Connecting Variations in Language Structures and Language Processing

A Cross-Linguistic Study of Homophones

Youtao Lu

JSPS Fellowship for Research in Japan
International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo

JSPS Science Dialogue, 岩手県立一関第一高等学校

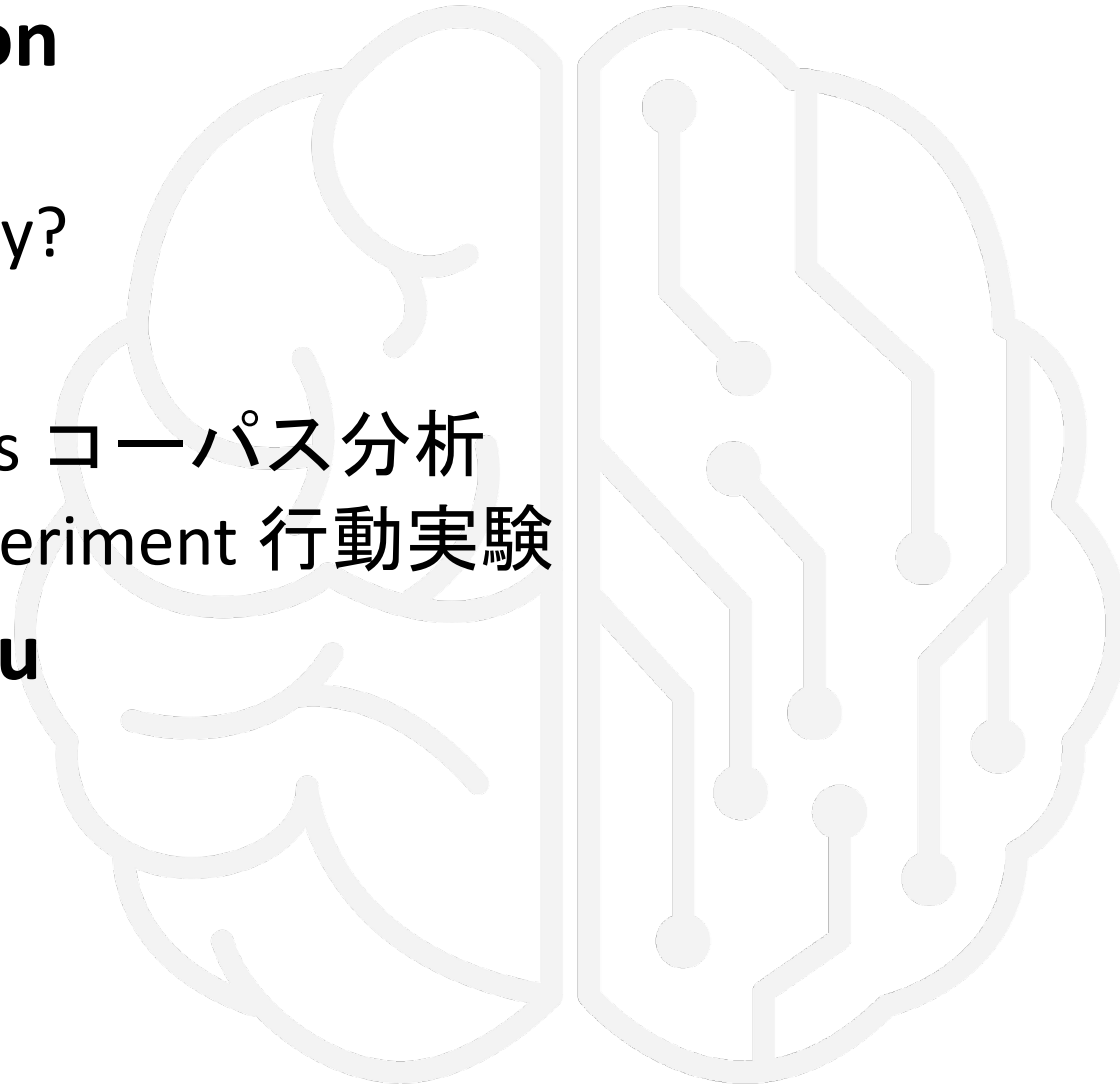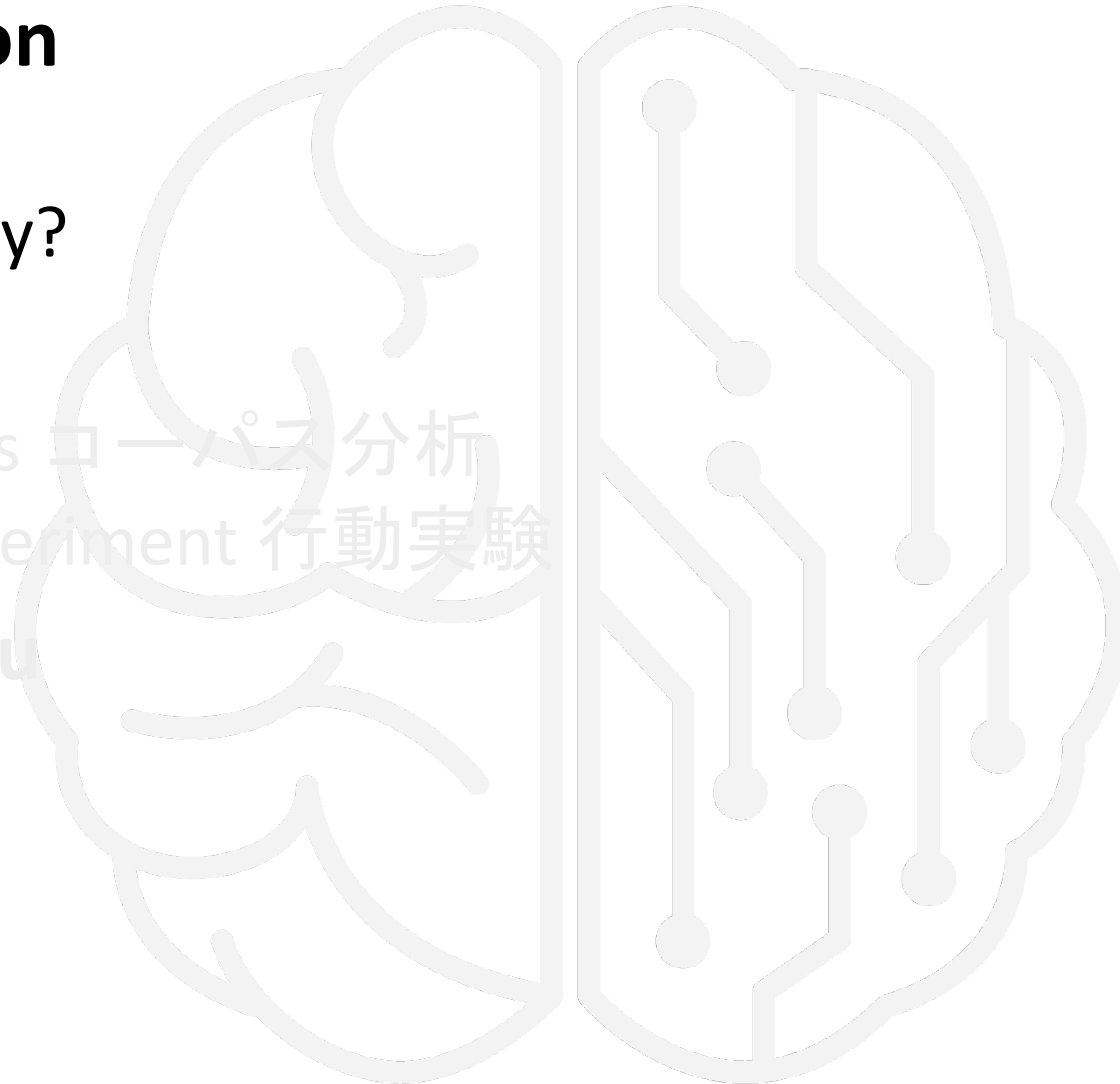*2025. 1. 29*

Contents

- **Self-Introduction**
  - Who am I?
  - What do I study?

