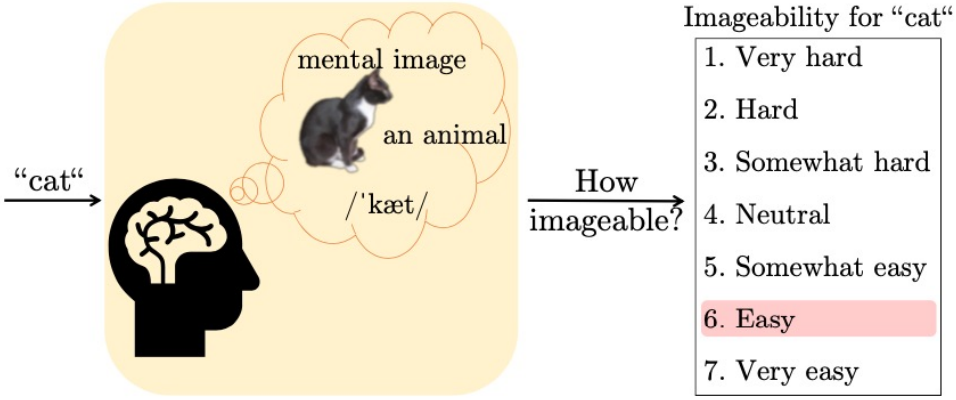
A Multi-modal Dataset for Analyzing the Imageability of Concepts Across Modalities

Marc A. Kastner[1], Chihaya Matsuhira[2], Ichiro Ide[2], Shin'ichi Satoh[1]

(1) National Institute of Informatics, Japan
(2) Nagoya University, Japan

Imageability

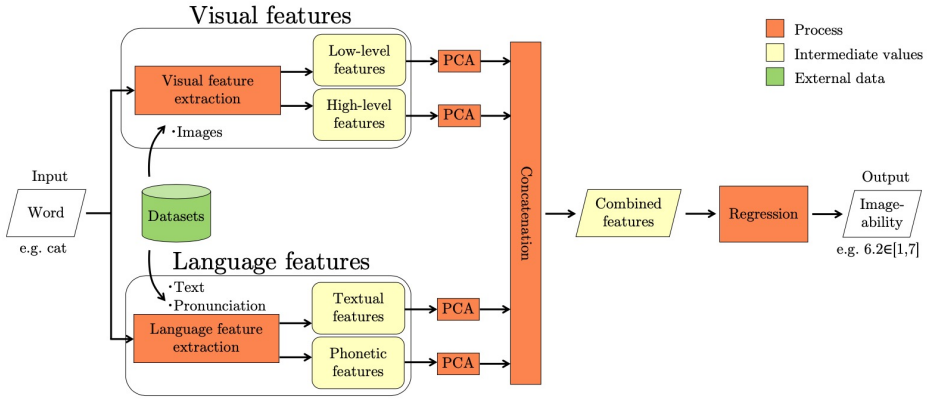
- Estimating the mental image of a human regarding a word or concept (“imageability”) [1]
  - For example:
    - ”*cat*” -> clear mental image -> “*easy*”
    - “*peaceful*” -> no clear mental image -> “*hard*”



[1] Paivio et al., “Concreteness, imagery, and meaningfulness values for 925 nouns.,” J. Exp. Psychol, 1968.

Tri-modal approach of this work

- Previous work[2] has been combining textual and visual knowledge to estimate imageability



- In this work, we want to look at different modalities seperately.
  - Can different modalities tell us different things about the same concept?

[2] Matsuhira et al., “Imageability Estimation using Visual and Language Features”, ICMR 2020.

