Eye can tell you how Japanese morphographic words are recognized

— A lexical decision with eye-tracking study —

Koji Miwa[1], Gary Libben[2], Ton Dijkstra[3], & Harald Baayen[1]

[1] University of Alberta, Canada

[2] University of Calgary, Canada

[3] Radboud Univeristy,

DCC language group meeting

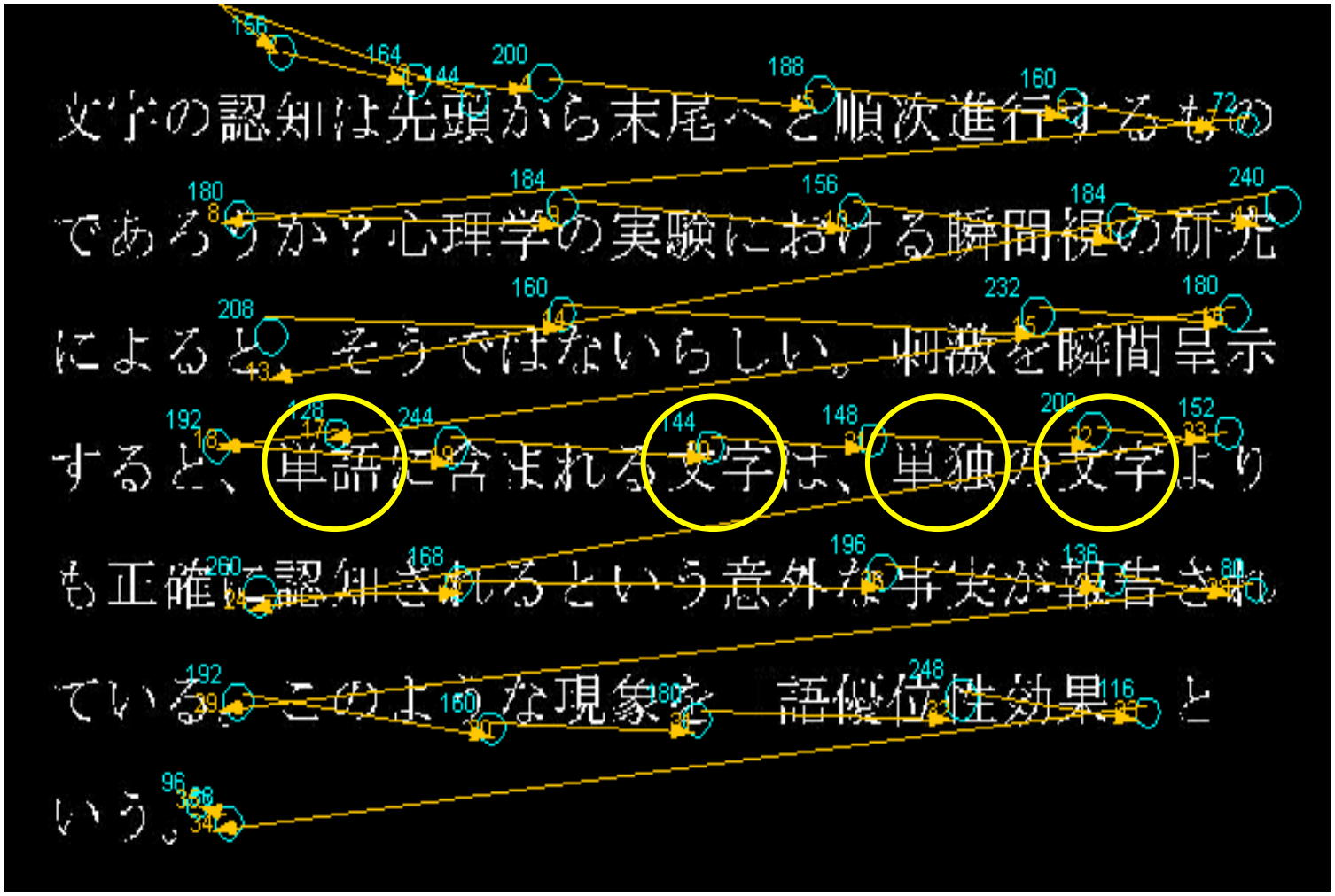
September 28, 2009

Reading a Japanese Text

- In Japanese, there are three types of orthography
  1. Morphographic kanji （阿蘭陀）
  2. Moraic hiragana （おらんだ）
  3. Moraic katakana （オランダ）
  (4. Alphabetic Romaji) （Nezarando)

- Roughly 70% of Japanese words are two-kanji-character words.

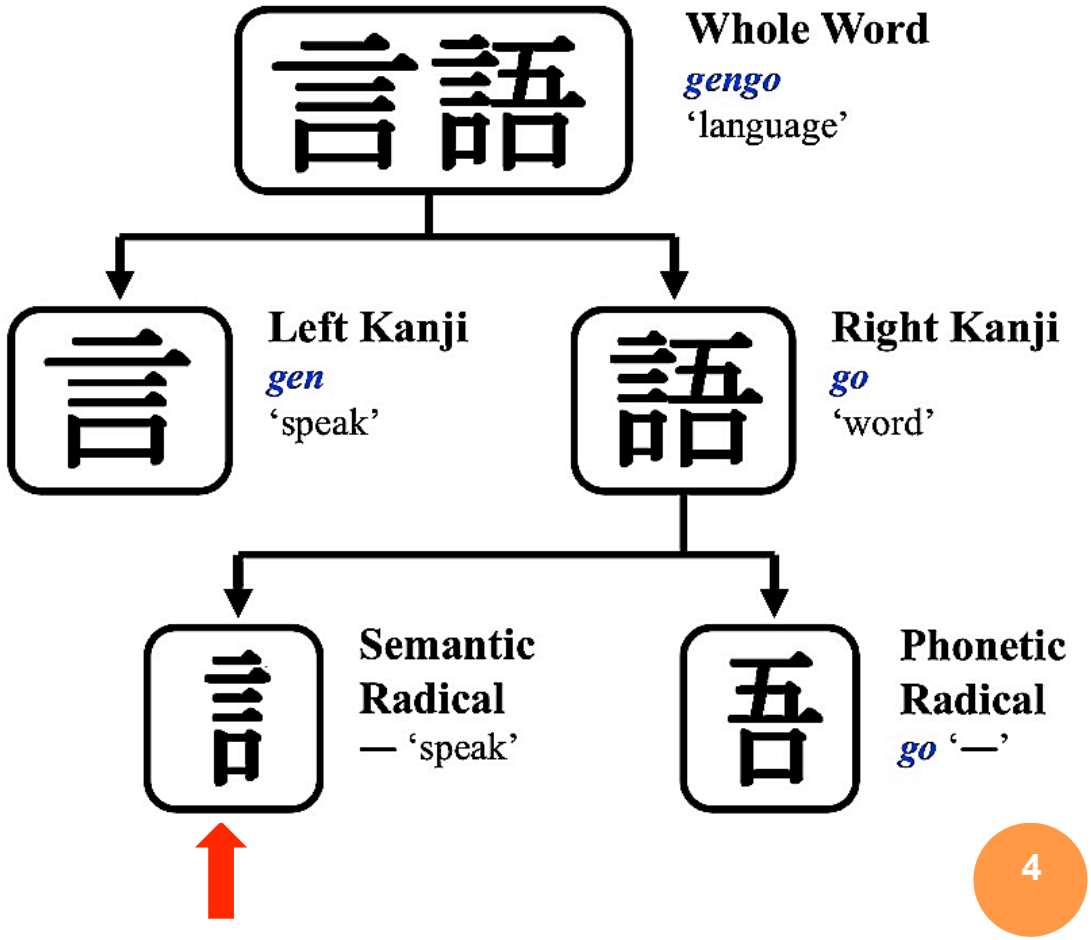
Reading a Japanese Text



c.f. Kajii,Nazir, & Osaka (2001)

