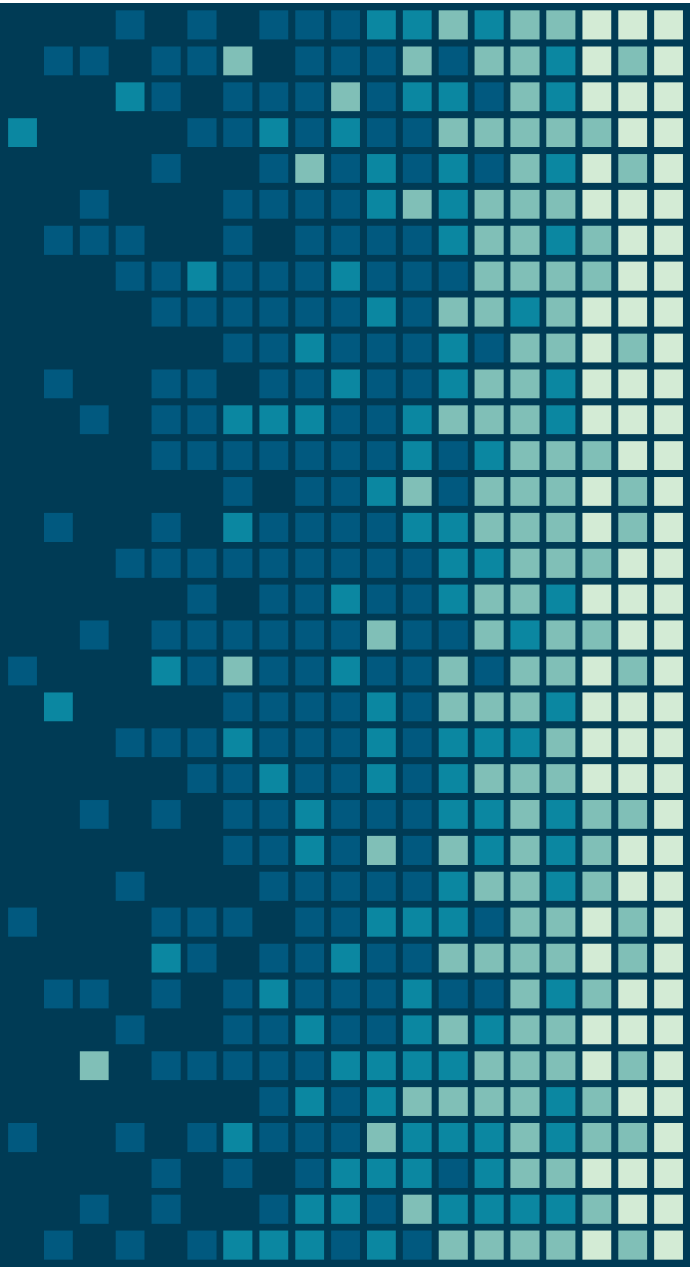
The Brain in Translation

Linguistic features of English research articles authored by Japanese neuroscientists

Mark Streer • TRAN-6602-080 • 2022-04-04



Context

- Technical translator
  - Japanese>English
  - ~400 research articles from 2013-2021

- “Japan-glish”
  - L2 English characteristic of L1 Japanese authors
  - L1 English translated by novice J-E translators

