信号在听觉系统被转换，信息传递到以角回为中心的顶颞枕联合皮质，再传递到 **Wernicke**区，在这里从语音信息中提取出单词表征 (word representation)。
- 信息流从 Wernicke 区经过弓形束(白质神经束)到达 Broca 区，这里是语法特征 (syntactic representation) 记忆之所。

white matter pathways

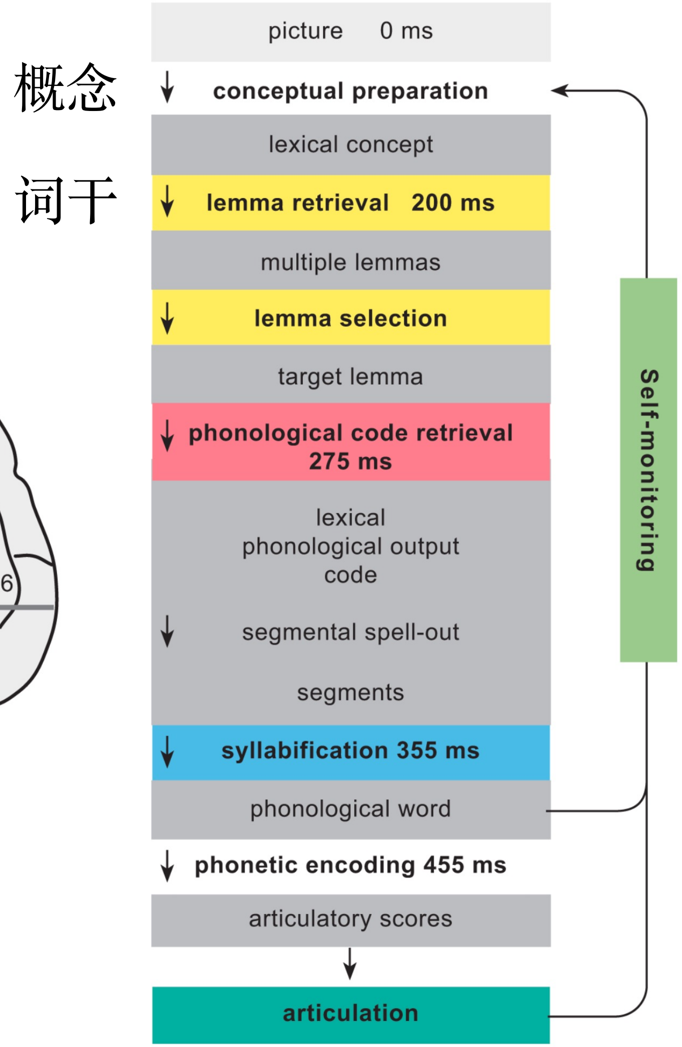
- 接着单词表征激活概念中心相关的概念 (concept)。

这样，听觉理解就发生了。

Example: Language Production and Brain



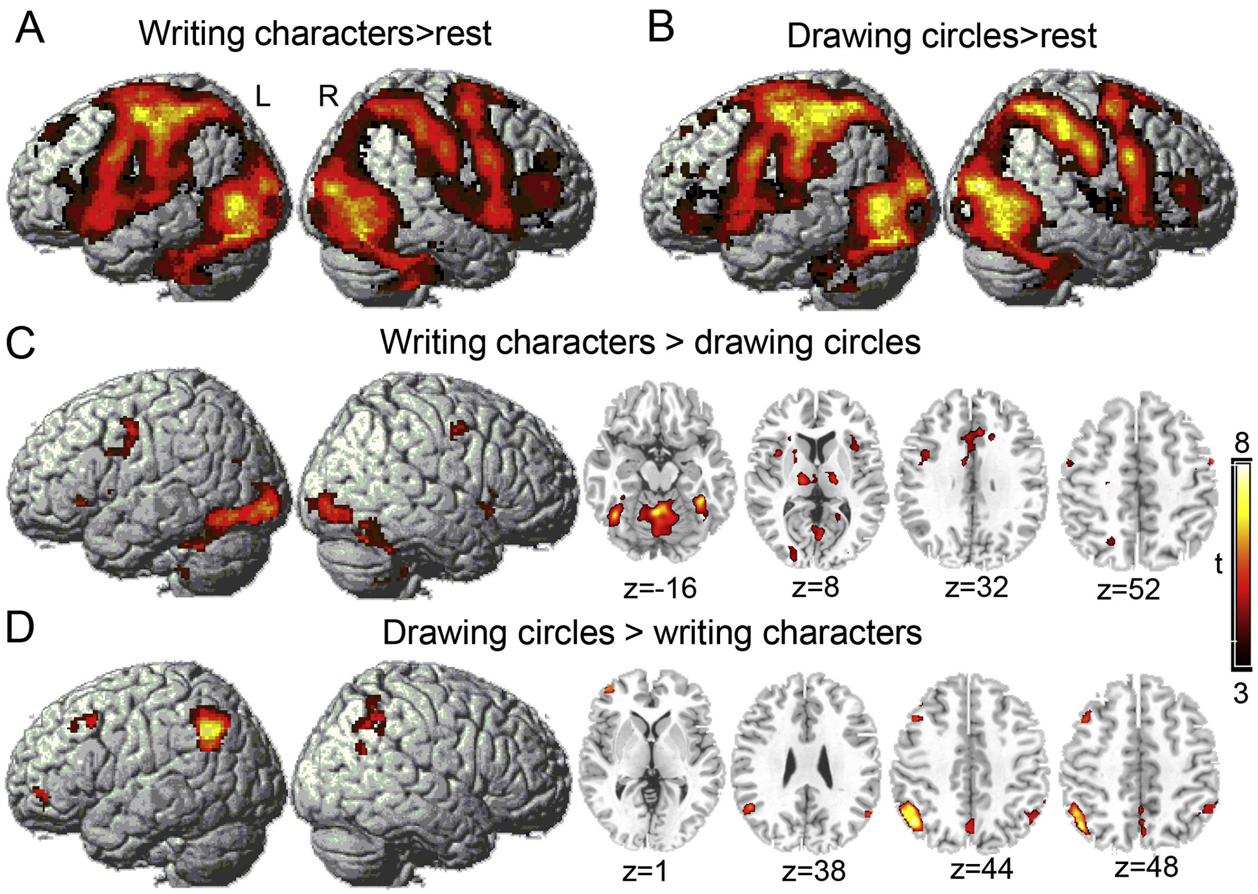
Activation time course of brain areas involved in word production



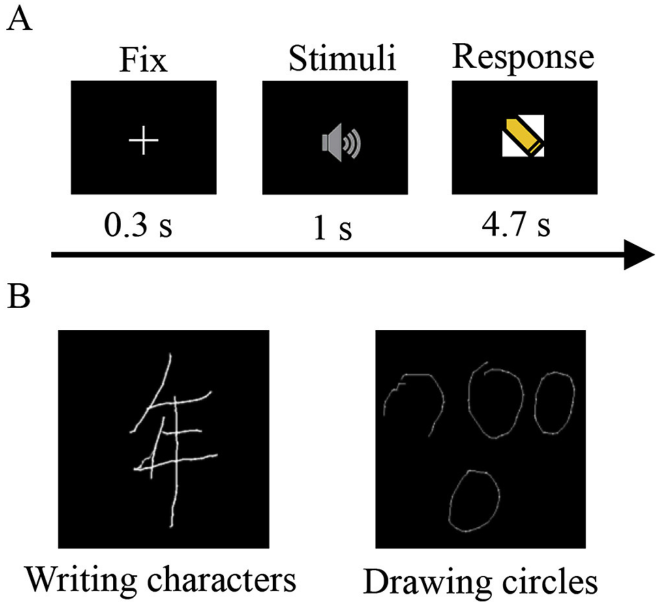
Task:
Naming a picture under fMRI, or EEG

Figure from Indefrey, 2011

Chinese Processing and Brain



Writing Chinese characters
⇔ drawing circles:
Different brain areas are activated



Yang et al., 2019, Journal of Neurolinguistics, https://doi.org/10.1016/j.jneuroling.2019.03.002.