Japanese Kanji Logography

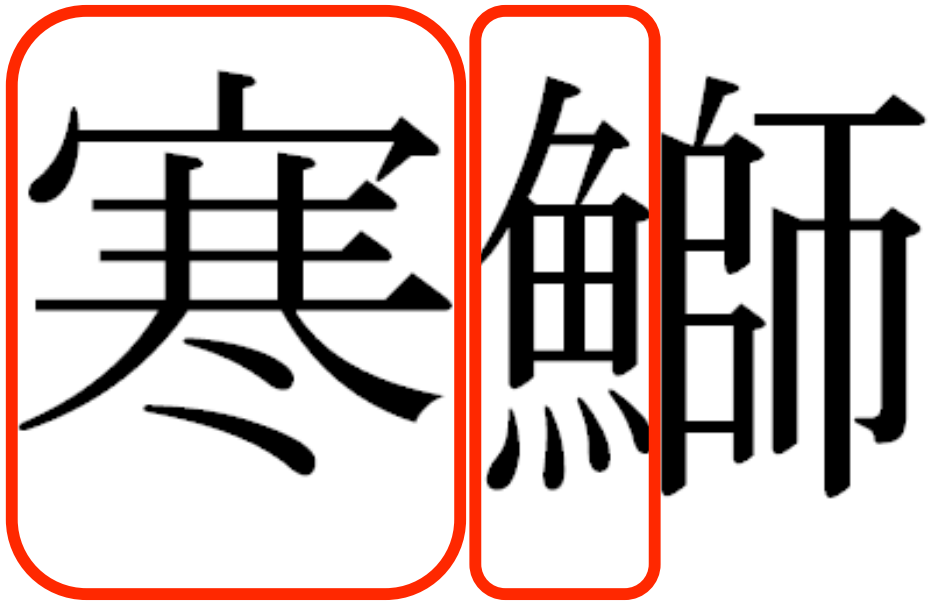
- The majority of Japanese "simplex" words consist of morphographic characters.
- Majority of characters further consist of a semantic radical and a phonetic radical.
- 1,945 basic characters in total
- How are two-character words recognized ?



Purely orthographic morpheme ?

Recognizing an Unfamiliar Word

- No need to know the whole
- No need to know its pronunciation



Studies on Compound word recognition

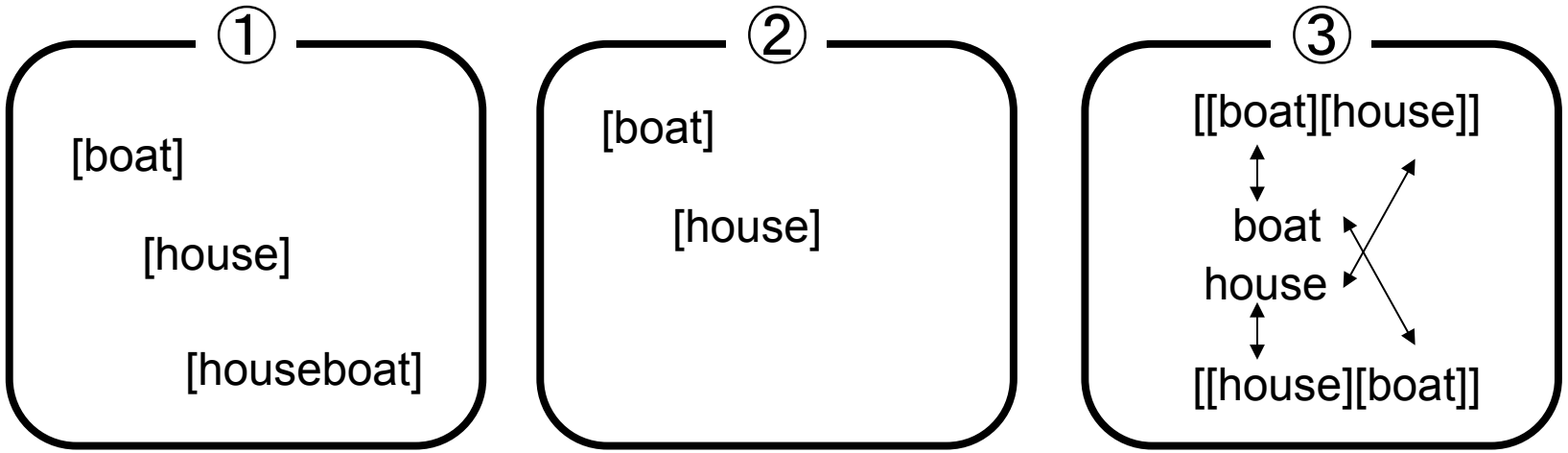
- How efficiently are complex words represented and processed ?

(Libben, 2005, OUP)

① Maximizing computational efficiency? => Pre-sorting is necessary

② Maximizing storage efficiency? => No advantage for HF words

③ Maximizing opportunities? => Lexicalize words but still process everything



- Sub-lexical vs. Supra-lexical Single-route vs. Dual-route vs. Multi-route
- What architecture best explains the complexity in Japanese ?

Whole Two-Character Words as Representational Units

- Morton, Sasanuma, Patterson, & Sakuma (1992)

  - Full word forms are recognition units

  - Only identical full forms cause priming effect

    (in a long-lag priming experiment)

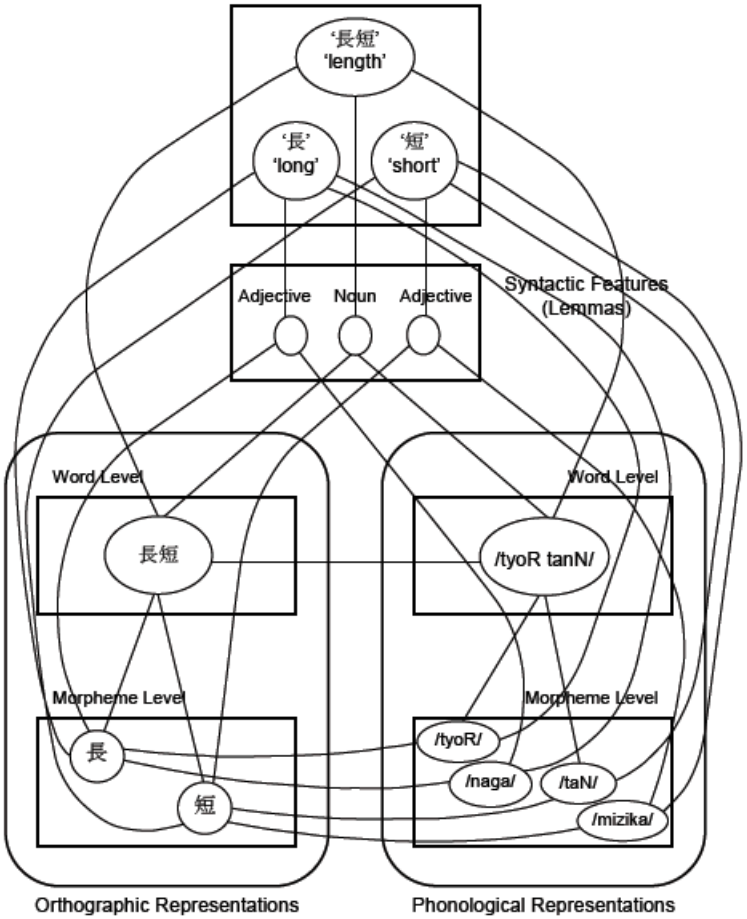
Characters as Representational Units

- Hirose (1992)'s partial constituent priming study
  - Left characters are important
  - Two-character words with the same left character are tied together in the mind

- Tamoka & Hatsuzuka (1995)'s lexical decision
  - Right characters are important
  - This is because of the left-to-right reading sequence