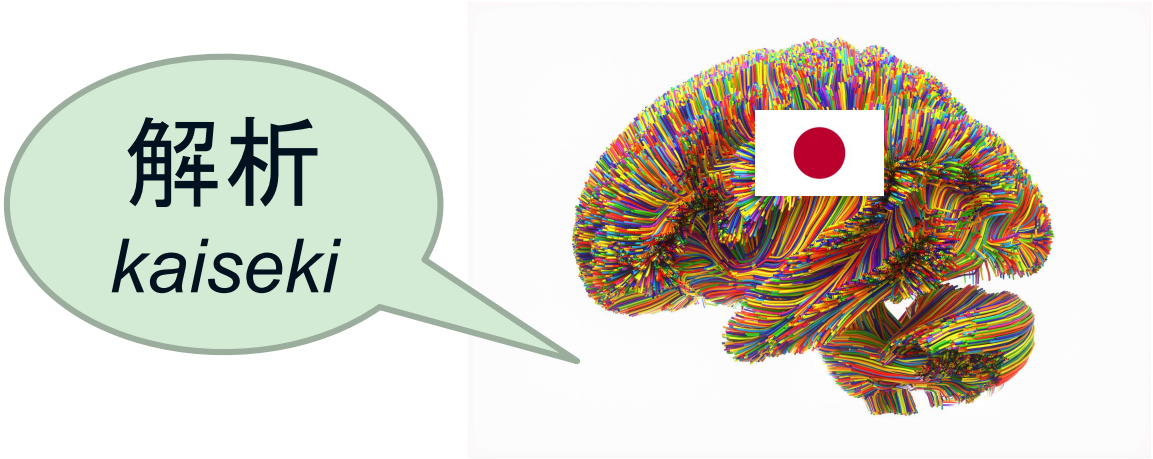
Feature #1: Lexical constraints

- ESP: Words of Latin/Greek origin connected by Germanic syntax
- JSP: Words of Chinese origin (*kanji, kango*) / Japonic syntax
  (and some European loanwords)

Brain	脳
Neuroscience	脳科学
Encephalography	脳造影
Cerebral cortex	大脳皮質
Cerebellar ataxia	小脳性運動失調

Feature #1: Lexical constraints

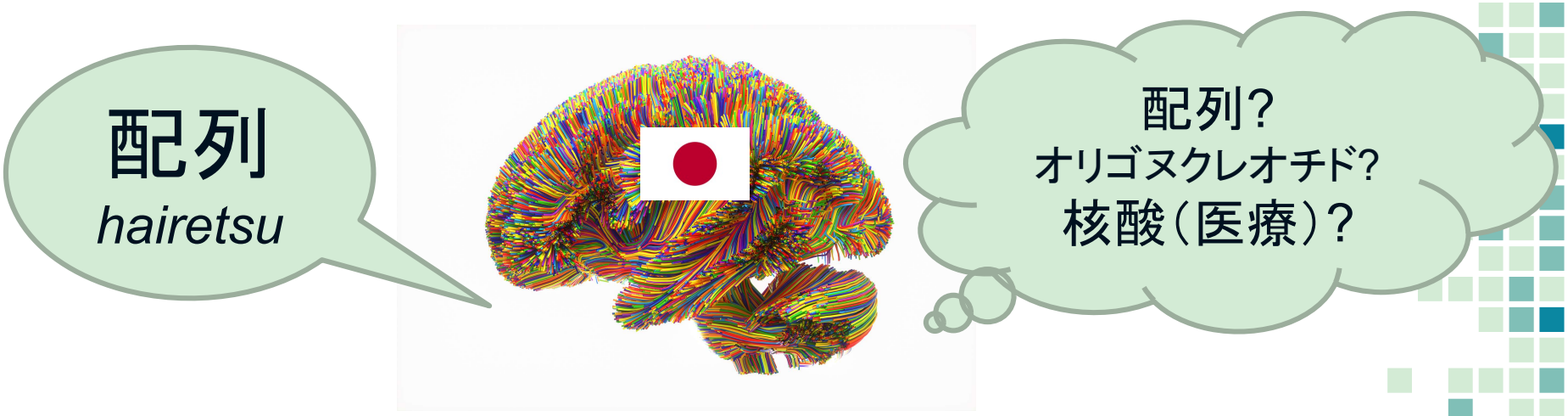
- …the **analysis** of spatial genetic variation within continuously…
  …describe **decomposition** of SNP weights used to construct…
  …program for **parsing** data files in the popular FastQ format…