Overview

- What is CogSci?
- Why CogSci matters to NLP?
- CogSci 101
- **Topics in CogSci in relation to Language -- Case studies:**
  - **1. Linguistic Alignment**
  - 2. Information in Language
  - 3. Turn-taking and Breathing

1. Linguistic Alignment

Phrase structure grammar (Chomsky, 1956)

Customer: *What time do you close today?*
Shopkeeper: *Six o'clock.*

Customer: ***At*** *what time do you close today?*
Shopkeeper: ***At*** *six o'clock.*

Same word "**at**" used, and same grammatical role *preposition*

VP **NP** **PP**

Teacher: *Who gave the book to Thomas?*
Student: *I gave it to him.*

VP **NP** **NP**

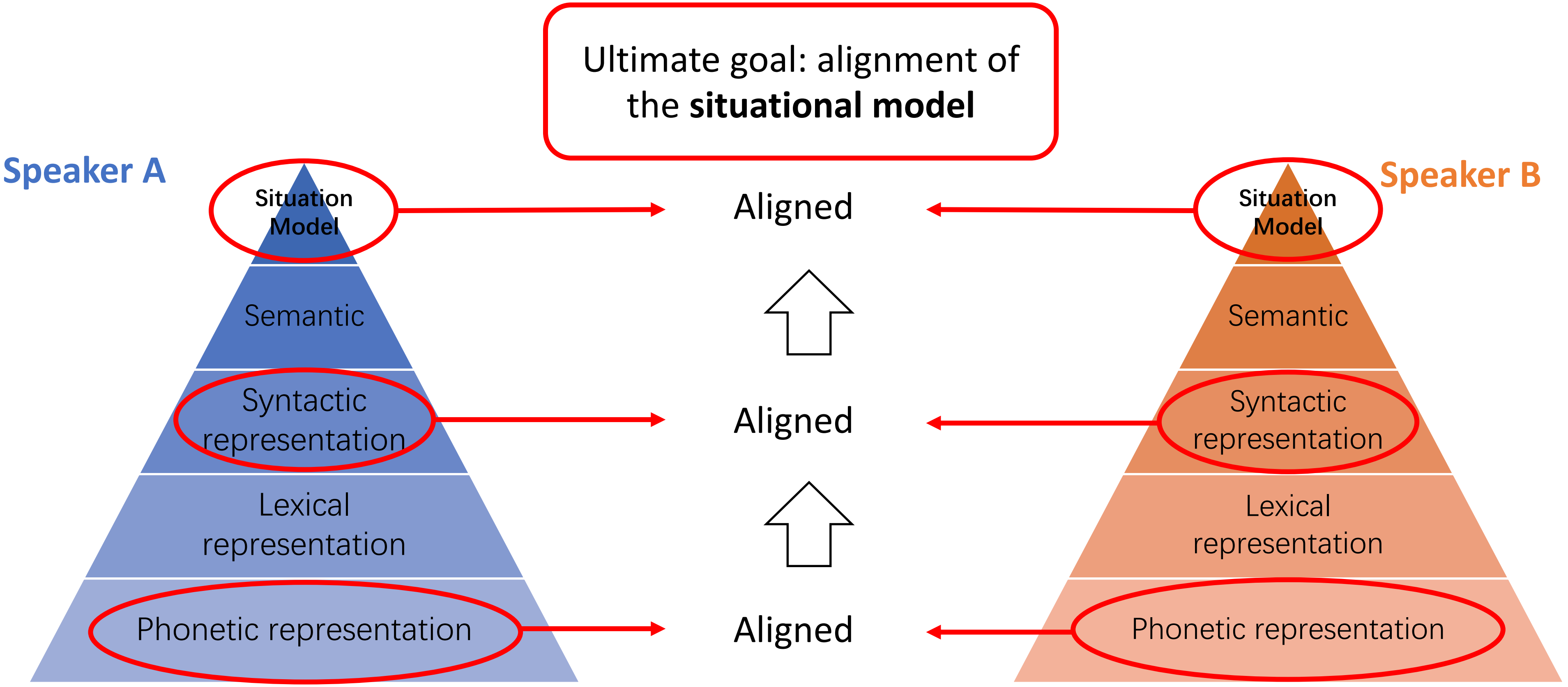
Teacher: *Who gave Thomas the book?*
Student: *I gave him the book.*

Same syntax used

Alignment: reuse of linguistic elements between speakers

Interactive Alignment Model (IAM)

Pickering & Garrod (2004)



Example of IAM: maze game (Garrod et al., 1987)

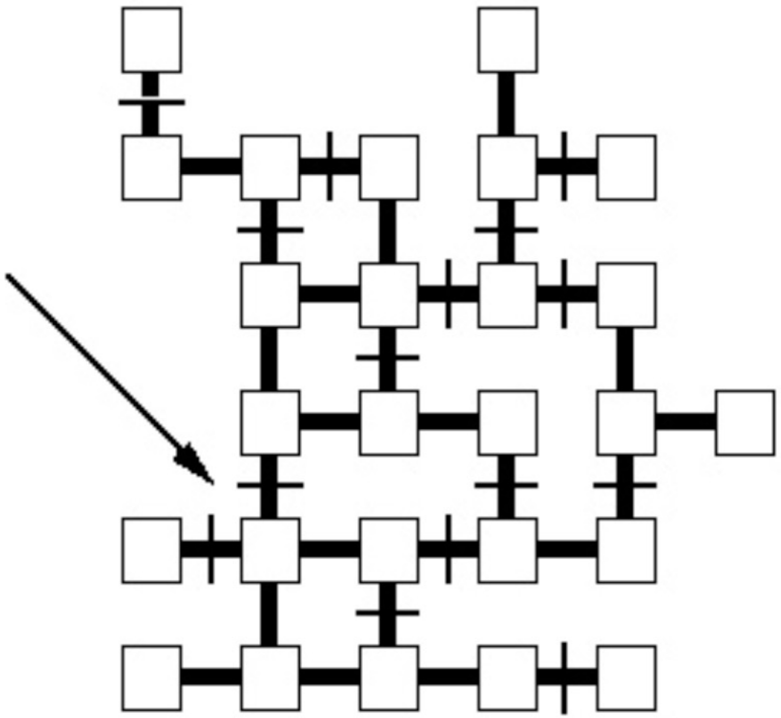
- Speaker A describes his position to B
- Multiple levels of alignment happen in dialogue

1——**B:** . . . Tell me where you are?
2——**A:** Ehm : Oh God (*laughs*)
3——**B:** (*laughs*)
4——**A:** Right : **two along from the bottom one up:***
5——**B:** Two along from the bottom, which side?
6——**A:** The left : going from left to right in the second box.
7——**B:** You're in the second box.
8——**A:** One up (*1 sec.*) I take it we've got identical mazes?
9——**B:** Yeah well : right, starting from the left, **you're one along:**
10——**A:** Uh-huh:
11——**B: and one up?**
12——**A:** Yeah, and I'm trying to get to . . .
[*28 utterances later*]
41——**B:** You are starting from the left, **you're one along, one up?** (*2 sec.*)
42——**A: Two along** : I'm not in the first box, I'm in the second box:
43——**B:** You're **two along:**
44——**A: Two up** (*1 sec.*) counting the : if you take : the first box as being one up:
45——**B:** (*2 sec.*) Uh-huh:
46——**A:** Well : I'm **two along, two up** (*1.5 sec.*)

Simple phonetic cues and words are aligned

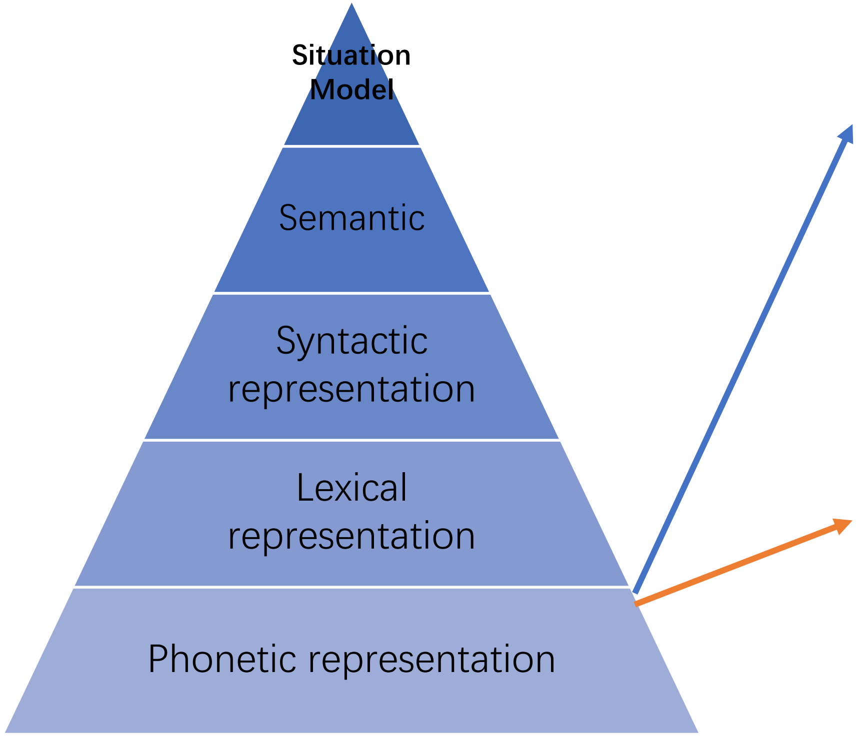
More complex phrases are aligned

Alignment of a larger set of structures



From Pickering & Garrod (2004)

Applications of alignment



Social computing

Lexical and **syntactic** alignment are used to classify user actions, and correlate with network structure in online community (Wang et al., 2015; Noble et al., 2015)

Dialogue system

Alignment of **acoustic cue**s, **prosodic cues**, and **words** are used to predict user behavior and improve system accuracy (Campbell et al., 2010; Fandrianto et al., 2012; Levitan, 2013, 2016).