Characters as Representational Units

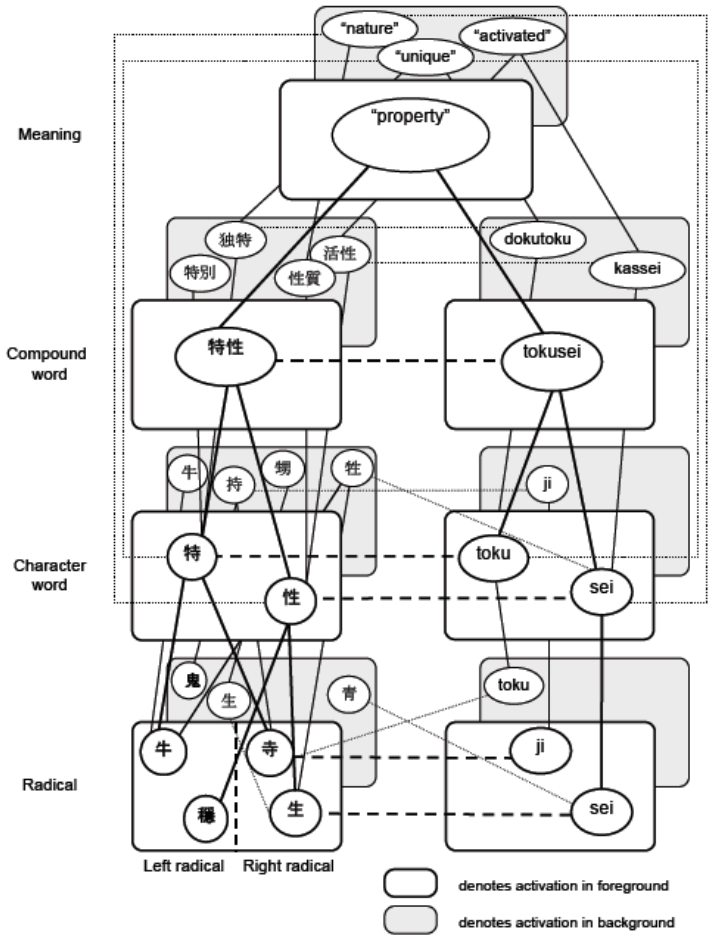
Tamaoka & Hatsuzuka (1998)



- Both the whole word unit and the character unit has their semantic representation.
- No radical-level representation
- The order of processing is under-specified
- Are radicals important ?

Radicals as Representational Units

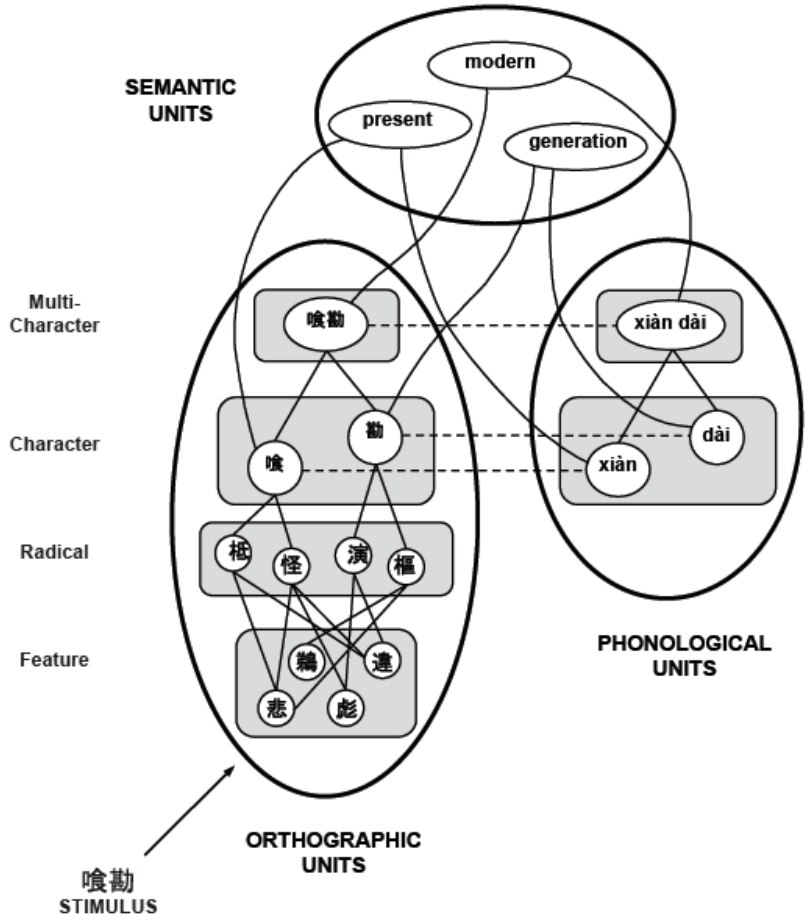
Saito (1997)



- There is a radical-level
- Only the whole word unit has a semantic representation
- Characters are not considered as morphemes
- A single-route model
- Do characters behave like morphemes?
- Are radicals processed first?

Radicals as Representational Units

Taft & Zhu (1997)



- Both the whole word unit and the character unit has their semantic representation.
- Semantic radicals are not considered as orthographic morphemes.
- A single-route model
- Do radicals behave like morphemes?
- Are radicals processed first?

What We Want to Study

- **Language-general issues**
  1. Locus of whole word frequency effect (c.f. Kuperman et al., 2009)
  2. Relative importance of the left and the right constituents in compounds
  3. Order of activating the constituents and the whole word representations
  4. Role of attention (what do eye fixations reflect?)

- **Language-specific issues**
  1. Whether semantic radicals contribute in the larger context of default two-character word recognition (no evidence so far).
  2. Status of characters (morphemes? or orthographic components?)
  3. Status of radicals (orthographic morphemes? or orthographic components?)

Lexical Decision with Eye-Tracking

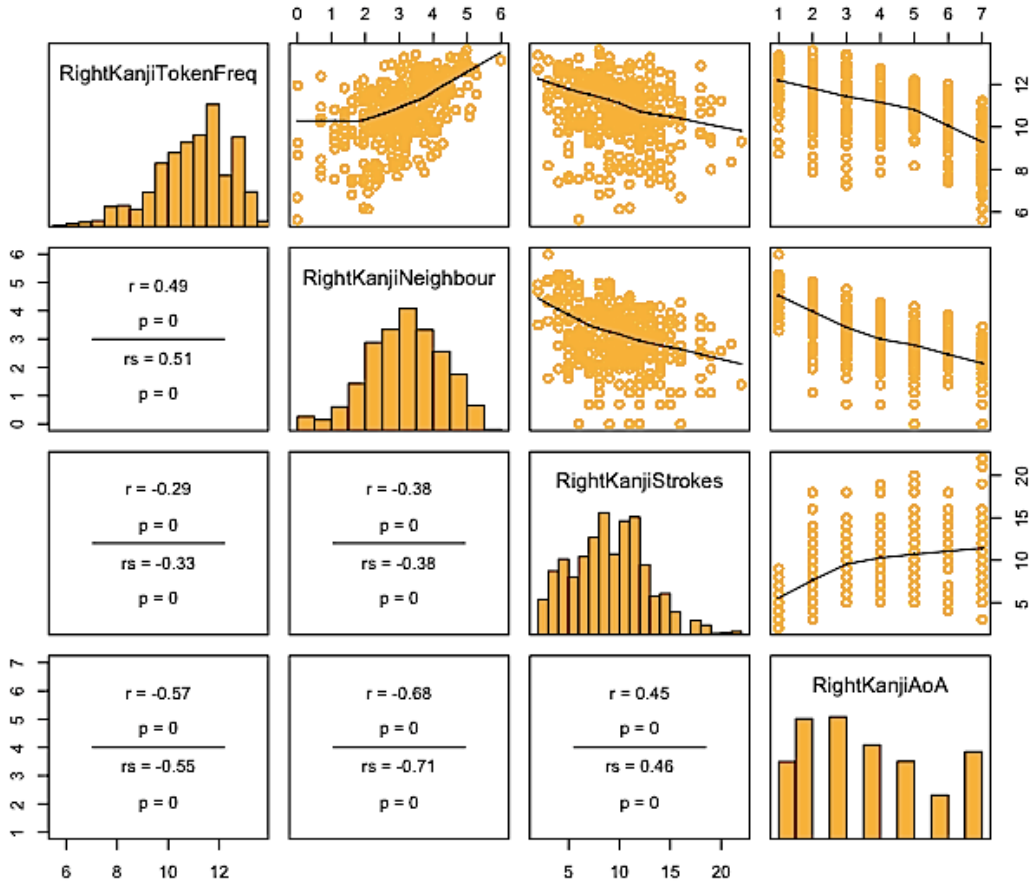
- Participants
  - 21 native Japanese readers in Canada
- Materials
  - **708** two-character words (=> 555)
  - **708** nonwords (various types)
- Procedure
  - 472 words in one session * 3 sessions
  - Fixation point manipulation
    (left, central, right)

20 Lexical Properties We Considered

Type	Predictors
Feature (、、✓)	· Left Kanji Radical Strokes (Resid), · Right Kanji Radical Strokes (Resid) · Left Kanji Strokes (Resid), · Right Kanji Strokes (Resid)
Radical (氵)	· Left Kanji Radical Combinability, · Right Kanji Radical Combinability · Left Kanji Radical Token Freq (Resid), · Right Kanji Radical Token Freq (Resid)
Character (港)	· Left Kanji Neighbour (Resid), · Right Kanji Neighbour (Resid) · Left Kanji Token Freq, · Right Kanji Token Freq · Left Kanji AoA (Resid), · Right Kanji AoA (Resid) · Left Kanji Homophones, · Right Kanji Homophones · Left Kanji OnKun, · Right Kanji OnKun
Whole word (空港)	· Whole Word Freq, · Google Doc Freq (Resid)
Individual	· Length of Stay Canada
Task	· PrevRT, · Trial, · Previous Trial Correct, · Eye Position, · Fixation

Residualization to Handle Multicollinearity

- Complex correlational structure is inherent among lexical distributional properties.