- **Research Talk**
  - Corpus analysis コーパス分析
  - Behavioral experiment 行動実験

- **Message for you**

Contents

- **Self-Introduction**
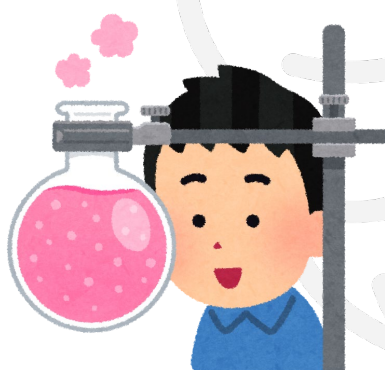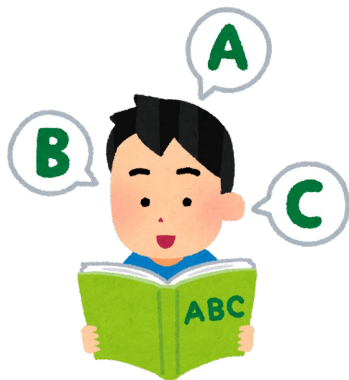 - Who am I?
  - What do I study?

- Research Talk
  - Corpus analysis コーパス分析
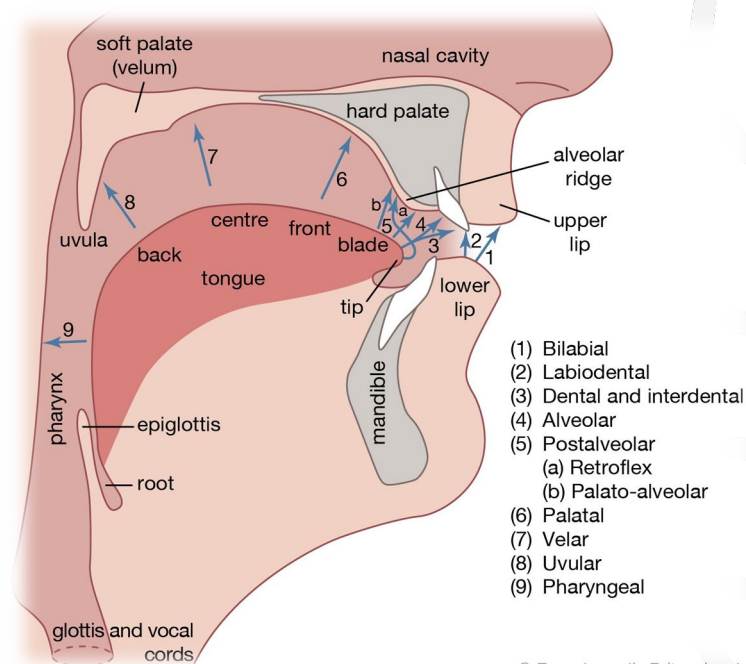 - Behavioral experiment 行動実験

- Message for you

Self Introduction

- Lu Youtao　ルウ ヨウタオ　陸 友陶

- Grew up in Nanjing, P. R. CHINA
  - I studied at a high school that emphasizes learning English
  - I was very interested in, and picked science in high school 理系
  - I went to study at Tsinghua University, Beijing
    - Foreign Languages and Literatures 外国語・文学
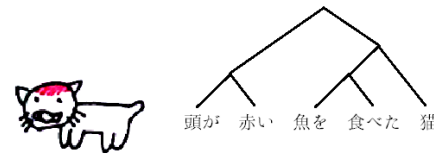   - Economics 経済学

Self Introduction

- Got to know and started to study **linguistics 言語学**
  - The scientific study of human languages

Self Introduction

- Got to know and started to study **linguistics** 言語学
  - The scientific study of human languages
  - In physics, there is the **Grand Unified Theory** 大統一理論
    - a single theory that is supposed to explain all physical phenomenon
  - In linguistics, there is the theory of **Universal Grammar** 普遍文法
    - Proposed by Noam Chomsky
    - Human children can learn any human language equally well
    - There must be a set of rules shared by all human languages
    - Children are born with the knowledge of these rules
    - Still a hypothesis 仮説

Self Introduction

- Learning and doing research as a graduate student at Brown University
  - Located in Rhode Island, the smallest state of the US

- Even if there is **Universal Grammar**, languages are still different
  - What are these differences?
    - Cross-linguistic comparison 言語間比較
  - How do those differences influence the way we understand and use language?
    - Psycholinguistics 心理言語学

見て！にゃんにゃんだ！

Self Introduction

- When learning Japanese, I found it difficult to memorize words that have the same pronunciation: 同音異義語
  - じしん：自信？地震？
- My research examined how these words are understood by native speakers of different languages
- The research made me realize that Japanese is a unique language
- After graduation, I decided to pursue cross-linguistic research in Japan
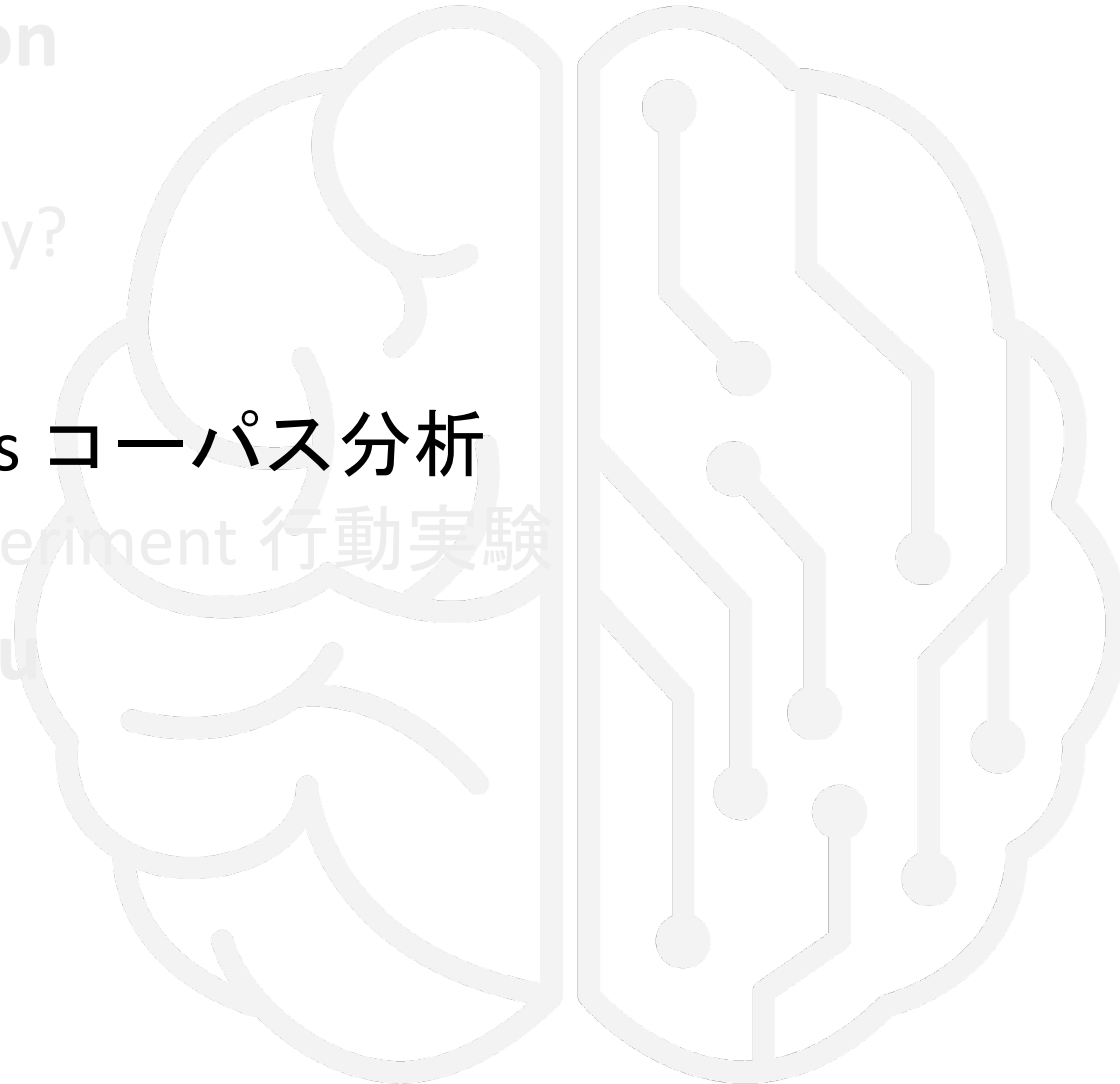
Contents