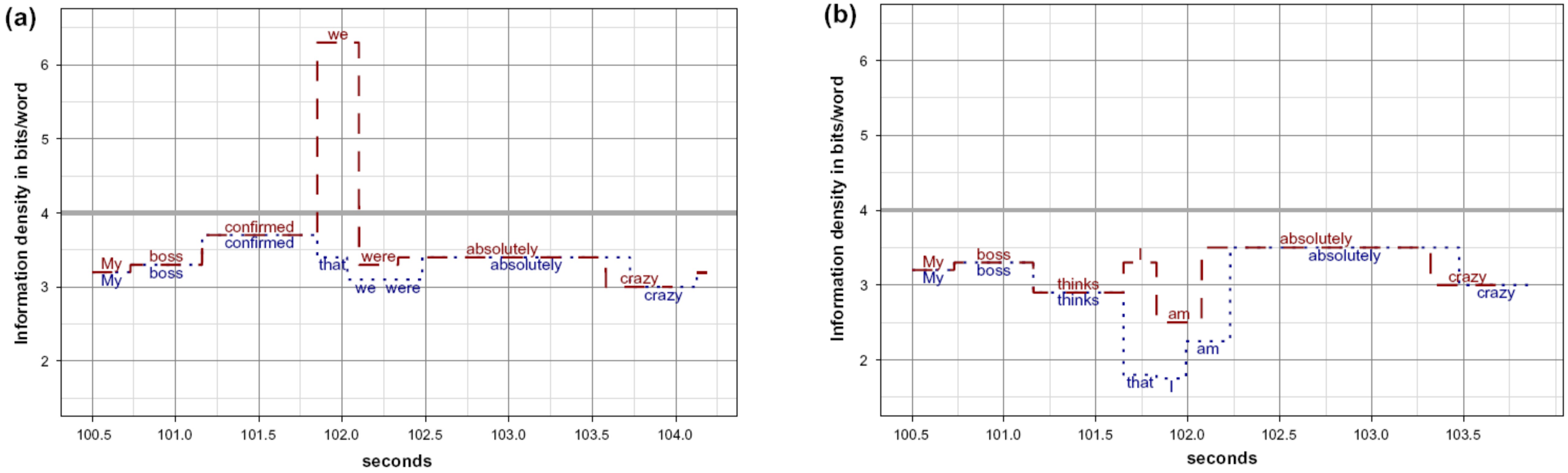
Overview

- What is CogSci?
- Why CogSci matters to NLP?
- CogSci 101
- **Topics in CogSci in relation to Language -- Case studies:**
  - 1. Linguistic Alignment
  - **2. Information in Language**
  - 3. Turn-taking and Breathing

Information Density

$$I(\text{word}) = -\log p(\text{word})$$

- Uniform Information Density predicts that language production is affected by a (human) preference to distribute information ***uniformly*** across the linguistic signal. (Jaeger et al., 2010)



Entropy ≠ …

$$H \triangleq -\sum p \log p \qquad NLL = -\sum \log p$$

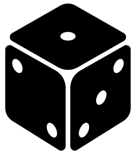
Shannon's entropy $\neq$

Negative log-likelihood (NLL)

Surprisal

Information density

Cross-entropy



almost equivalent concepts

On the wall I see a $\begin{cases} \textit{picture} \\ \textit{pitcher} \end{cases}$

$$p(picture|\dots) = 10^{-4}$$

$$p(pitcher|\dots) = 10^{-6}$$

NLL(picture) < NLL(pitcher)

More information in "pitcher"??
(not necessarily)

$$H(X|On\ the\ wall\ I\ see\ a) =$$

$$-p(picture) \cdot \log p(picture|\dots)$$

$$-p(pitcher) \cdot \log p(pitcher|\dots)$$

…

$$-p(dog) \cdot \log p(dog|\dots)$$

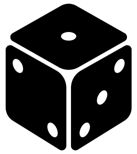
…

Information, probability, entropy

Classical Shannon's Information Theory (1948)

Entropy of a random variable X:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log p(x_i)$$

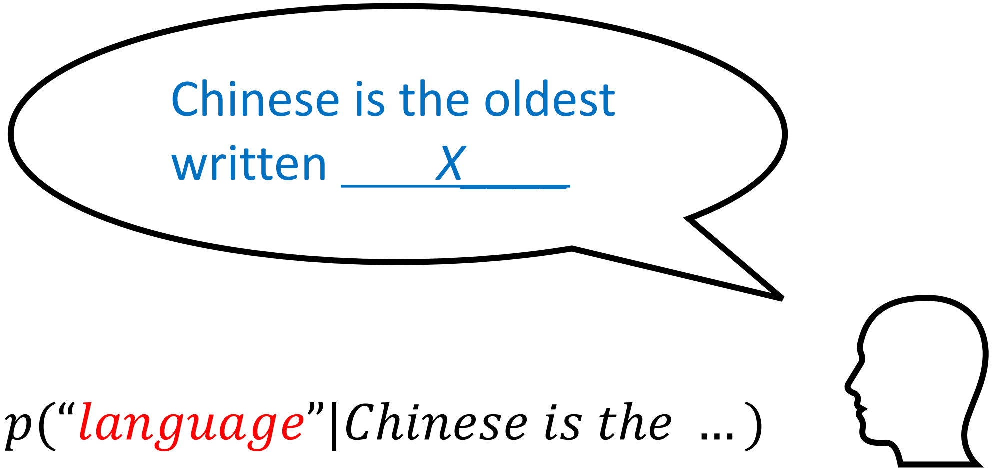


A normal dice with 6 sides:

$$p(1) = p(2) = \cdots = p(6) = \frac{1}{6}$$

$$H = -6 * (\frac{1}{6})\log_2(\frac{1}{6}) \approx 2.585$$

Treating ***word*** as a random variable



$p(\text{“}language\text{”}|Chinese\ is\ the\ \ldots)$

$\gg p(\text{“}culture\text{”}|Chinese\ is\ the\ \ldots)$

$\gg p(\text{“}people\text{”}|Chinese\ is\ the\ \ldots)$

Context matters! How to quantify?

Decompose context

- *Context* is difficult to model, further decompose (Shannon, 1948)

$$\begin{aligned} H(X|Context) &= H(X|C, L) \\ &= H(X|L) - I(X; C|L) \end{aligned}$$

- C, **global context**
    - → All preceding words *before* the current sentence
- L, **local context**
    - → All preceding words *within* the current sentence
- $I(X; C|L)$, **mutual information**
    - → How much do we know about the word by reading all its global context.

Which terms can be computed?

$$H(X|Context) = \underbrace{H(X|C, L)}_{\text{No}} = \underbrace{H(X|L)}_{\text{Yes}} - \underbrace{I(X;C|L)}_{\text{No}}$$

A sentence of N words, $s = \{w_1, w_2, \cdots, w_N\}$

$$H(X|L) \approx \boxed{\mathrm{NLL}(s)} = -\frac{1}{N}\sum_{k=1}^{N} \log p(w_k|w_{<k})$$

Average **negative log-probability** of all words in sentence

Information (negative log likelihood)

$$p(w_k|w_1, \ldots, w_{k-1})$$

Can be estimated using language models (n-gram, RNN, GPT, …)

Question: Which term is uniformly distributed?

$$H(X|Context) = H(X|C,L) = \boxed{H(X|L)} + I(X;C|L)$$

Approximated by $\boxed{\mathrm{NLL}(s)}$

- Does it conform to the principle of **Uniform Information Density (UID)**, or from the perspective of another line of work: **Entropy Rate Constancy (ERC)?**

Assuming UID/ERC is true

Constant $\leftarrow$ $\boxed{H(w|Context)} = H(w|\mathbf{G},\mathbf{L}) = \boxed{H(w|\mathbf{L}) - I(w;\mathbf{G}|\mathbf{L})}$ Constant

Expected observation:

- $I(w_i; G\ |L)$ increases with i
- $I(w_i; G\ |L)$ decrease when topic shifts

NLL ($-\log p$) measures the difficulty (cognitive load) of understanding word w

Good speakers put hard words later (☺)