RT Analysis

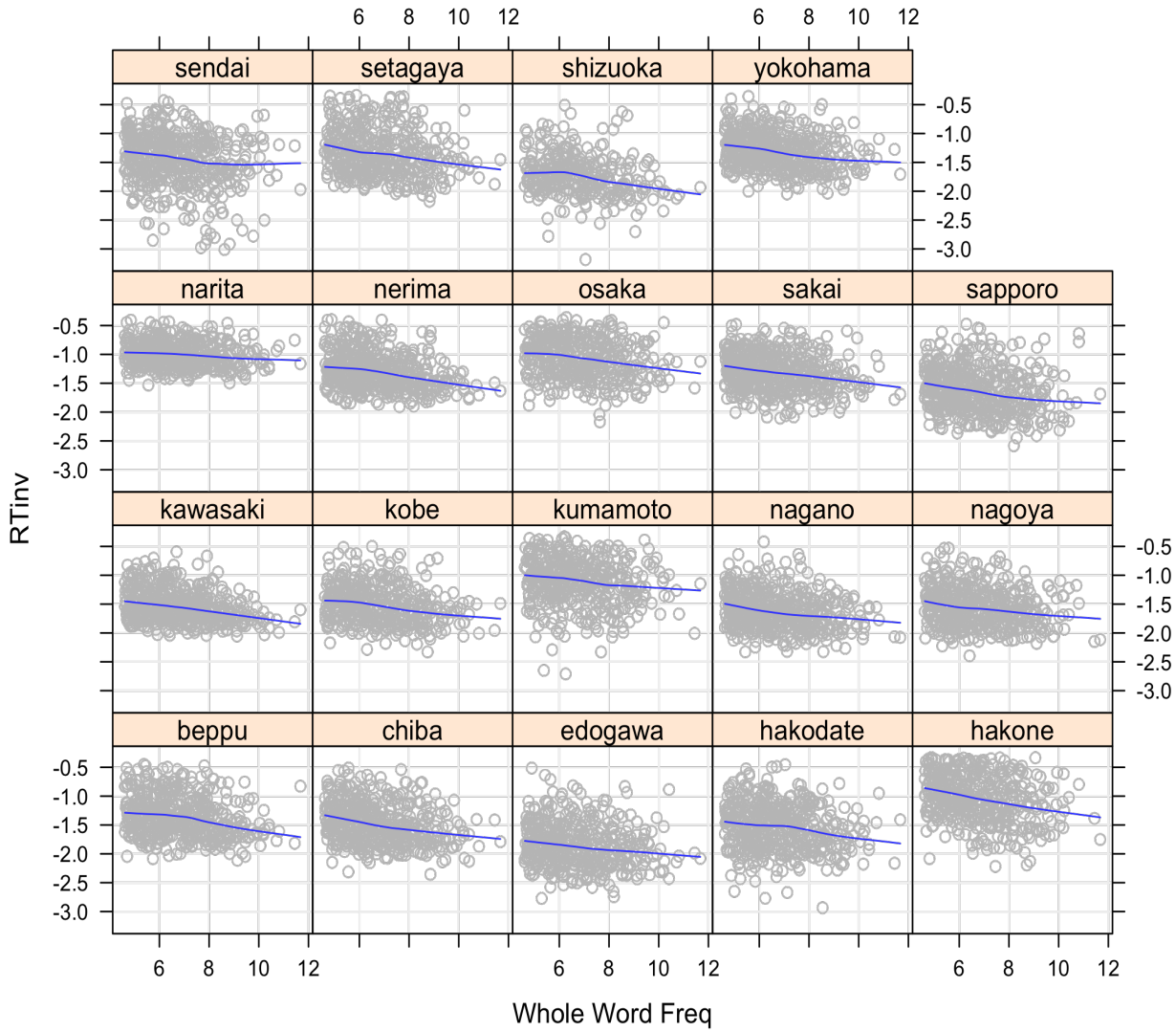
- A model that stores frequent complex word forms for computational efficiency

  => there should be non-linear/ catastrophic whole word frequency effect

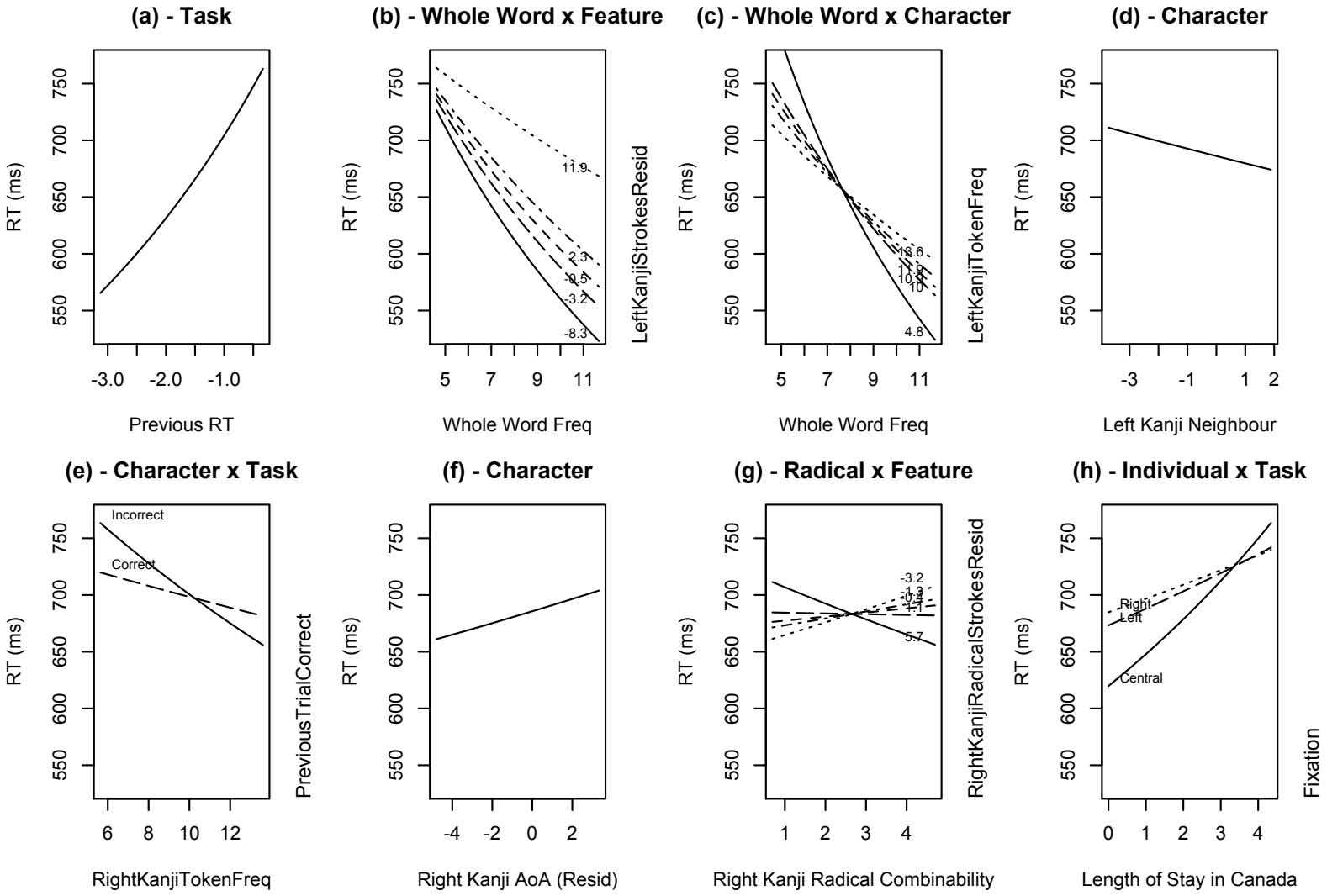
- A model that always compute complex words from the scratch without storing complex word forms