- Self-Introduction
  - Who am I?
  - What do I study?

- **Research Talk**
  - Corpus analysis コーパス分析
  - Behavioral experiment 行動実験

- Message for you

Words & Homophones

- Let's start from the question: what are words?
  - Mappings between **phonological sequences** and **meanings**
    音素配列と**意味**の対応付け

ねこ
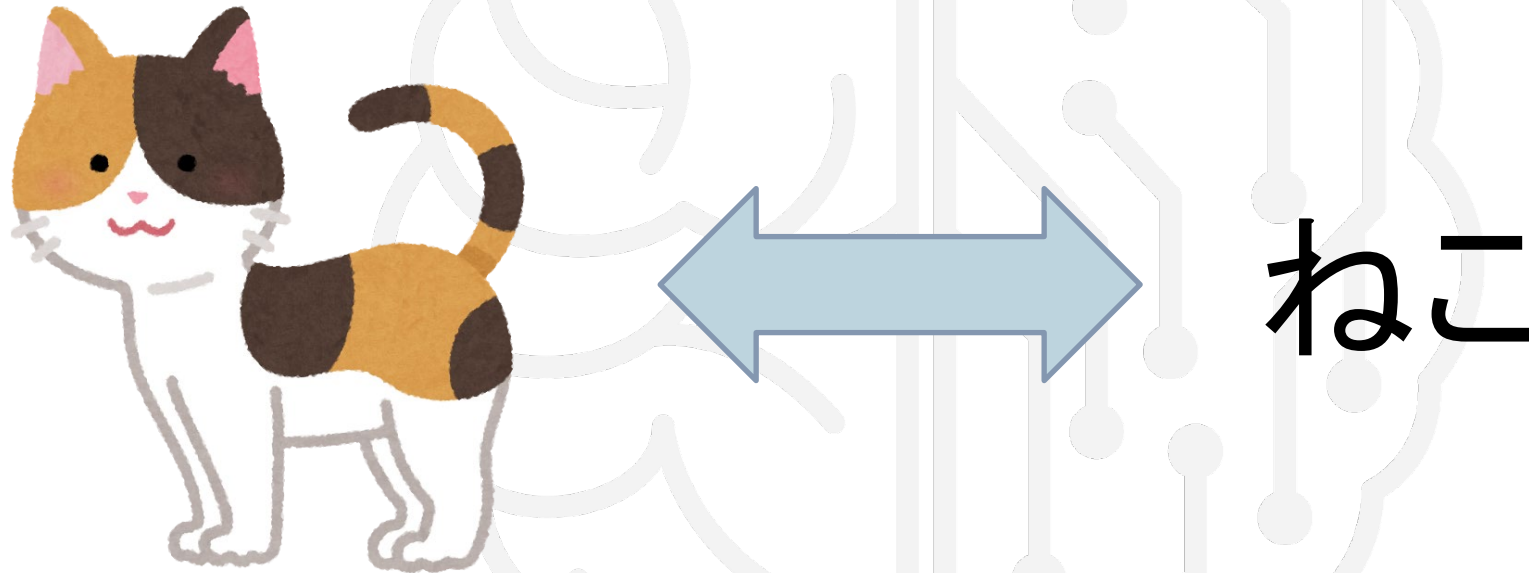
Words & Homophones

- Let's start from the question: what are words?
  - Mappings between **phonological sequences** and **meanings**
    音素配列と**意味**の対応付け
  - The mapping is not always 1-to-1

- **Homophones 同音異義語**: words with different meanings that share the phonological sequence
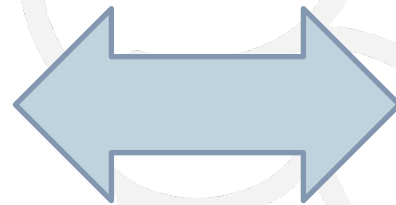
じしん

Words & Homophones

- Let's start from the question: what are words?
  - Mappings between **phonological sequences** and **meanings**
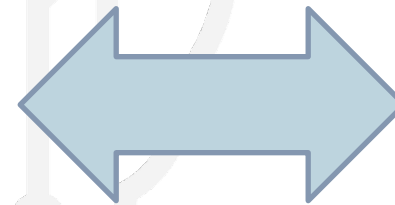   **音素配列**と**意味**の対応付け
  - The mapping is not always 1-to-1

- **Homophones 同音異義語**: words with different meanings that share the phonological sequence
  - They exist in most if not all languages

night
/naɪt/
knight
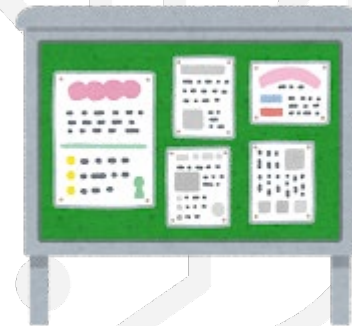
海报
hai3 bao4
海豹

Words & Homophones

- However, are homophones equally common in different languages?
  - Probably not?
  - Affected by how words are formed

- Remember that words are mappings between **phonological sequences** and **meanings**

- How many **meanings** do we need to construct words for?
  - We live in the same world
  - Should be similar across languages

- How many **phonological sequences** do we have?
  - Differ a lot by languages!

Phonemes & Combination

- Why do languages have different number of phonological sequences?