Sentence 1: $\{w_1, w_2, \dots, w_6\}$

Sentence 2: $\{w_1, w_2, \dots, w_8\}$

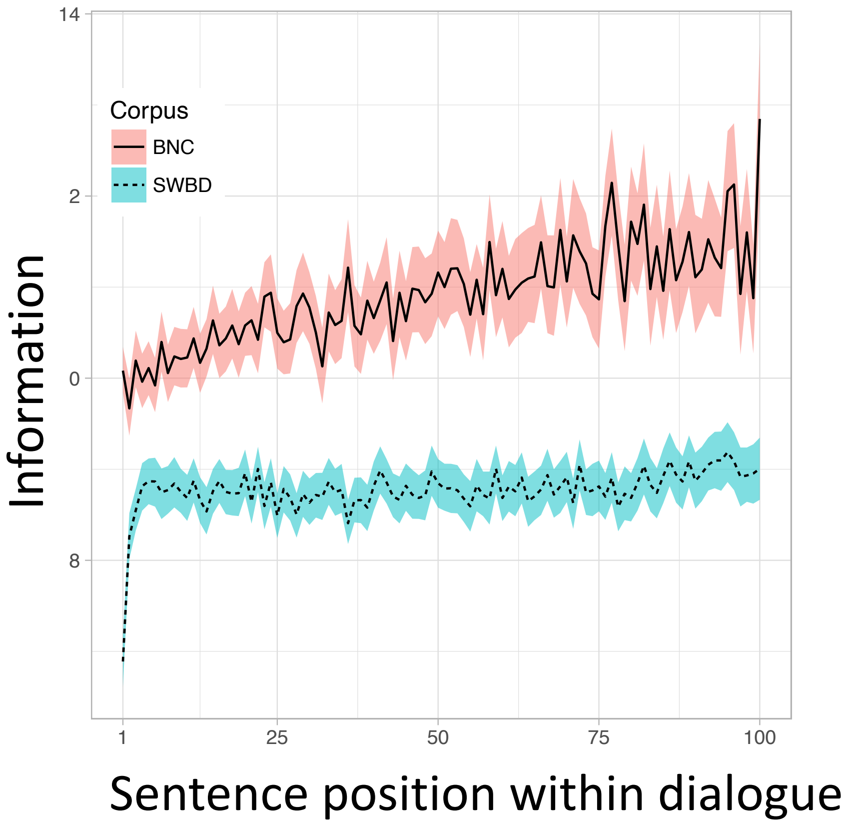
Sentence 2: $\{w_1, w_2, \dots, w_{10}\}$

…

Sentence *N*: $\{w_1, w_2, \dots w_{i-1}, \boxed{w_i}, \dots, \}$

Expected observation 2 confirmed (in dialogue)

- $H(X_i|L_i)$, $\mathrm{NLL}(s_i)$, increases with i



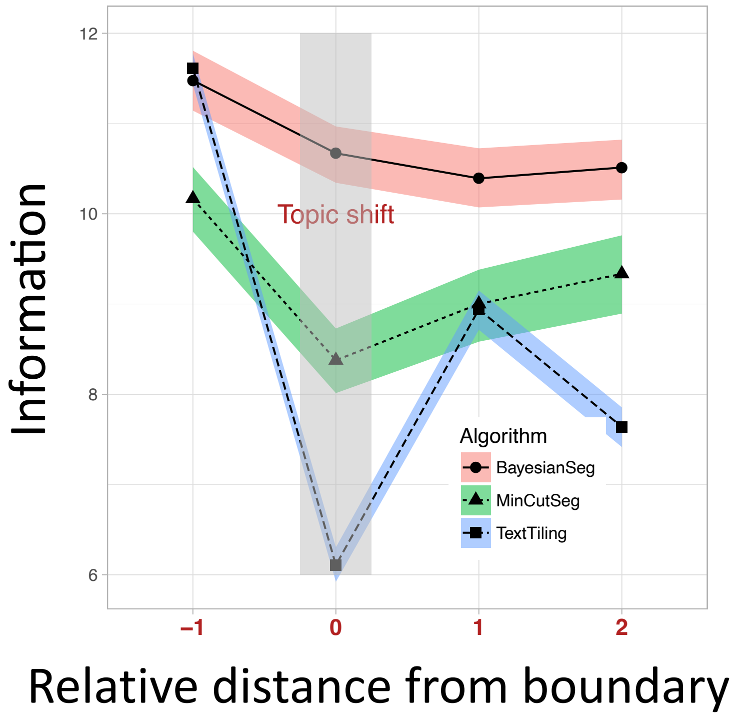
$$H(w|Context) = H(w|\boldsymbol{G}, \boldsymbol{L}) = H(w|\boldsymbol{L}) - I(w; \boldsymbol{G}|\boldsymbol{L})$$

Figure from Xu et al., 2016

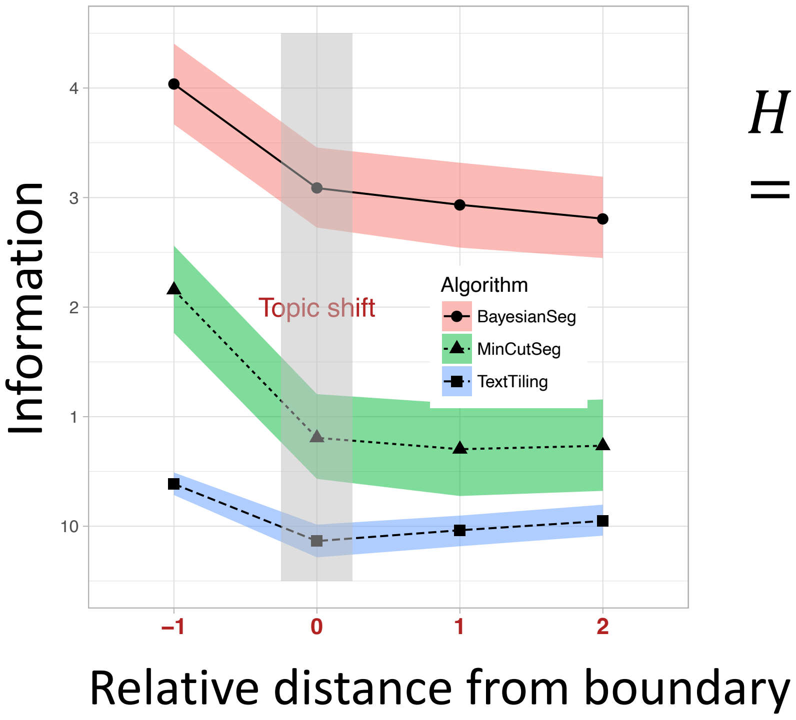
Expected observation 2 confirmed (in dialogue)

- Information suddenly **decrease** at topic boundaries.