  => Complex words' frequency is not important

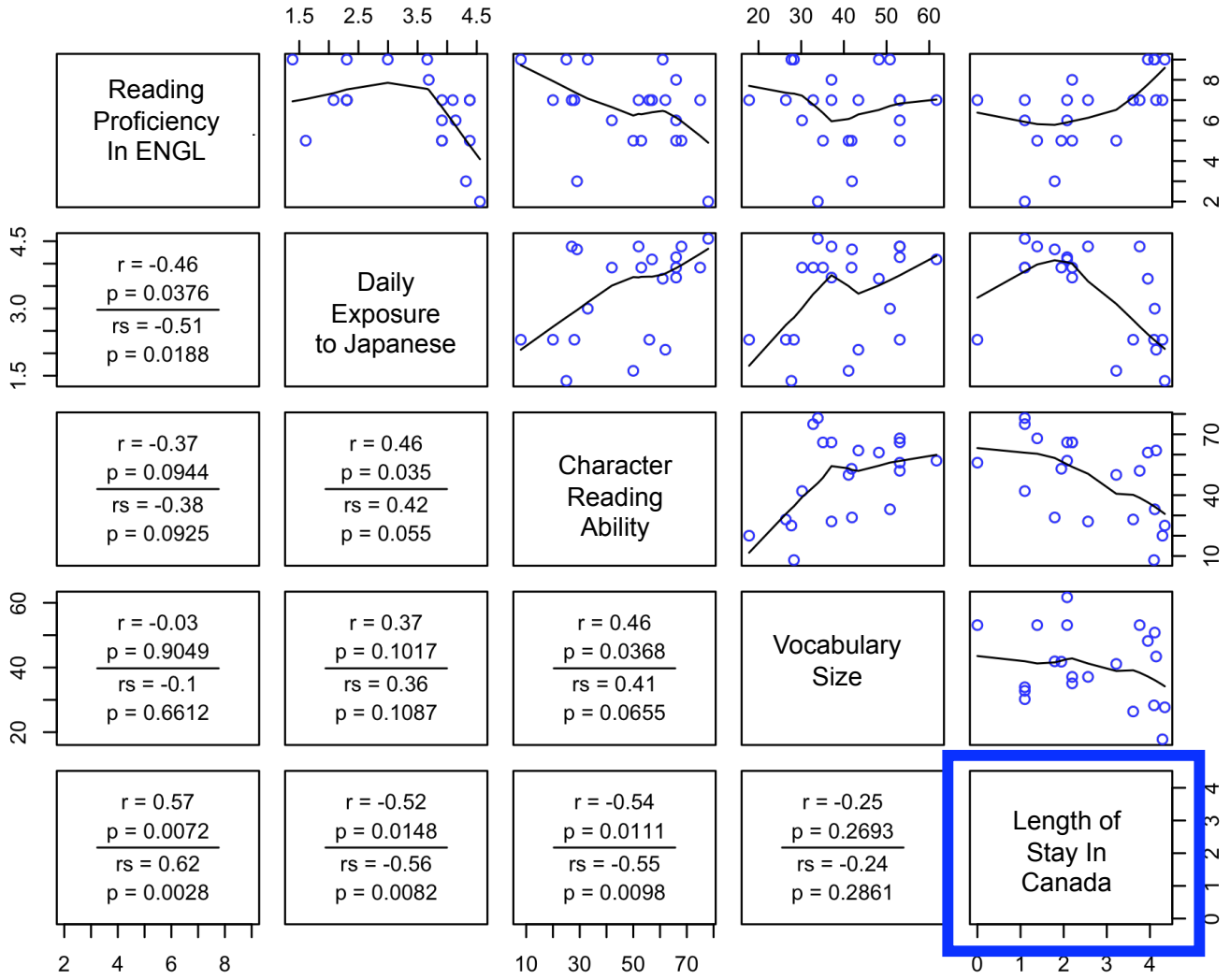
RT Analysis


sendai
setagaya
shizuoka
yokohama
narita
nerima
osaka
sakai
sapporo
kawasaki
kobe
kumamoto
nagano
nagoya
beppu
chiba
edogawa
hakodate
hakone
RTinv
Whole Word Freq
6
8
10
12
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0

RT Analysis


(a) - Task
RT (ms)
Previous RT
(b) - Whole Word x Feature
RT (ms)
Whole Word Freq
LeftKanjiStrokesResid
11.9
2.3
-0.5
-3.2
-8.3
(c) - Whole Word x Character
RT (ms)
Whole Word Freq
LeftKanjiTokenFreq
13.6
4.8
(d) - Character
RT (ms)
Left Kanji Neighbour
(e) - Character x Task
Incorrect
Correct
RT (ms)
RightKanjiTokenFreq
PreviousTrialCorrect
(f) - Character
RT (ms)
Right Kanji AoA (Resid)
(g) - Radical x Feature
RT (ms)
Right Kanji Radical Combinability
RightKanjiRadicalStrokesResid
-3.2
5.7
(h) - Individual x Task
RT (ms)
Length of Stay in Canada
Fixation
Right
Left
Central

What is “Length of stay in Canada” ?



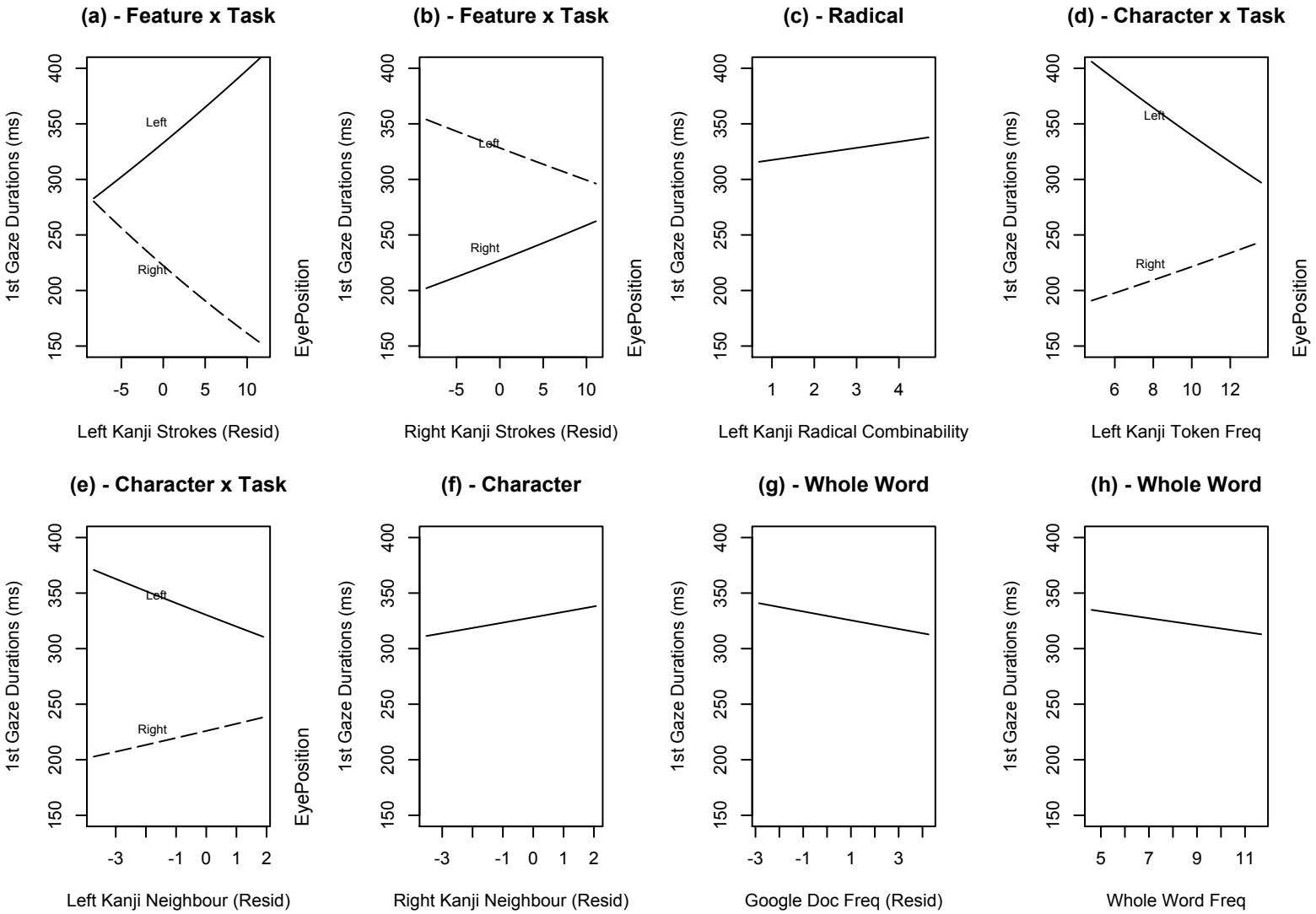
Interim Summary: RT Analysis

- Character frequency still exerts significant effect when AoA is considered together.
- Both the left character and the right character are important, and they are qualitatively different in their role
- The right character's semantic radical effect is manifested in RTs
- Semantic radicals may not be fully morphemic.