- …the **analysis** of spatial genetic variation within continuously…
  …describe **analysis** of SNP weights used to construct…
  …program for **analysing** data files in the popular FastQ format…

Feature #1: Lexical constraints

- …the known amino acid **sequence** of bovine pancreatic trypsin…
  …5′-modified **oligonucleotides** were examined as substrates…



- …the known amino acid **sequence** of bovine pancreatic trypsin…
  … 5′-modified **sequences** were examined as substrates …

Feature #2: Long noun compounds

- BRCA1遺伝子変異
  BRCA1-idenshi-hen'i
  lit. 'BRCA1-gene-mutation'
  BRCA1 [gene] mutation
  mutant *BRCA1*

- 経口血糖降下剤治療
  keikō-kettō-kōka-zai-chiryō
  lit. 'oral-blood-sugar-drop-agent-therapy'
  oral hypoglycemic [drug] therapy
  "treated with an oral hypoglycemic…"

- 唾液腺摘出マウス
  daeki-sen-tekishutsu-mausu
  lit. 'saliva-gland-excision-mouse'
  "parotid-ectomized mouse"
  "mouse model produced by salivary gland excision"

Feature #3: Characteristic suffixes

- ~量　　*ryō*　　amount, quantity
- 課題の難易度と前頭連合野の**活性化量**は関係している.
  Task difficulty was associated with **[the amount of] activation** in frontal association cortex.
- ~値　　*chi*　　level, value
- 投与前の平均朝食前空腹時**血糖値**を共変量とした.
  Baseline mean fasting **blood sugar [level]** was set as a covariate.

- ~群 → (study) group, arm; ~液 → (liquid) solution;
  ~基 → (functional) group, moiety, unit; etc.

Mini-Corpora

- JP-EN: RAs from Japanese institutions (n=10)
- EN-EN: RAs by U.S./European* institutions (n=10)
- Constructed/analyzed using Python (NLTK, scispaCy)

First author	Date	Title	Subject Area 1	Subject Area 2	Subject Area 3
Yomogida	2014	The Neural Basis of Event Simulation: An fMRI Study	fMRI	Cognition	Prefrontal cortex
Dixon	2012	The Decision to Engage Cognitive Control Is Driven by Expected Reward-Value: Neural and Behavioral Evidence	fMRI	Cognition	Prefrontal cortex
Minamoto	2014	Extrapunitive and Intropunitive Individuals Activate Different Parts of the Prefrontal Cortex under an Ego-Blocking	fMRI	Aggression	Prefrontal cortex
Pawliczek*	2013	Anger under Control: Neural Correlates of Frustration as a Function of Trait Aggression	fMRI	Aggression	Prefrontal cortex
Ohta	2013	Syntactic Computation in the Human Brain: The Degree of Merger as a Key Factor	fMRI	Language	Memory
Himmelstein	2018	Linguistic analysis of the autobiographical memories of individuals with major depressive disorder	fMRI	Language	Memory
Watanabe	2012	Diminished Medial Prefrontal Activity behind Autistic Social Judgments of Incongruent Information	fMRI	Emotions	Facial expressions

Hypothesis testing

Compared with EN-EN, JP-EN has…

- H1: Lower lexical diversity (type/token ratio)
- H2: Longer noun *n*-grams
- H3: Greater frequency of lexemes:
  - /amount/, /level/, /value/
    (corresponding to Japanese suffixes)
- H4: Greater frequency of phrases:
  - "in this study", "in addition", "therefore"
    (corresponding to Japanese discourse markers)
- H5: Greater frequency of lexeme: /consider/
- H6: Lower frequency of "BOLD response"

streerm
/METIS-Porfolio

The Brain in Translation

- scispaCy
- Text mining
- Corpus analysis

myCorpus: NLP for Translation Providers

- Topic modeling (LSA / NMF / LDA)
- Clustering (k-means)
- Text preprocessing (‘data wrangling’)