Switchboard

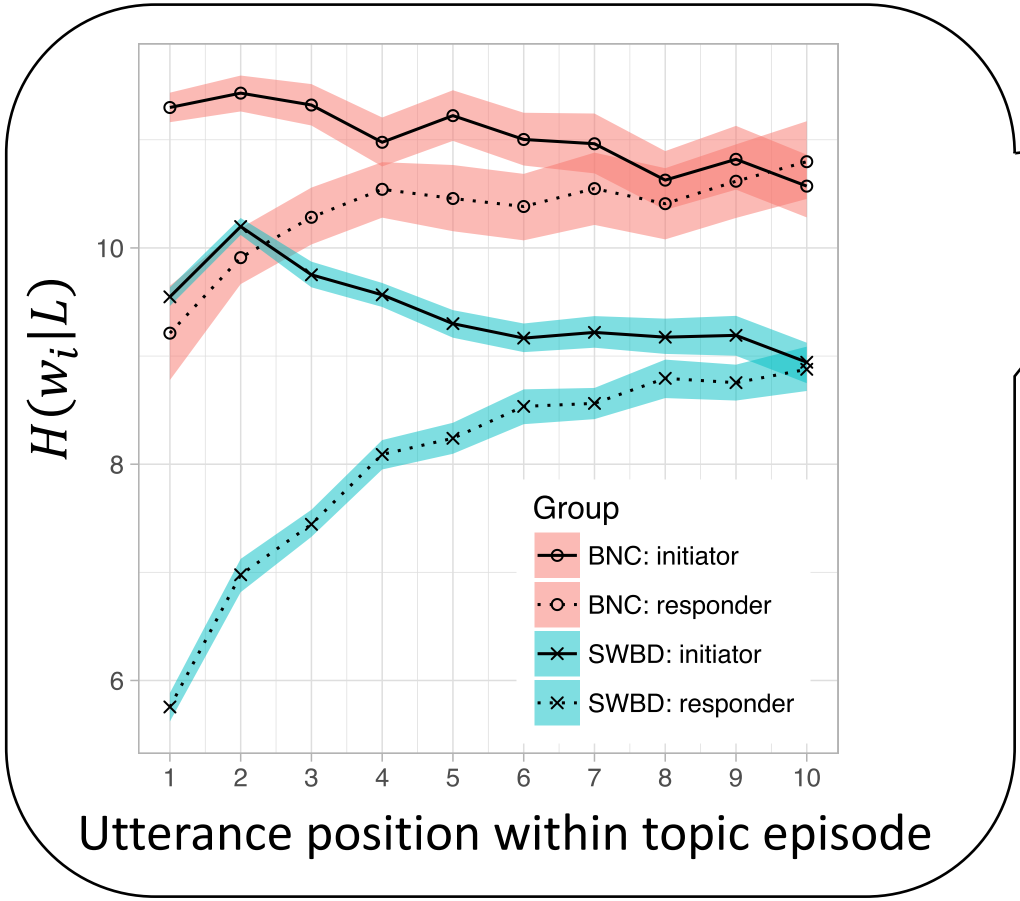


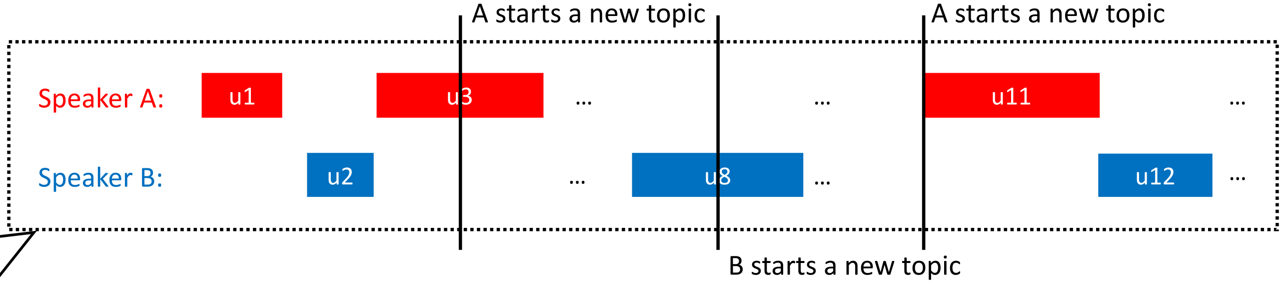
BNC



$$H(w|Context) = H(w|\color{red}{\boldsymbol{G}}, \color{blue}{\boldsymbol{L}})$$
$$= H(w|\color{blue}{\boldsymbol{L}}) - I(w; \color{red}{\boldsymbol{G}}|\color{blue}{\boldsymbol{L}})$$

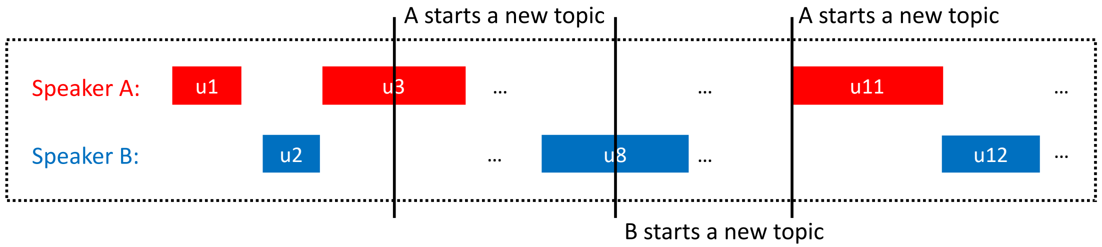
NLL term $H(w|L)$ converges



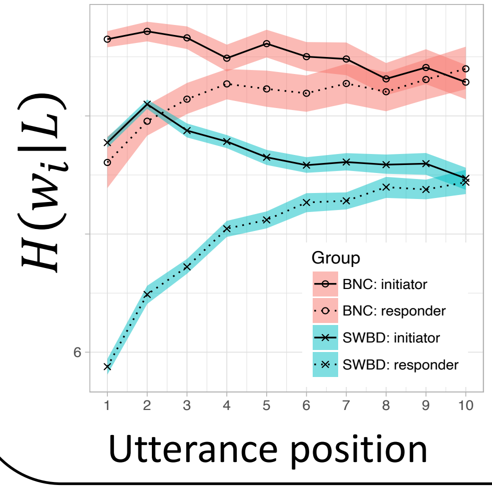


The local entropy of topic initiators and responders increases form a *convergence* pattern within topic episode

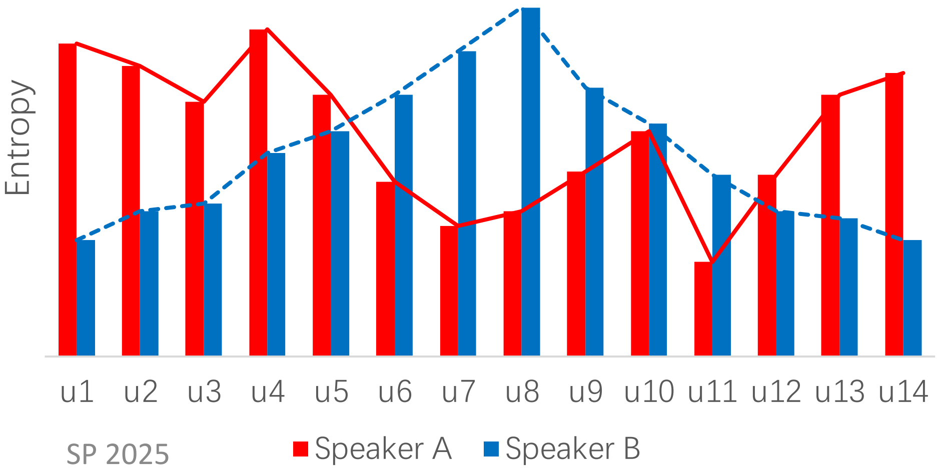
Periodicity of information in dialogue



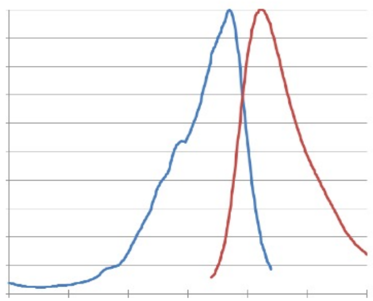
Topic shift occurs ***periodically*** → Aggregate → Convergence



Time domain



Frequency domain

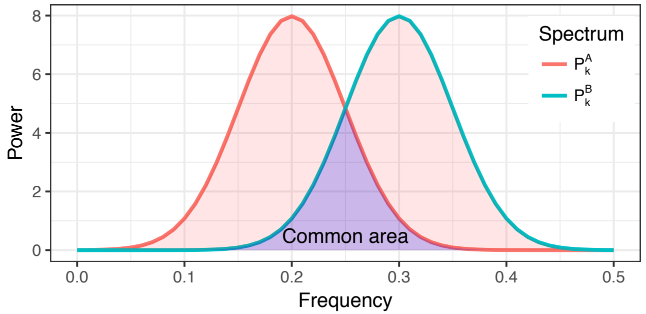


Features from frequency domain may reflect high-level dialogue properties

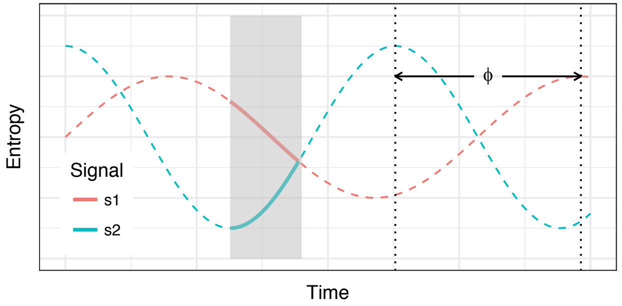
Frequency domain features

1. Power Spectrum Overlap, **PSO**
   Similarity between two spectra
2. Relative Phase, **RP**
   Delay between two signals

PSO



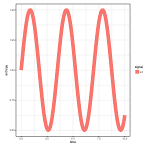
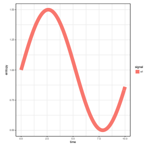
RP



PSO is a **negative** predictor of **task success** ☹
Larger overlap in spectra, poorer conversation

RP is **positive** predictor of **task success** ☺
Larger relative phase, better conversation

Interpret PSO

- Frequency ⇔ rate of information change: high f →  low f 

High PSO

Speaker A: high f
Speaker B: high f

Both change information **rapidly**

Competition for channel

Speaker A: low f
Speaker B: low f

Both change information **slowly**

Inefficient usage of channel

Bad strategy!