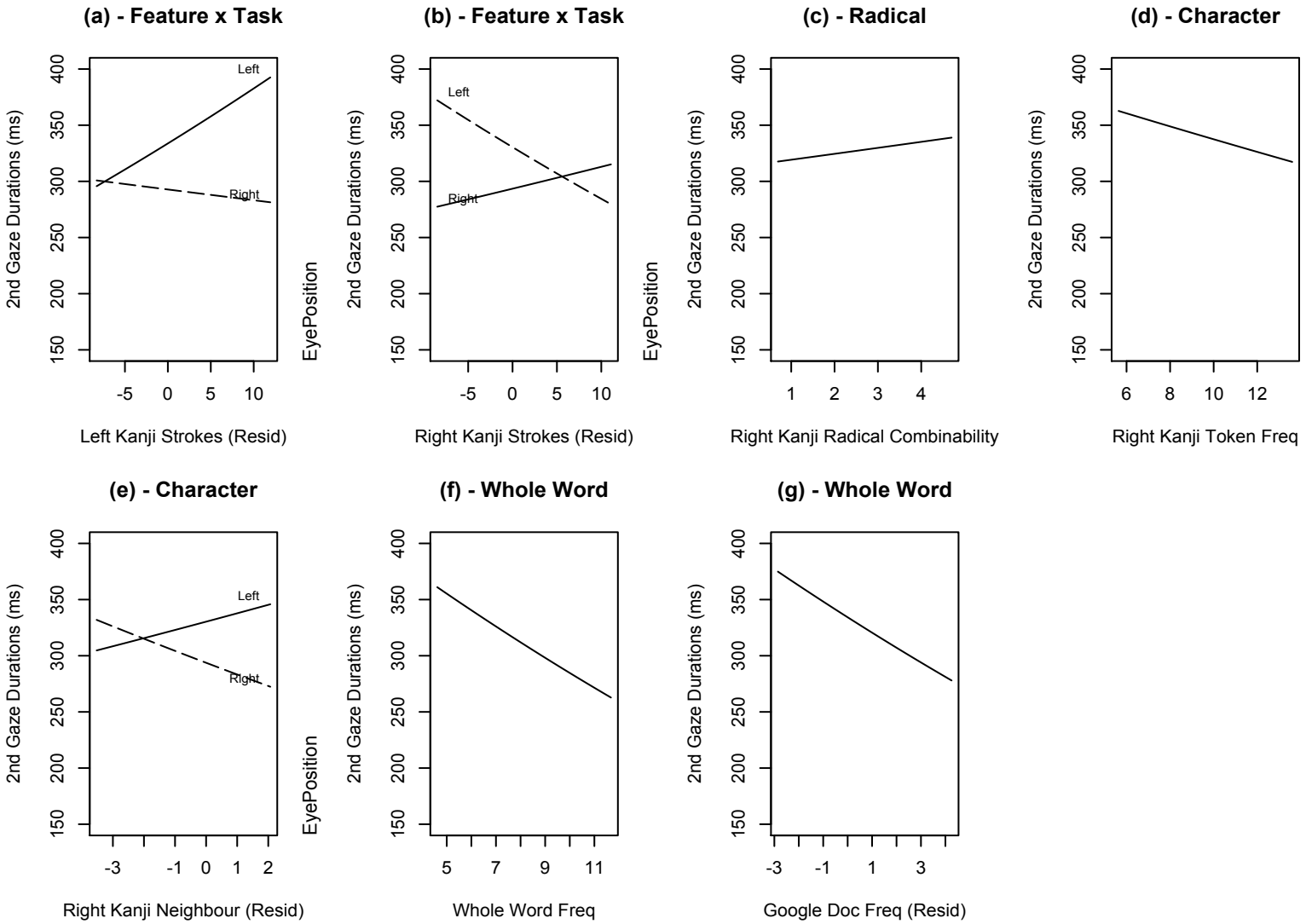
Eye-Movements during Lexical Decision

- Eyes try to optimize word recognition process by moving to optimal positions.
- Sentence reading vs. Isolated word reading

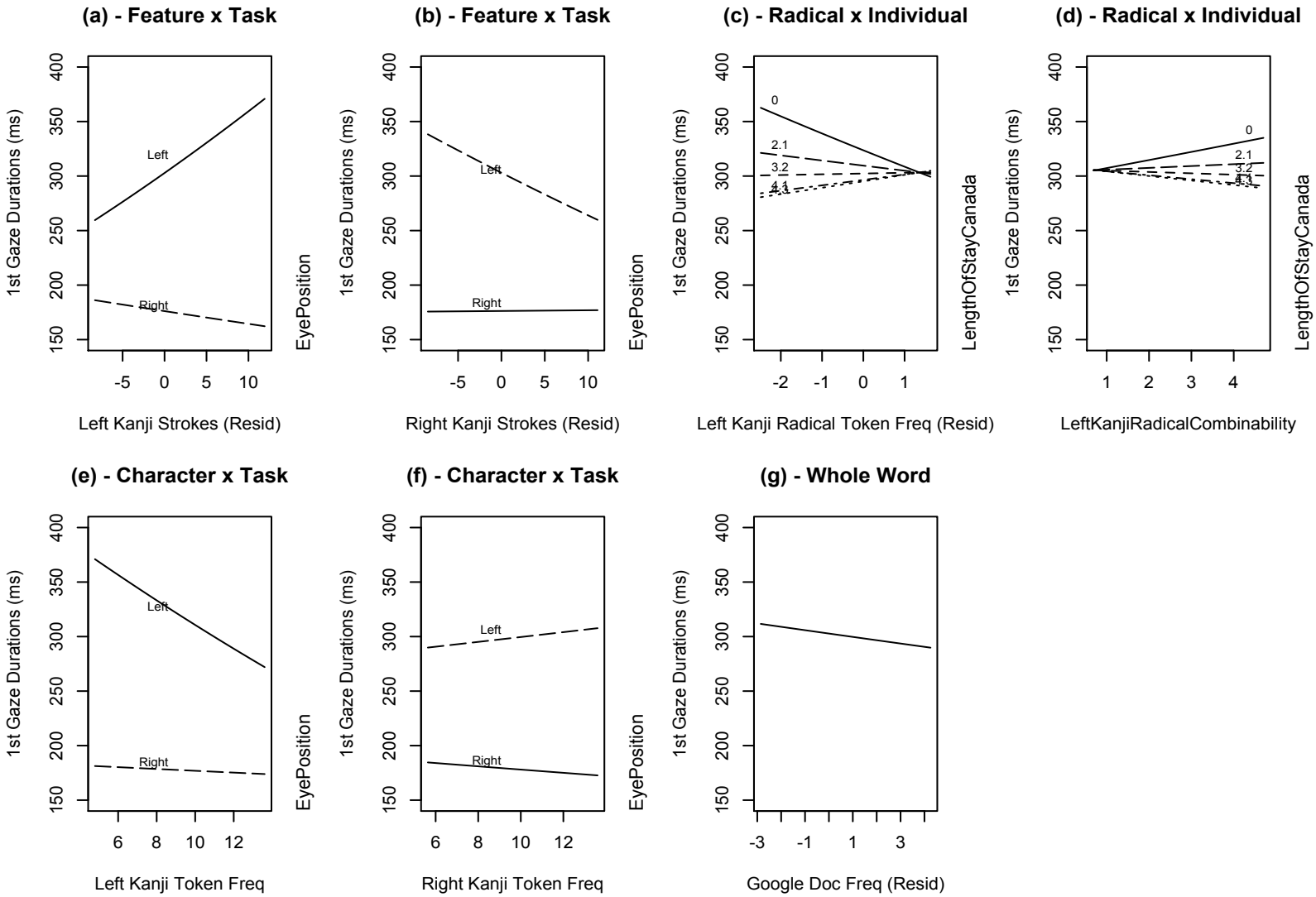
1st gaze duration for trials with 2 fixations