- They have different **phonemes 音素**
  - English have "r" and "l", but there is only one "r/l" phoneme in Japanese
  - Japanese distinguishes between long and short vowels, but English doesn't
  - Extreme cases: 13 phonemes in Hawaiian, 69 phonemes in Irish

- They have different **restrictions 制限 for phonological sequences**
  - Consonant clusters 子音連結
  - S C R I P T S
  - **Stra**wberry -> ストロベリー
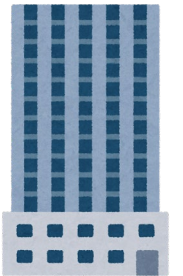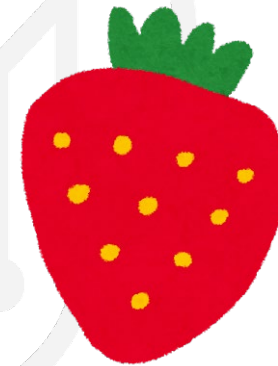
Phonemes & Combination

- What will happen if a language only has a few phonological sequences, but a lot of meanings to express?
  - Reuse phonological sequences for different meanings -> **homophones**
- **A hypothesis 仮説**:
  - A language will have more homophones if it has fewer phonemes or more phonological restrictions
  - Problem: it is difficult to "measure" phonological restrictions
- Most restrictions are about how phonemes are combined to form **syllables 音節**
  - More restrictions -> fewer syllables that are possible
  - Also: fewer phonemes -> fewer syllables that are possible

Corpus Analysis of Homophones

- A better, testable hypothesis:
  - A language will have more homophones if it has fewer possible syllables

- To count homophones: check all the words in a dictionary!
  - CALLHOME Lexicon: words people use during phone calls
  - Six languages: Arabic, English, German, Mandarin, Japanese, Spanish
  - Calculate the proportion

- To count possible syllables: do a simple math!
  - Syllable: consonant(s) + vowel + consonant(s)
  - Number of possible syllables = number of possible consonants before a vowel
    - * number of possible vowels
    - * number of possible consonants after a vowel

Corpus Analysis of Homophones

- Languages really differ a lot!

- Japanese is unique!

- Almost 1/3 of words in Japanese CALLHOME lexicon are homophones
  - A lot more than other languages

- Phonemes and restrictions can be a reason
  - About 1,000 different syllables are possible in Japanese
  - Much fewer than languages like German and English
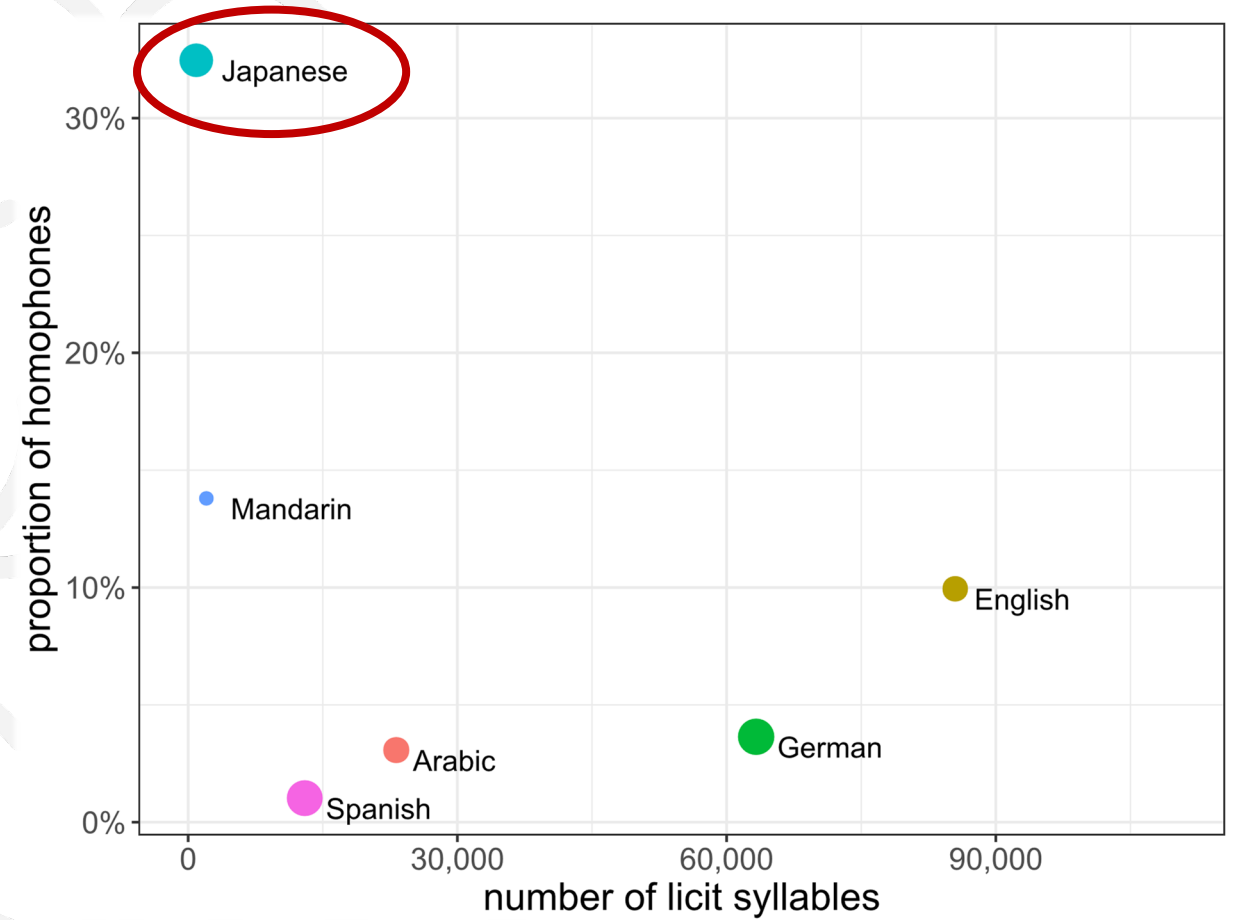
Corpus Analysis of Homophones

- The rule is not perfect

- Compared to German, English has both more homophones and more possible syllables
  - Bonus question: what can be the cause of this unusual pattern?

- Now we know that Japanese has many homophones, but what is the impact?
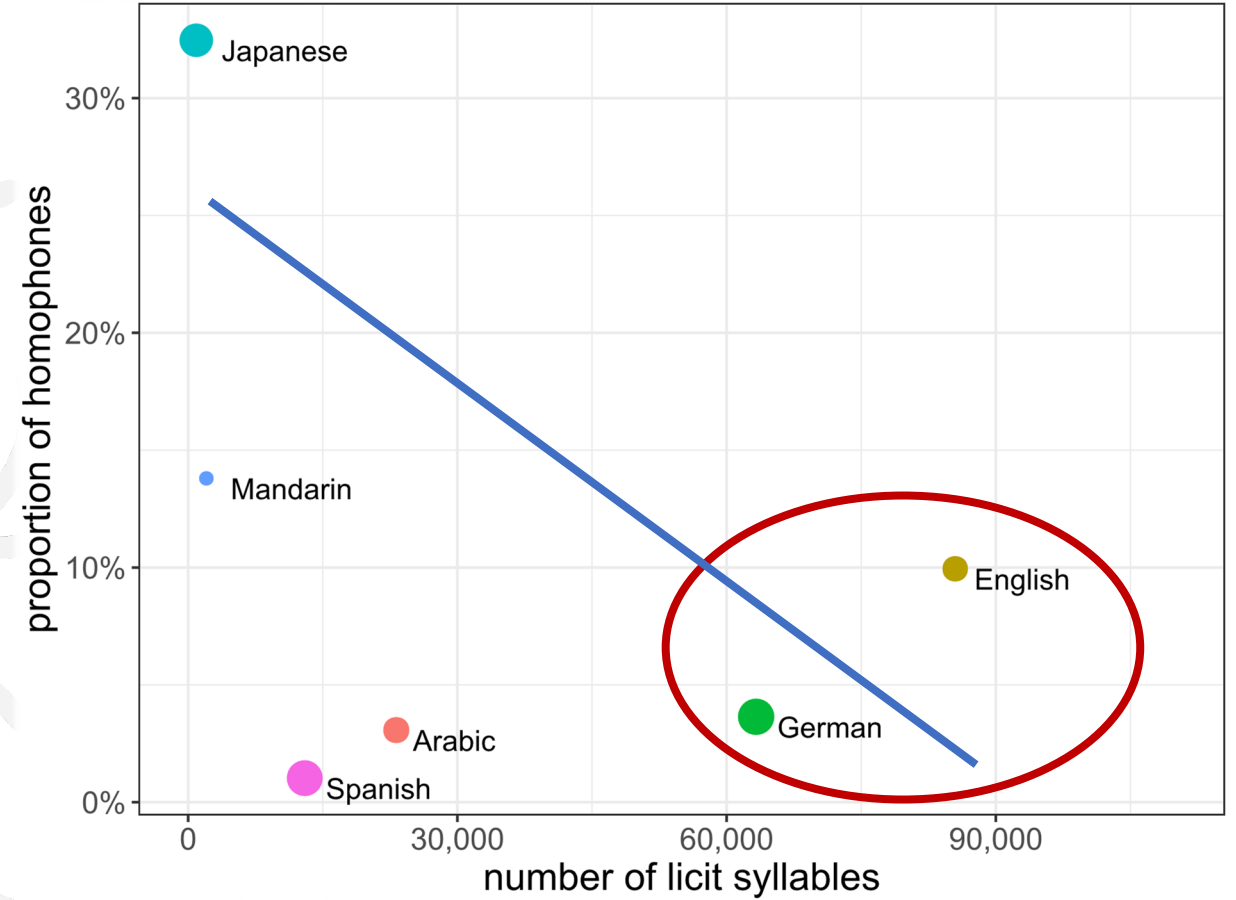
Contents