Low PSO

Speaker A: high f
Speaker B: low f

Speaker A: low f
Speaker B: high f

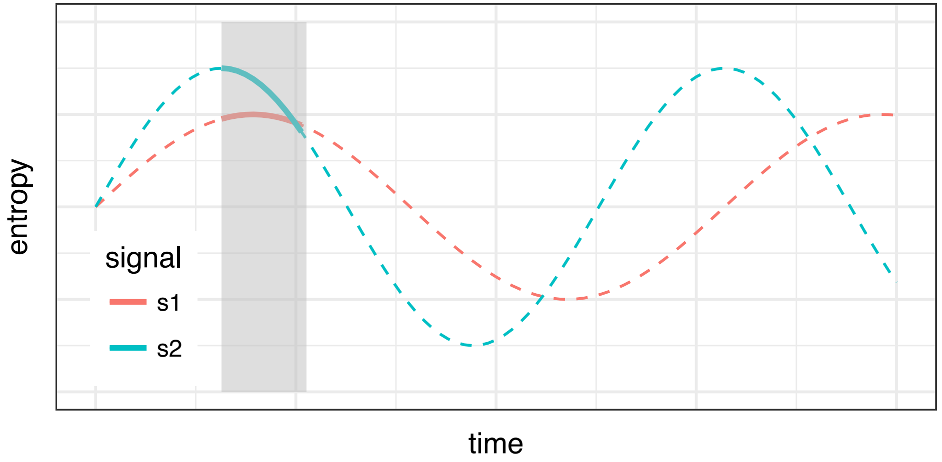
Taking turns for rapid information change

Balanced usage of channel

Good strategy!

Interpret RP

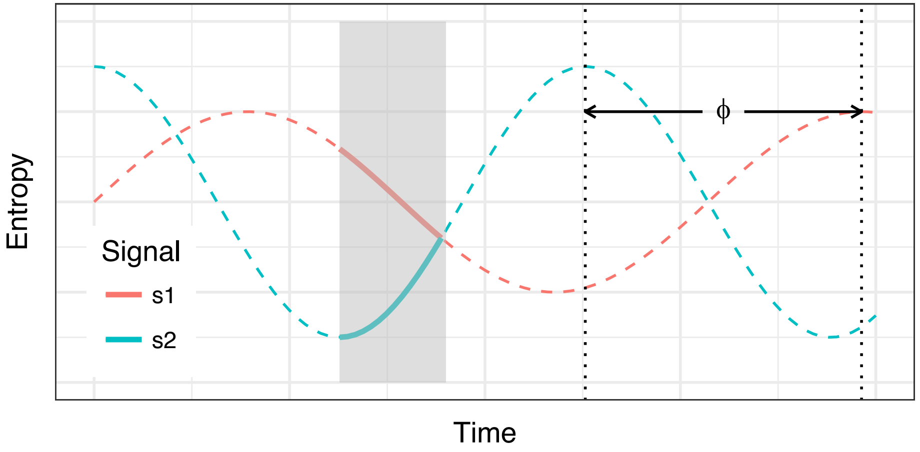
Low RP



Shorter delay between information peaks, **blocking** the channel

Bad strategy!

High RP



Longer delay between information peaks: **avoid simultaneous** occupying of channel

Good strategy!

Overview

- What is CogSci?
- Why CogSci matters to NLP?
- CogSci 101
- **Topics in CogSci in relation to Language -- Case studies:**
  - 1. Linguistic Alignment
  - 2. Information in Language
  - **3. Turn-taking and Breathing**

Turn-Taking 话轮转换

- Language is interactive, involving rapid turn-taking
- Turns are short and responses are remarkably rapid
- Part of a **universal infrastructure** for language

Responses in conversation are fast

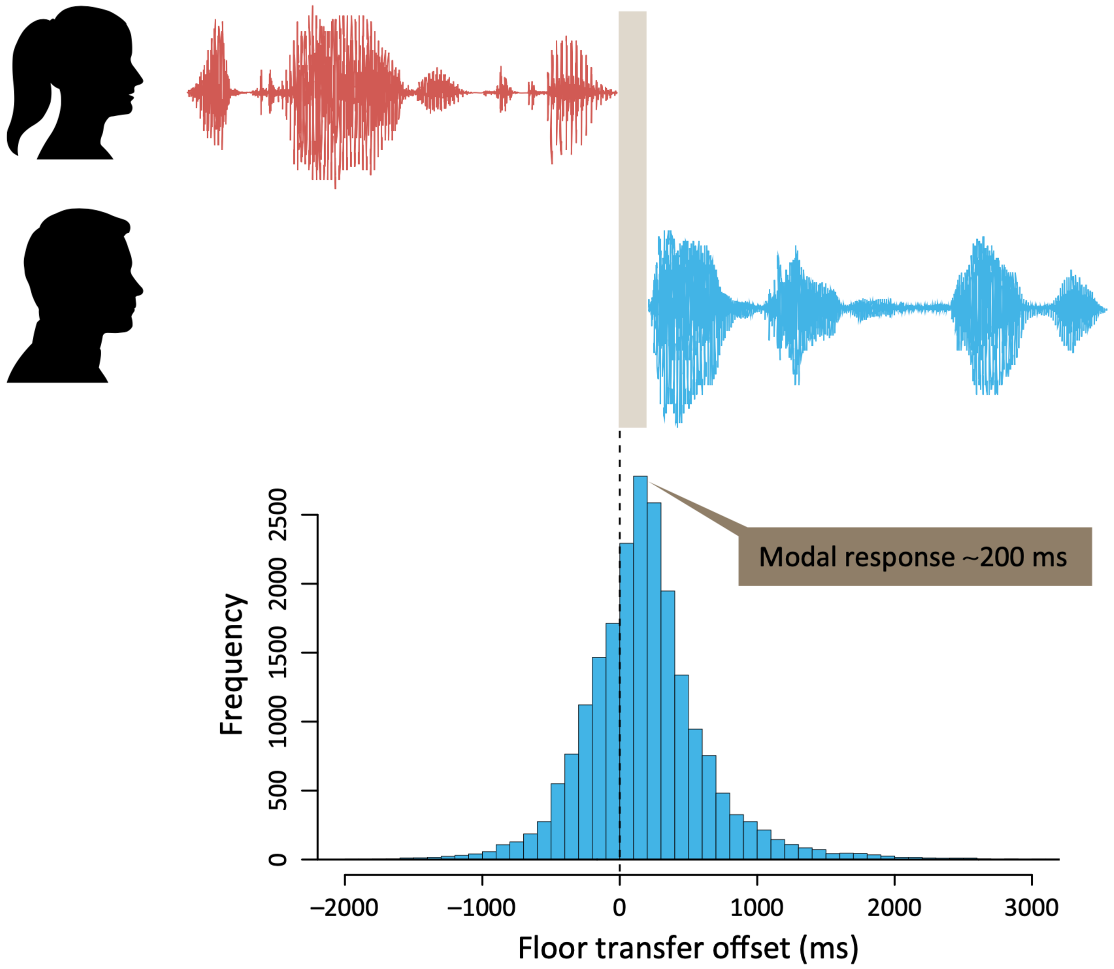
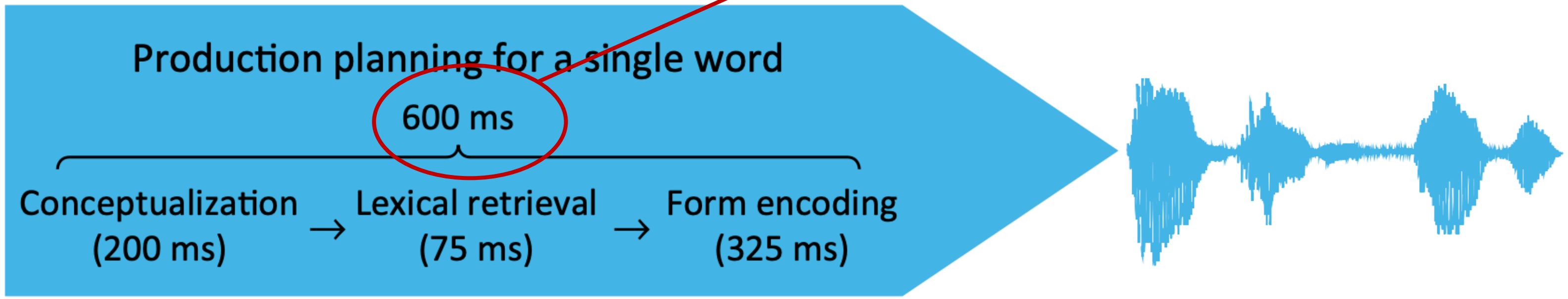


Figure from Levinson, 2016

Turn-Taking

Latencies in production are **three fold**:

Longer than the 200ms modal gap!