2nd gaze duration for trials with 2 fixations



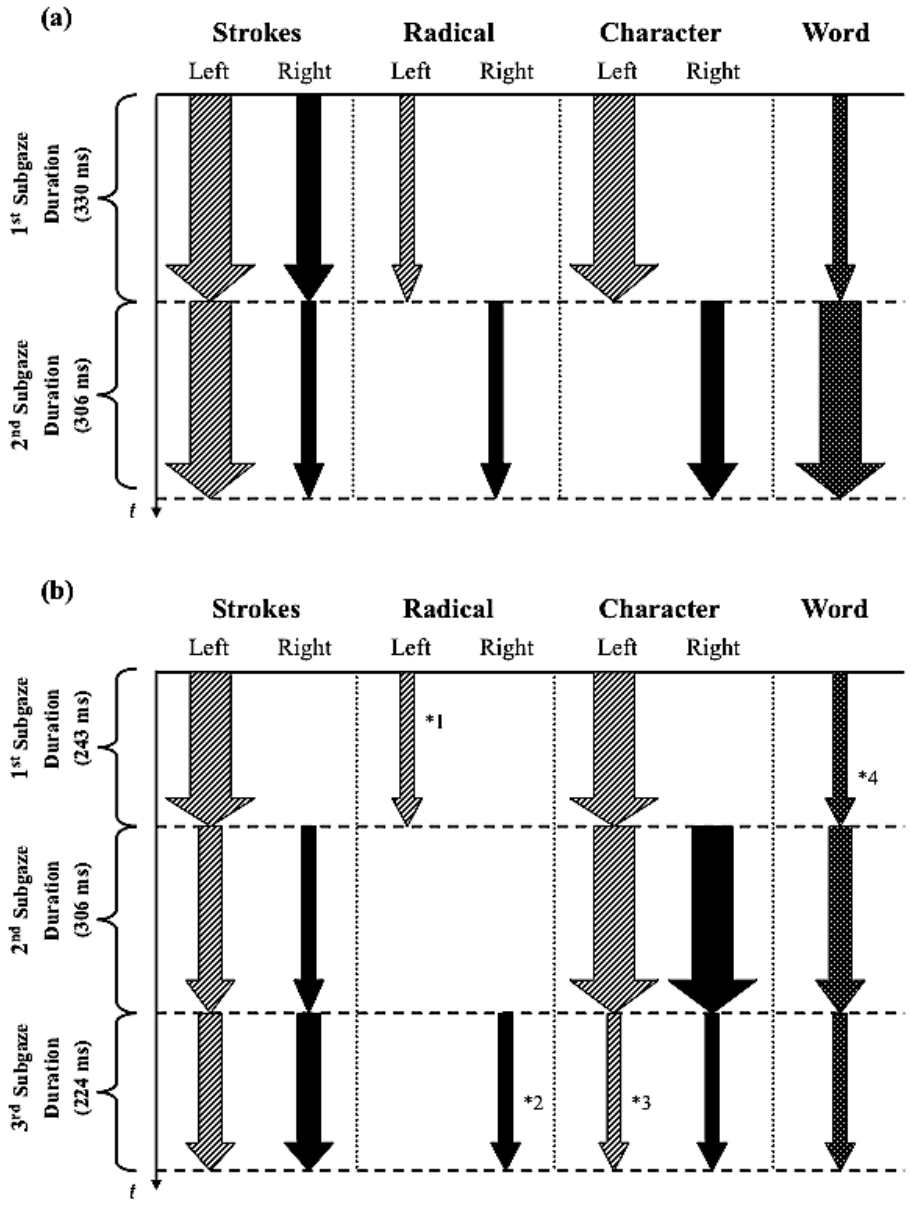
1st gaze duration for trials with 3 or more fixations



Summary

- All representations that can be activated can be activated (maximization of opportunity principle)
- Whole word frequency effect arises early before the whole word is scanned, but nature of this effect is yet to be investigated
- The relative importance of the left and the right constituents should not be just understood quantitatively. There is qualitative difference in their role (c.f. Gagne & Shoven, 1997, CARIN theory)
- Semantic radicals are processed but to a slight extent
- It is preferable to process from left to right
- What we are fixating ≠ what we are processing (the eye-mind assumption does not hold perfectly).
- Attending what is supposed to be processed at an appropriate point in time is, however, important.

Now, What is the Architecture !!



- The eye-tracking results do not fully support the current IA models of morphographic word recognition in Chinese and Japanese.

Now, What is the Architecture !!

- The radical-level is skipped, and radicals are activated through characters.
- Semantic radicals are activated to maximize opportunities, but they are not always useful
- Characters and semantic radicals have their own semantic representation.

- This IA architecture still needs to explain the early whole word frequency effect and resolve the issue of a feature template.

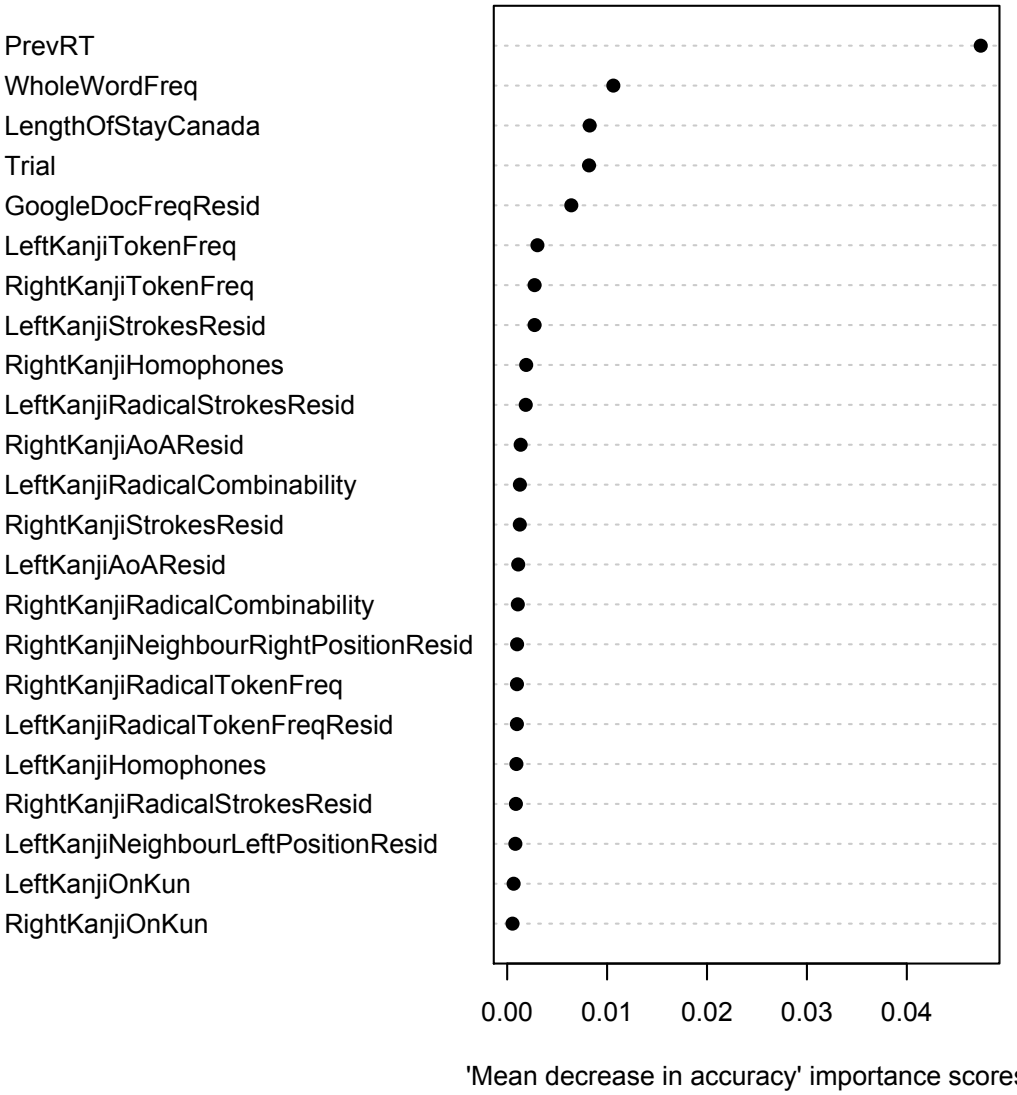
Issues that can make my life even more enjoyable in the future

- What happened during the 1st gaze duration ?
- What is qualitative nature of the early whole word frequency?