- **Self-Introduction**
  - Who am I?
  - What do I study?

- **Research Talk**
  - Corpus analysis ローパス分析
  - Behavioral experiment 行動実験

- **Message for you**

IRCN
International Research Center for Neurointelligence
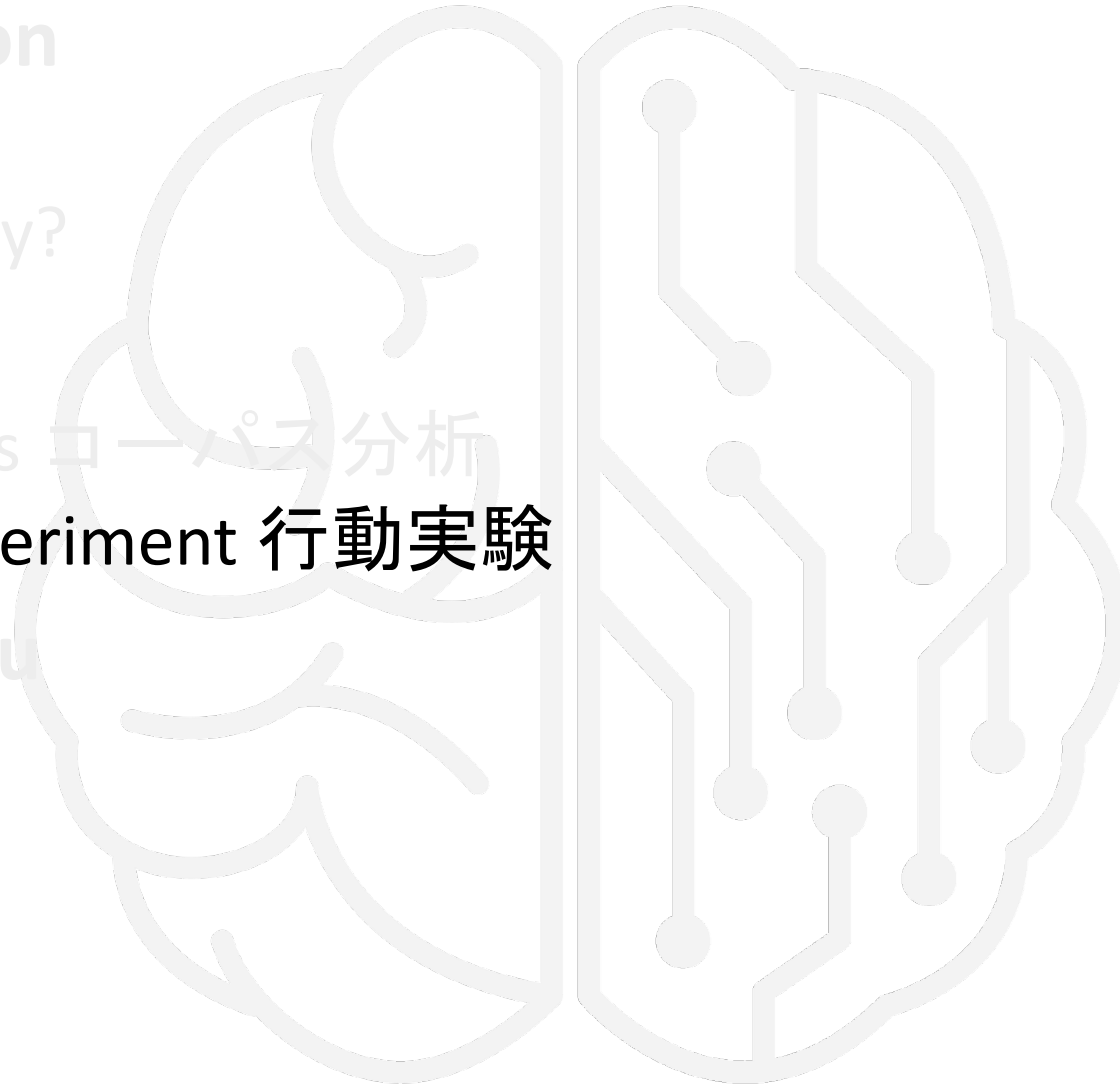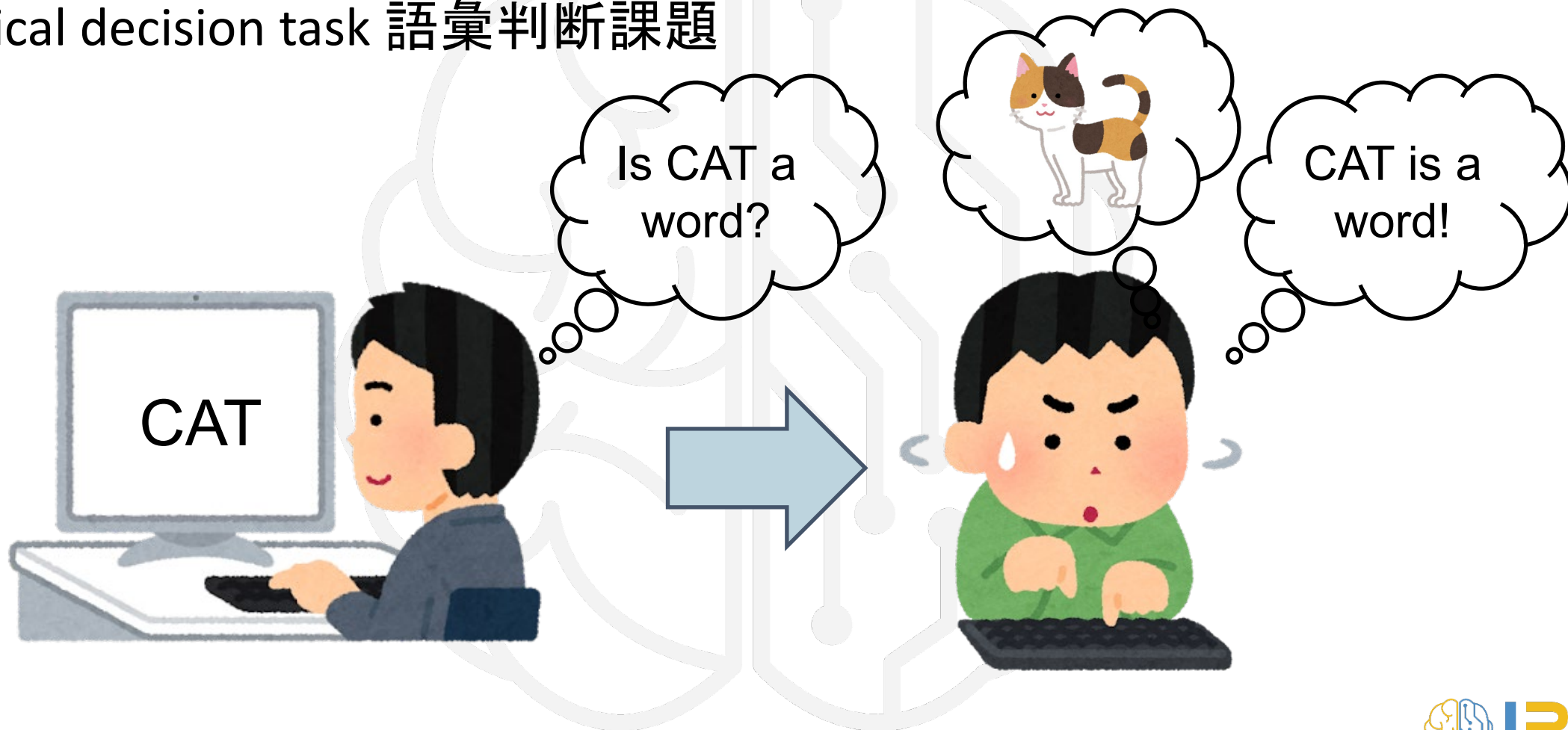
Homophones and Language Processing