From **dense vectors** to one-hot

Lookup the vocabulary table

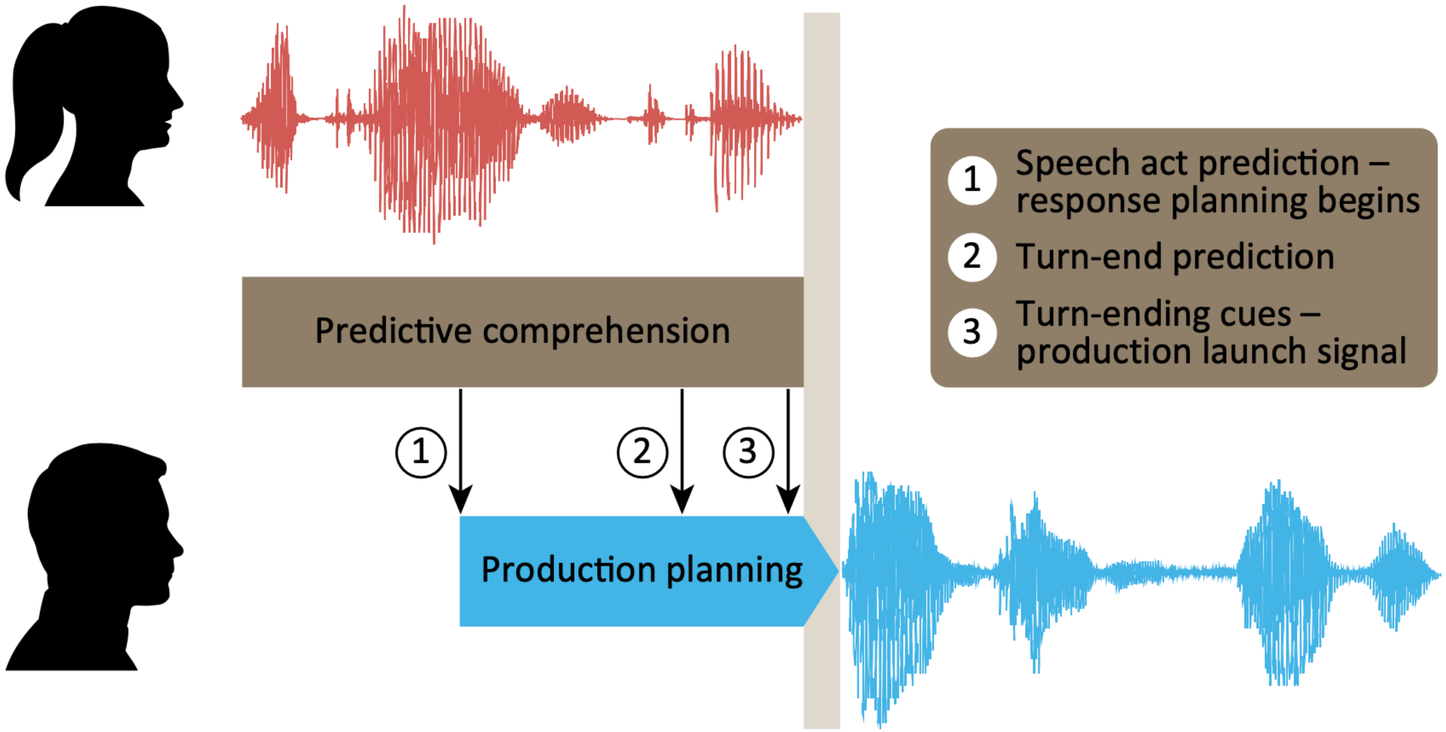
text-to-audio

Analogy to generation with LM

Figure from Levinson, 2016

Turn-Taking

Production of response must therefore overlap with comprehension of the incoming turn



The slow production may be compensated for by predicting the **continuation** or **termination** of the incoming turn, and launching production early

Figure from Levinson, 2016

Not just humans do turn-taking

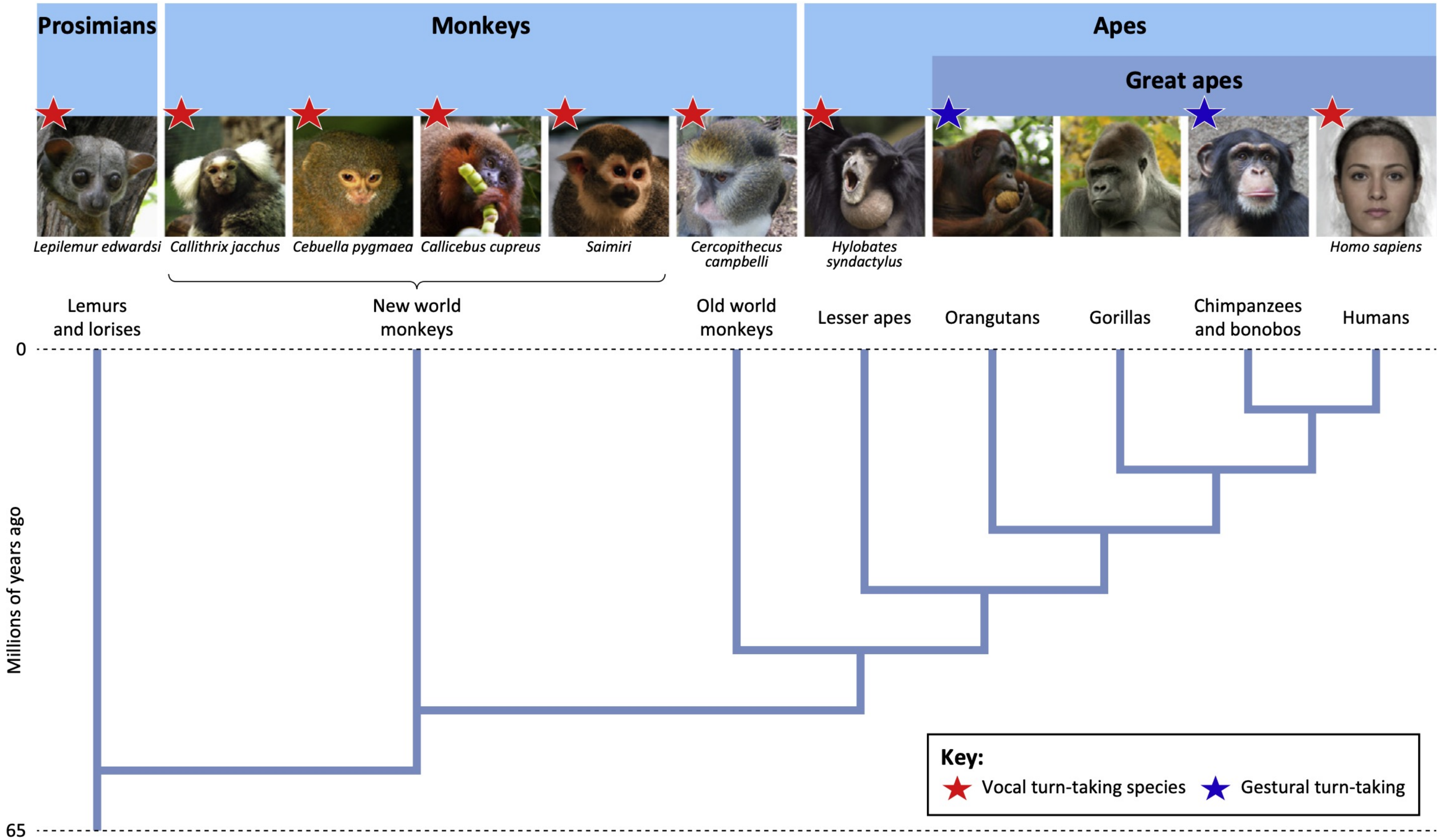


Figure from Levinson, 2016

Question

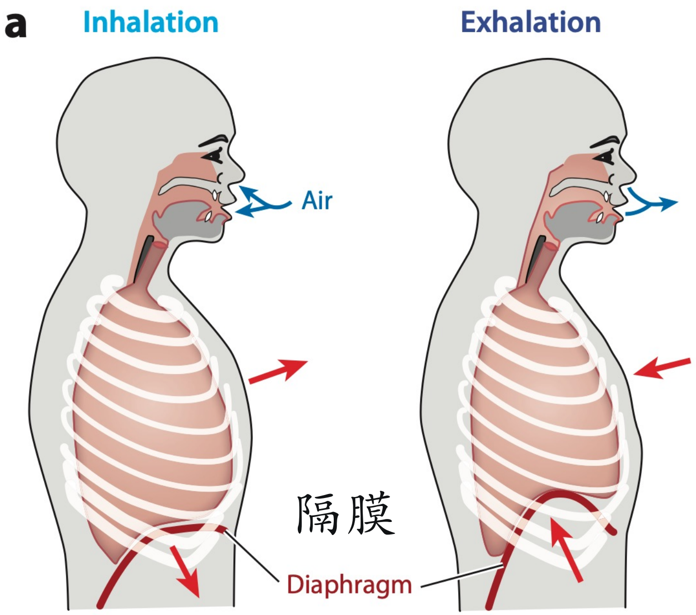
- What’s the difference in GPT-to-Human vs. Human-to-Human conversations?
- How is GPT-4o trained so that it recognizes the window for “turn-taking”?