Dataset

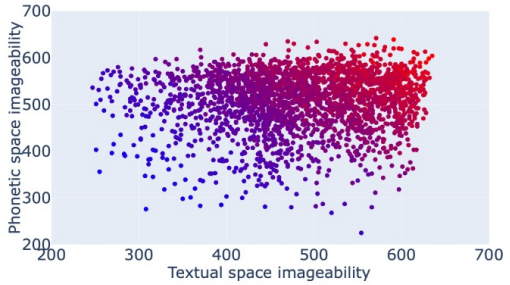
- We generated three imageability estimations for 2,430 words
  1. Image data mining (Visual)
  2. Text data mining (Language)
  3. Pronunciation data analysis (Phonetic)

- Data is available at: https://github.com/mkasu/imageabilitycorpus

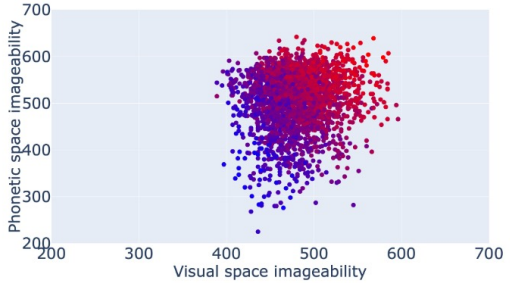
Analysis

- Looking into
  - Cross-modal outliers
  - Visualization of phonetic space
  - Relationship of etymology and imageability
  - Generating per-word mozaics to understand visual space

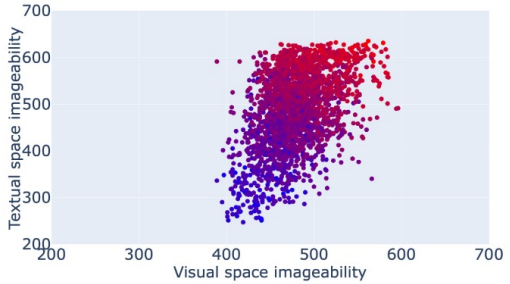
	Visual space imageability	Textual space imageability	Phonetic space imageability
Middle English	477 (35)	480 (85)	513 (61)
English	483 (36)	485 (88)	492 (65)
Old English	471 (30)	476 (80)	513 (60)
Latin	480 (28)	485 (70)	489 (68)
French	486 (30)	497 (75)	504 (55)
Old French	478 (29)	481 (76)	497 (65)
Middle French	486 (29)	489 (76)	507 (62)
Ancient Greek	478 (25)	497 (58)	499 (64)
Italian	490 (33)	521 (95)	500 (64)
Anglo-Norman	481 (25)	486 (75)	521 (58)
Hebrew	473 (22)	510 (72)	518 (73)
Spanish	489 (34)	499 (94)	513 (51)
Old Norse	483 (37)	480 (73)	516 (33)
Other	481 (33)	485 (65)	511 (59)



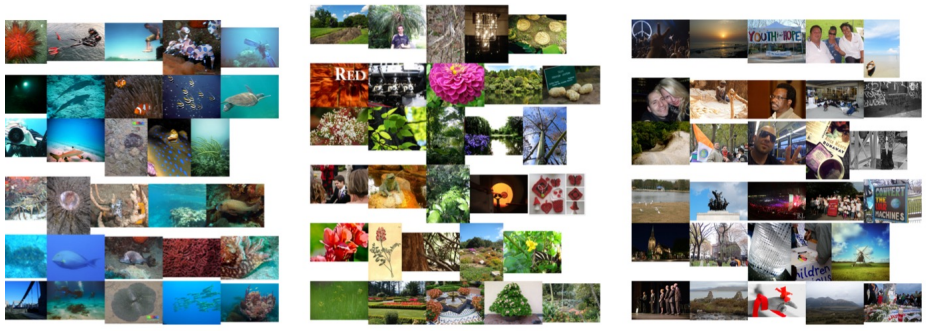
(a) 2D spatial distribution of imageabilty values across phonetic and textual modalities.



(b) 2D spatial distribution of imageabilty values across phonetic and visual modalities.



(c) 2D spatial distribution of imageabilty values across textual and visual modalities.



running (408): R AH1 N IH0 NG
nature (563): N EY1 CH ER0
command (414): K AH0 M AE1 N D
underwater (454): AH1 N D ER0 W AO2 T ER0
botanical (355): B AH0 T AE1 N IH0 K AH0 L

Thank you!

Looking forward to your visit during the poster session

mkastner@nii.ac.jp **https://www.marc-kastner.com** **@mkasu**