- How do homophones influence language processing 言語処理?
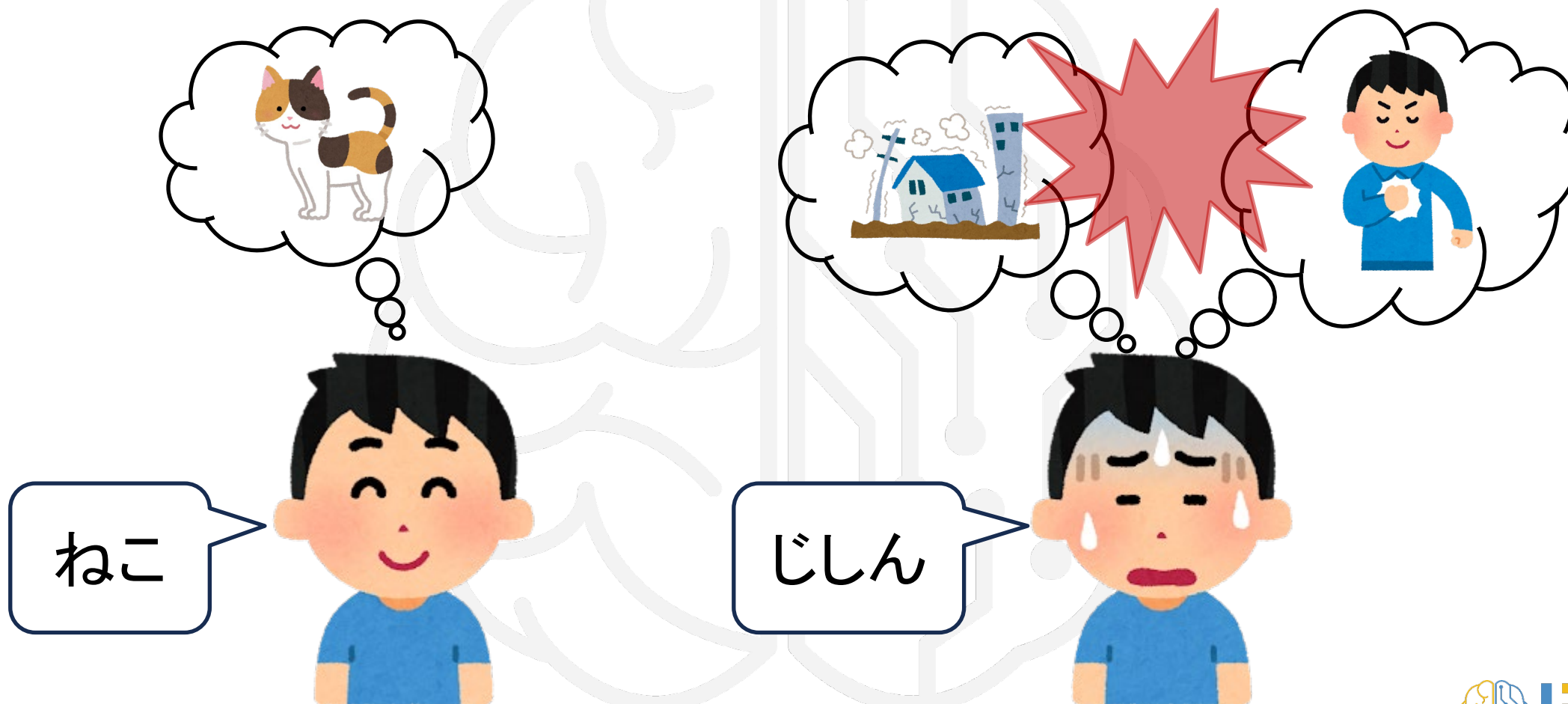
Homophones and Language Processing

- Homophones make language processing more difficult in English
- Lexical decision task 語彙判断課題

Homophones and Language Processing

- Homophones make language processing more difficult in English
- Lexical decision task 語彙判断課題