Breathing is related to spoken language

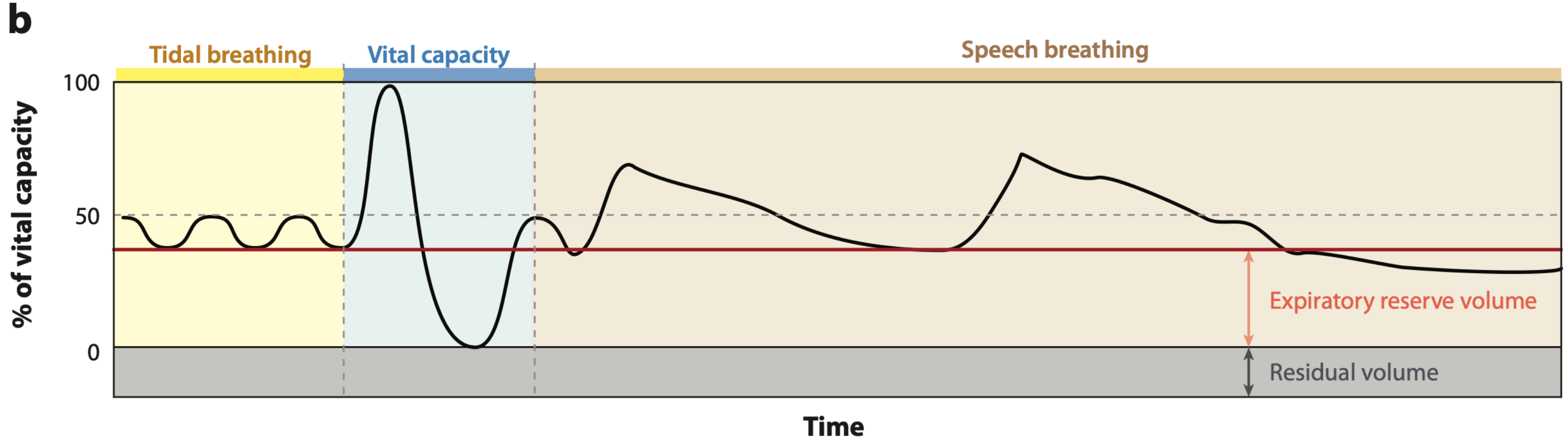
肺活量

Vital capacity: the maximum amount of air that can be breathed out after breathing in as much air as possible.

Speech breathing: involves more variable and deeper inhalations (更深的吸气) than tidal breathing



/ ˈdaɪəfræm /



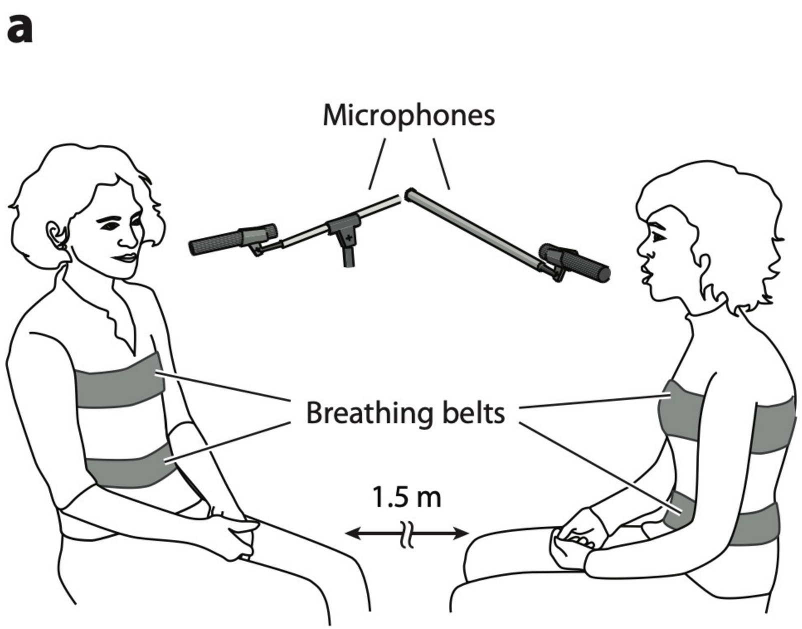
Tidal breathing: the term for inhaling and exhaling while you are relaxing

Speech breathing: short inhalation and a long exhalation

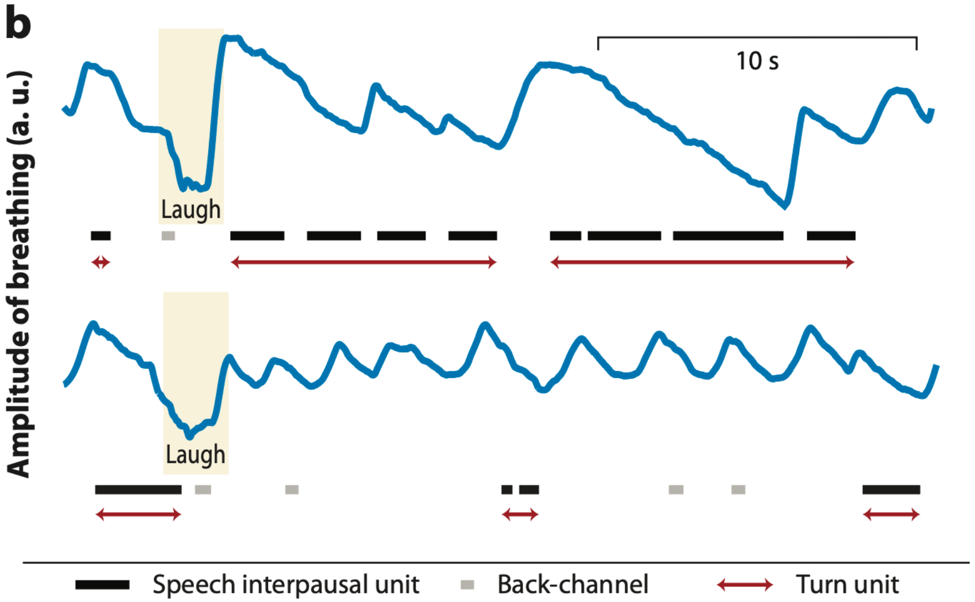
Figure from Fuchs and Rochet-Capellan, 2020

Breathing during conversation/dialogue

- Techinique: Inductance plethysmography (电感体积描记法)



Breathing signals are recorded using breathing belts



Black bars indicate stretches of speech; red arrow indicates a speaker's complete turn

Figure from Fuchs and Rochet-Capellan, 2020

Breathing adapts to speech

- Inhalation (吸气) at **syntactic boundaries** 完整句法结构的边界
  - Rather than occurring randomly in the speech stream, inhalation very often coincides with constituent boundaries
- Anticipating the length of the upcoming sentence
  - Deeper inhalation for longer sentences than for shorter ones

长句 短句

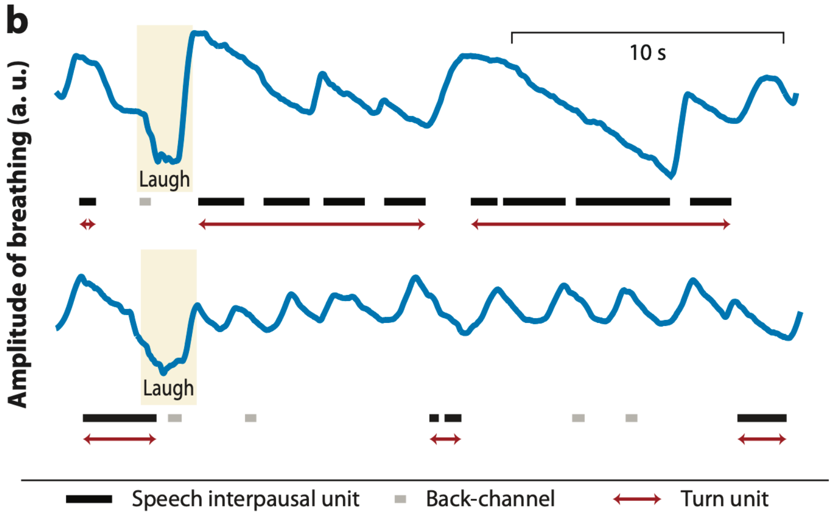


Figure from Fuchs and Rochet-Capellan, 2020

Breathing adapts to speech

- Breathing is related to prosody (语气；韵律), specifically prominence (凸显)

- Breathing when listening
  - Listeners breathe **faster** when listening to **loud** speech compared with normal-level speech
  - breathe **slower** when listening to **slow** speech compared with normal-rate speech

- Interpersonal coordination of breathing
  - Speakers tend to display in-phase coordination of breathing when reading together
  - Synchronization (同步) is stronger during simultaneous reading

Why caring about breathing?

- Question: How is breathing related to information in language?
- Open-ended question: Do LLMs “breathe”? 😳😮‍💨

References

- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23-62.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. Frontiers in psychology, 2, 255.
- Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. Trends in cognitive sciences, 20(1), 6-14.
- Xu, Y., & Reitter, D. (2018). Information density converges in dialogue: Towards an information-theoretic model. Cognition, 170, 147-163.
- Fuchs, S., & Rochet-Capellan, A. (2021). The respiratory foundations of spoken language. Annual Review of Linguistics, 7, 13-30.