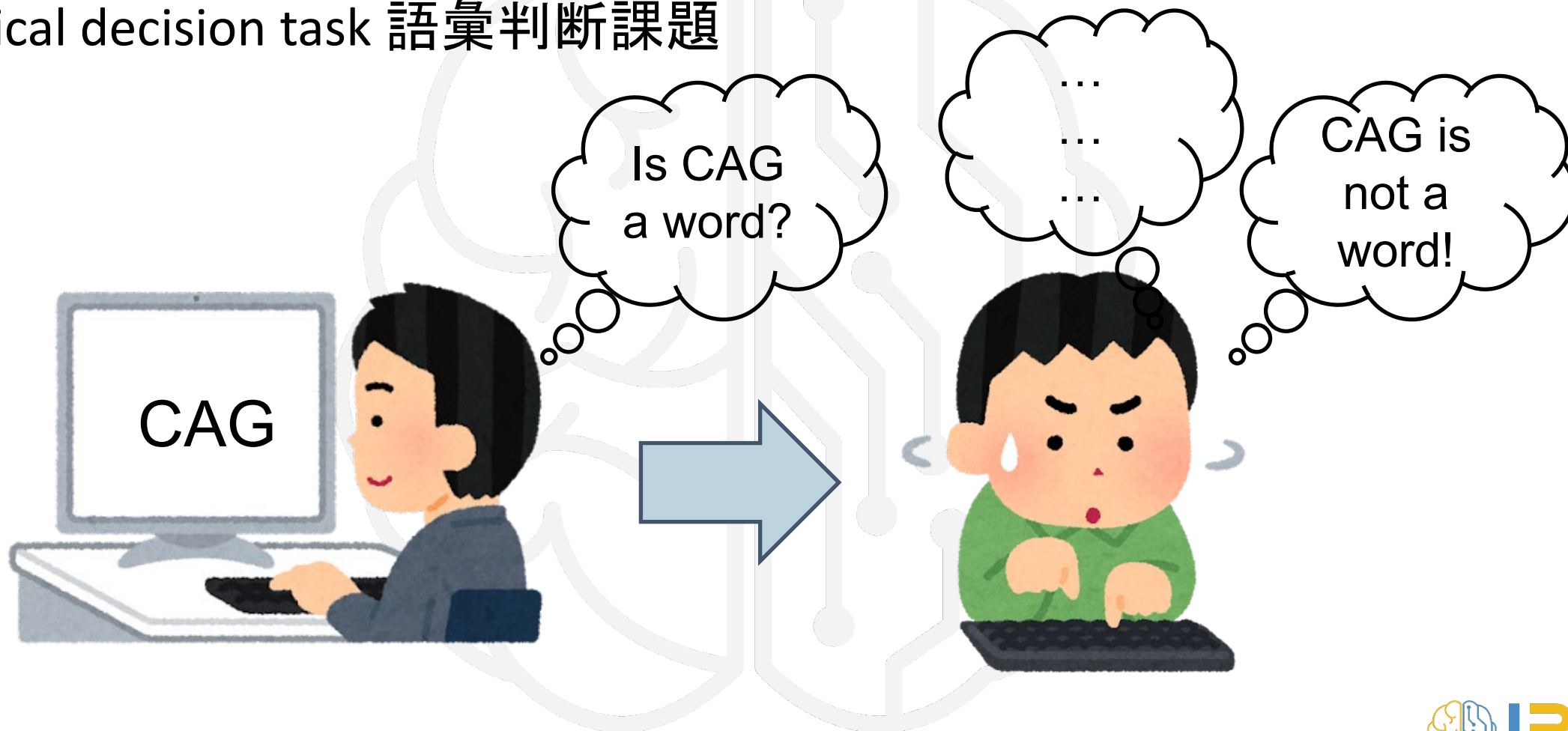# Homophones and Language Processing

- Homophones make language processing more difficult in English

- Lexical decision task 語彙判断課題
    - Decisions on homophones are longer
    - Deciding "NIGHT" is a word is more difficult than deciding "CAT" is a word

- Things are different for Japanese speakers
    - Deciding "地震" is a word is **NOT** more difficult than deciding "猫" is a word

- WHY?

Homophones and Language Processing

- Now we know that there are many more homophones in Japanese
  - Maybe Japanese speakers just have more experience

- But homophones in English also **LOOKS** more similar
  - KNIGHT – NIGHT
  - 地震 – 自信
  - Maybe the similar looking confuses English speakers

- What if they **hear**, rather than **see** the words?

Homophones and Language Processing

- Auditory lexical decision task 聴覚語彙判断課題

Homophones and Language Processing

- In this behavioral experiment, I examined how homophones influence language processing in English, Mandarin Chinese, and Japanese

- **Hypothesis**: the influence of homophones is smaller in languages with more homophones

- Homophones

  Japanese >> Mandarin > English

- Prediction: native speakers of Japanese should have the least difficulty, while native speakers of English should have the most difficulty
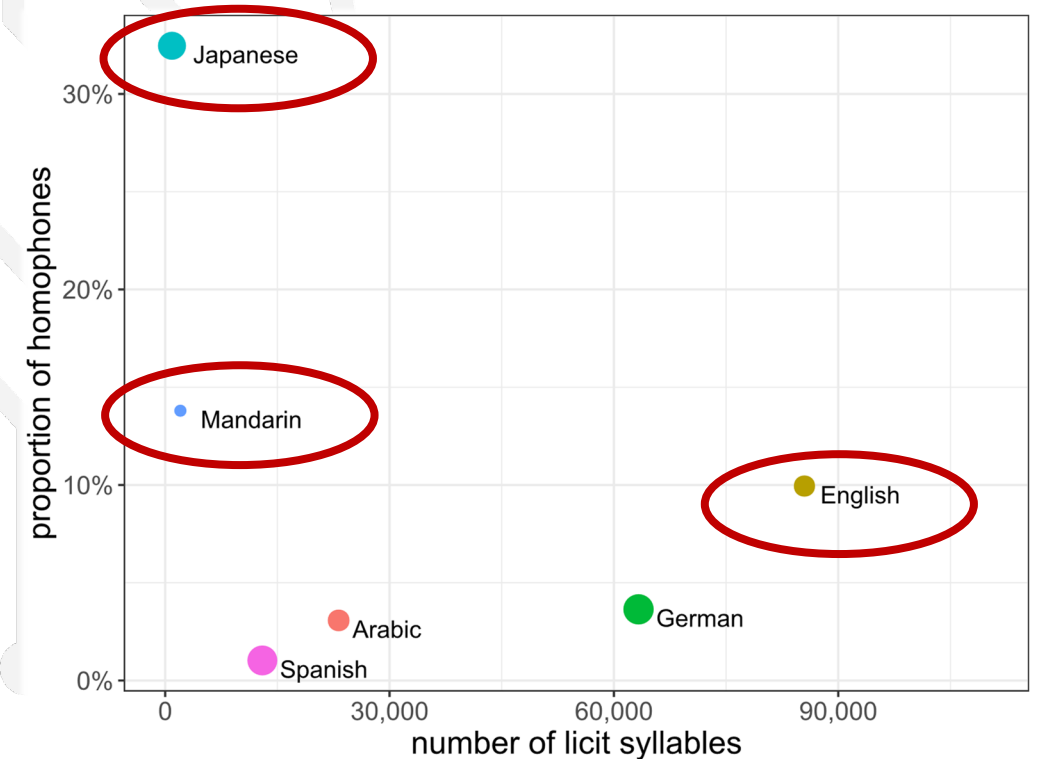
Corpus Analysis of Homophones

- The x-axis shows a measure of processing difficulty, which is based on response time 反応時間

- The result confirmed our prediction
  - Homophones did not cause any difficulty in language processing to Japanese speakers
  - But they slow down language processing of native speakers of Mandarin Chinese and English
  - The difficulty is even larger in English

$\beta = 0.057$

English

$\beta = 0.040$

Mandarin

$\beta = -0.025$

Japanese

-0.10   -0.05   0.00   0.05   0.10   0.15

more difficulty
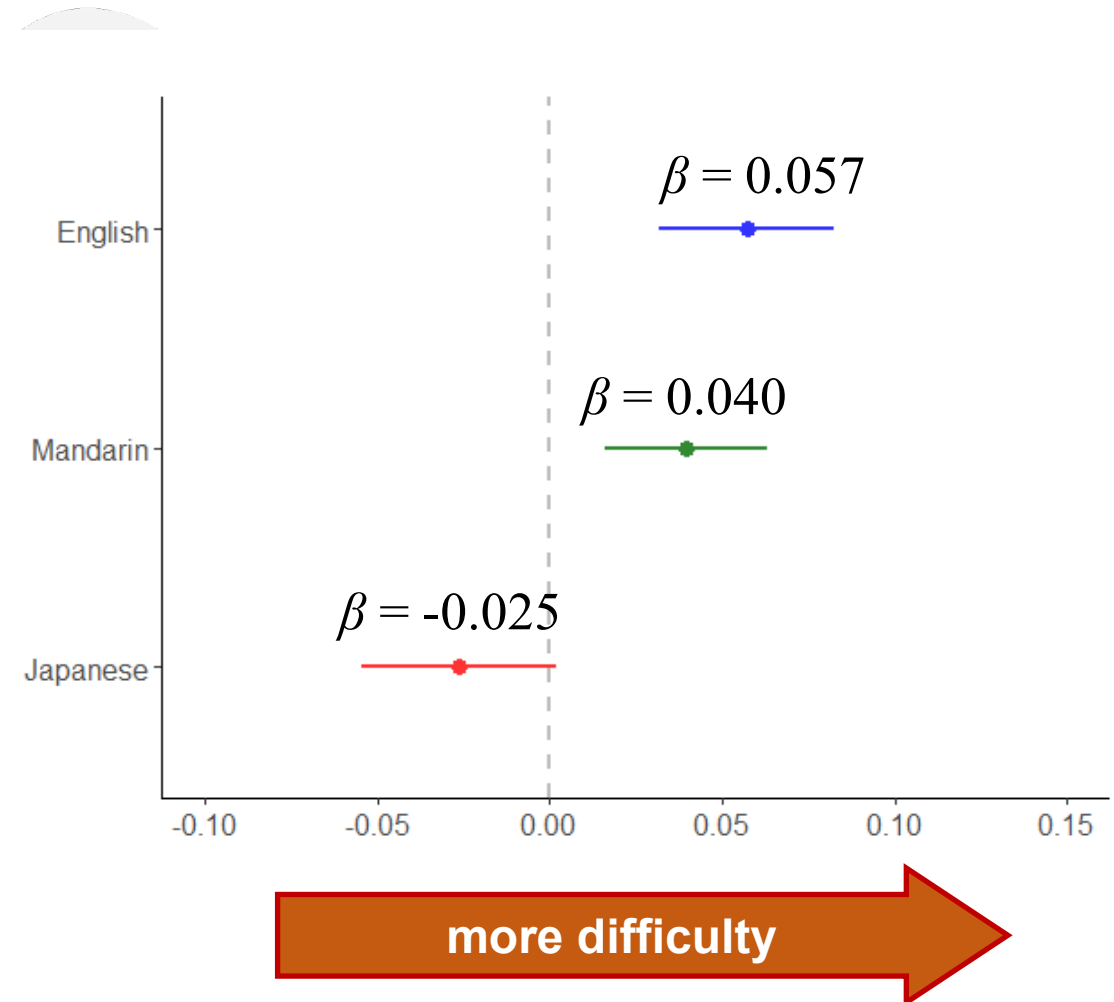
Conclusion of Research Talk

- There are many differences across languages
  - Just as we love to see different animals, we should appreciate the differences across languages
  - No language is superior or inferior to another
- Each language is a well-organized system
  - The sounds of a language and how they are combined can influence the vocabulary of this language
- Features of a language can influence how we process the language
  - Having less difficulty in one aspect of language processing does not mean being more clever
  - We all use our native language very efficiently
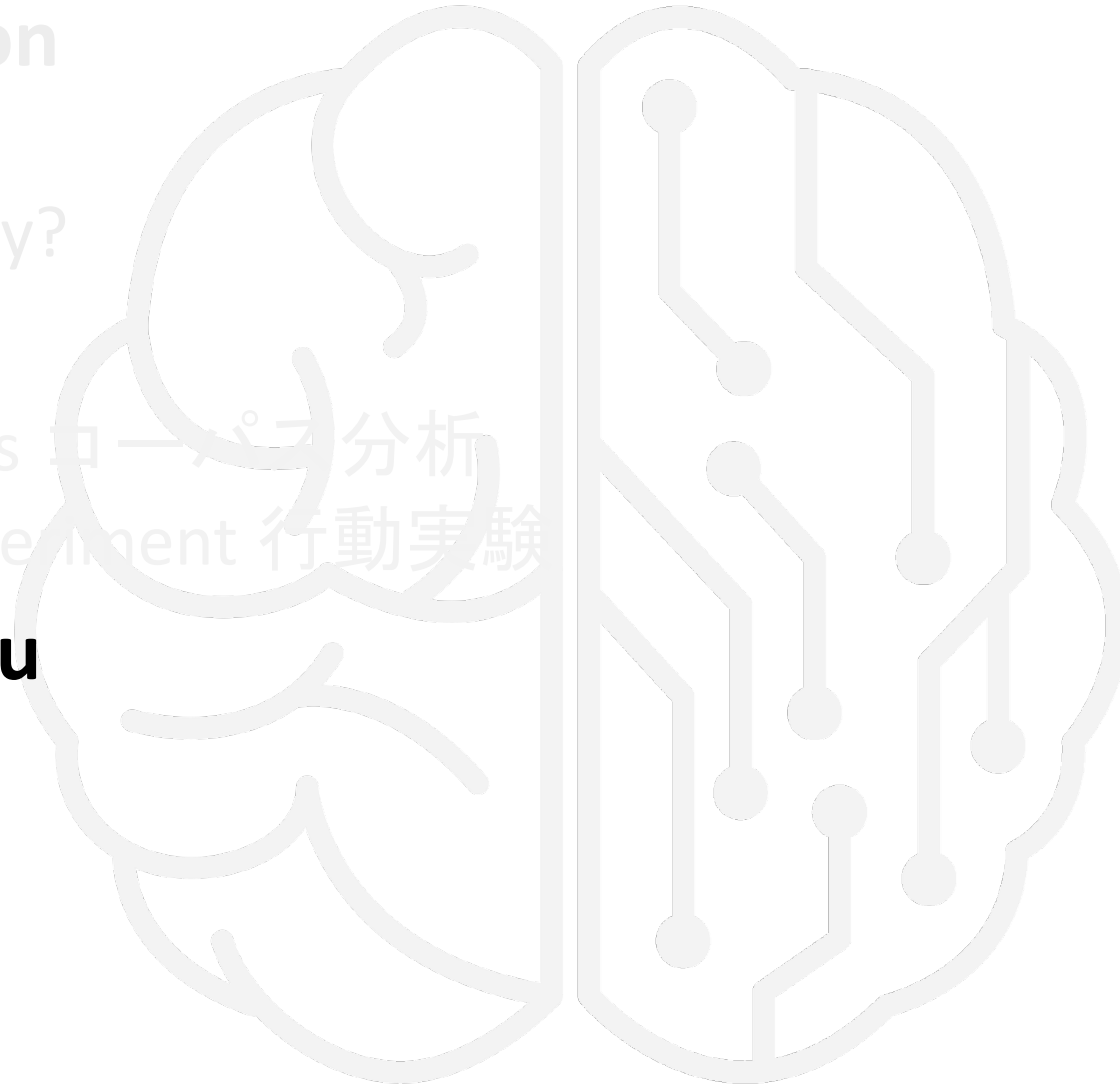
Contents

- **Self-Introduction**
  - Who am I?
  - What do I study?
- **Research Talk**
  - Corpus analysis コーパス分析
  - Behavioral experiment 行動実験

- **Message for you**

IRCN
International Research Center for Neurointelligence

Final Remarks and Messages

- I don't think I'm old/wise enough to give you any suggestion
  - Be skeptical about what I say :)
- Try to learn more, but not just about those taught in school
- Learn about yourself: what makes me "ME"?
  - What is my real passion? What do I value most?
- Learn about the world (from the world around you)
  - Everything in the world has its own logic
  - The more you see, the more you can see
- Learn to balance happiness in the short and long term
  - What gives you fulfillment in the long term may not make you happy at the moment
  - But happiness can go away if you wait too long

IRCN
International Research Center for Neurointelligence

Thank you for listening!

ご清聴ありがとうございました！