  (perceptual ? or conceptual ?)
- Role of other properties of compounds ? (e.g. transparency)
- Word / Character identification speed
- What about other tasks ? (e.g. naming, PDM)
- Chinese vs. Japanese ?
- What about other groups of individuals ?

…. and many more

Language Effects are Small ☹



Random forests

(Strobl et al, 2007, 2008)

Thank You

Correspondence

Koji Miwa Email: kmiwa@ualberta.ca

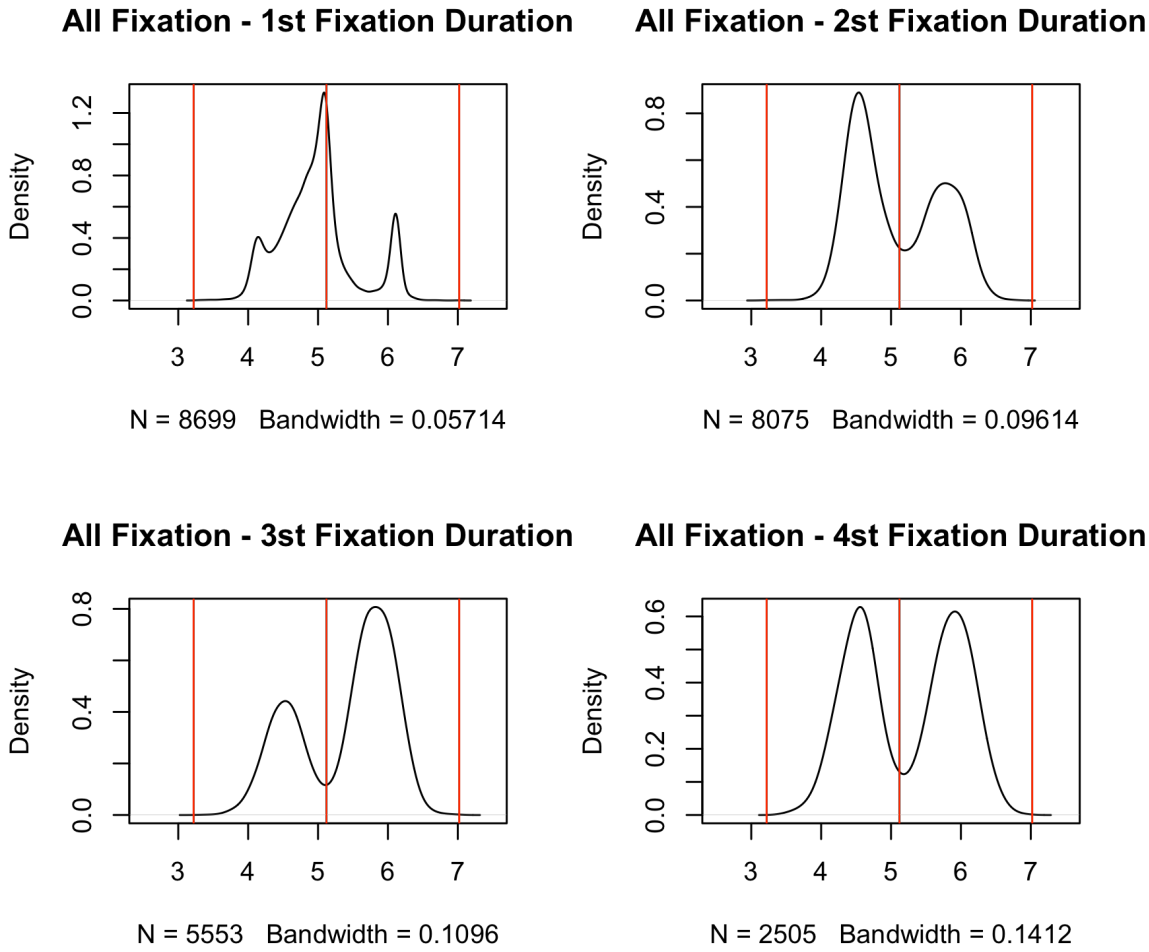
This research was supported by a Major Collaborative Research Initiative Grant (#412-2001-1009) from the Social Sciences and Humanities Research Council of Canada to Gary Libben (Director), Gonia Jarema, Eva Kehayia, Bruce Derwing, and Lori Buchanan (Co-investigators).

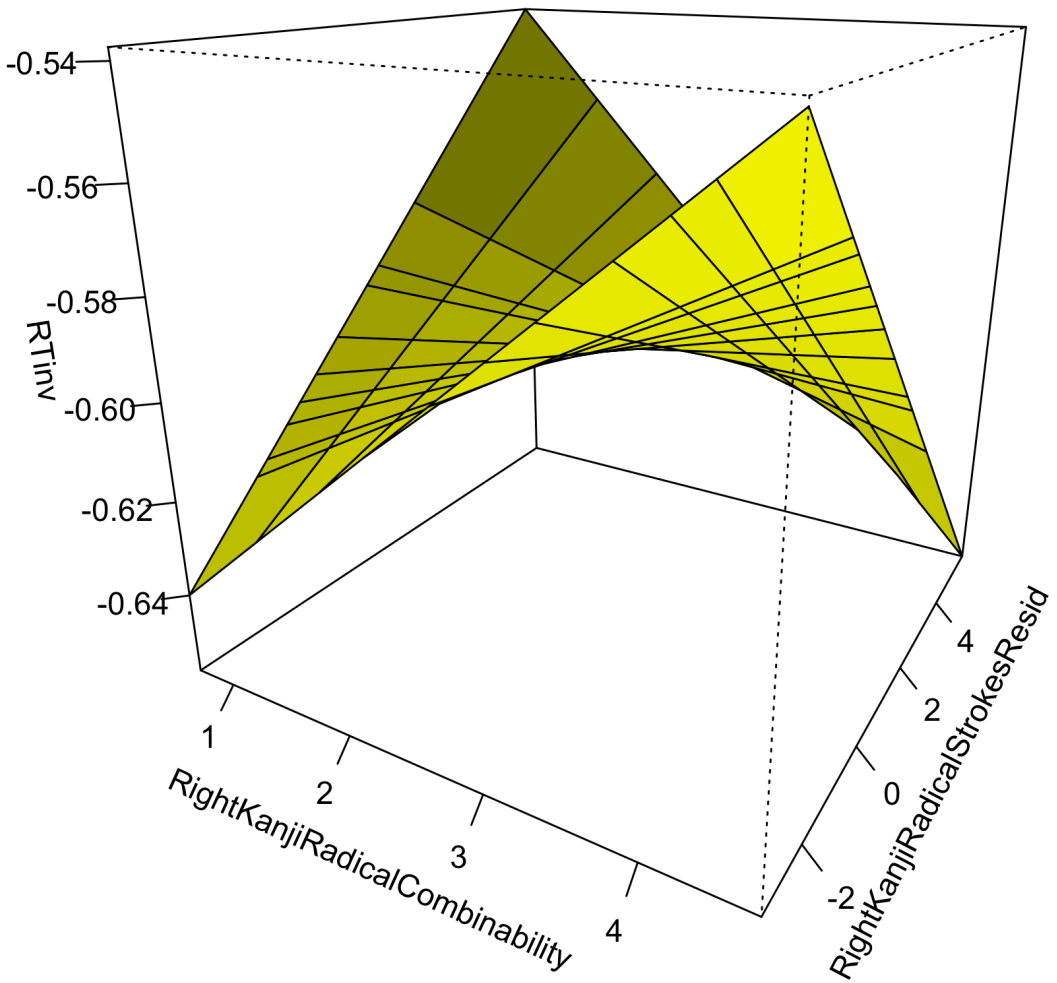
Social Sciences and Humanities Research Council of Canada

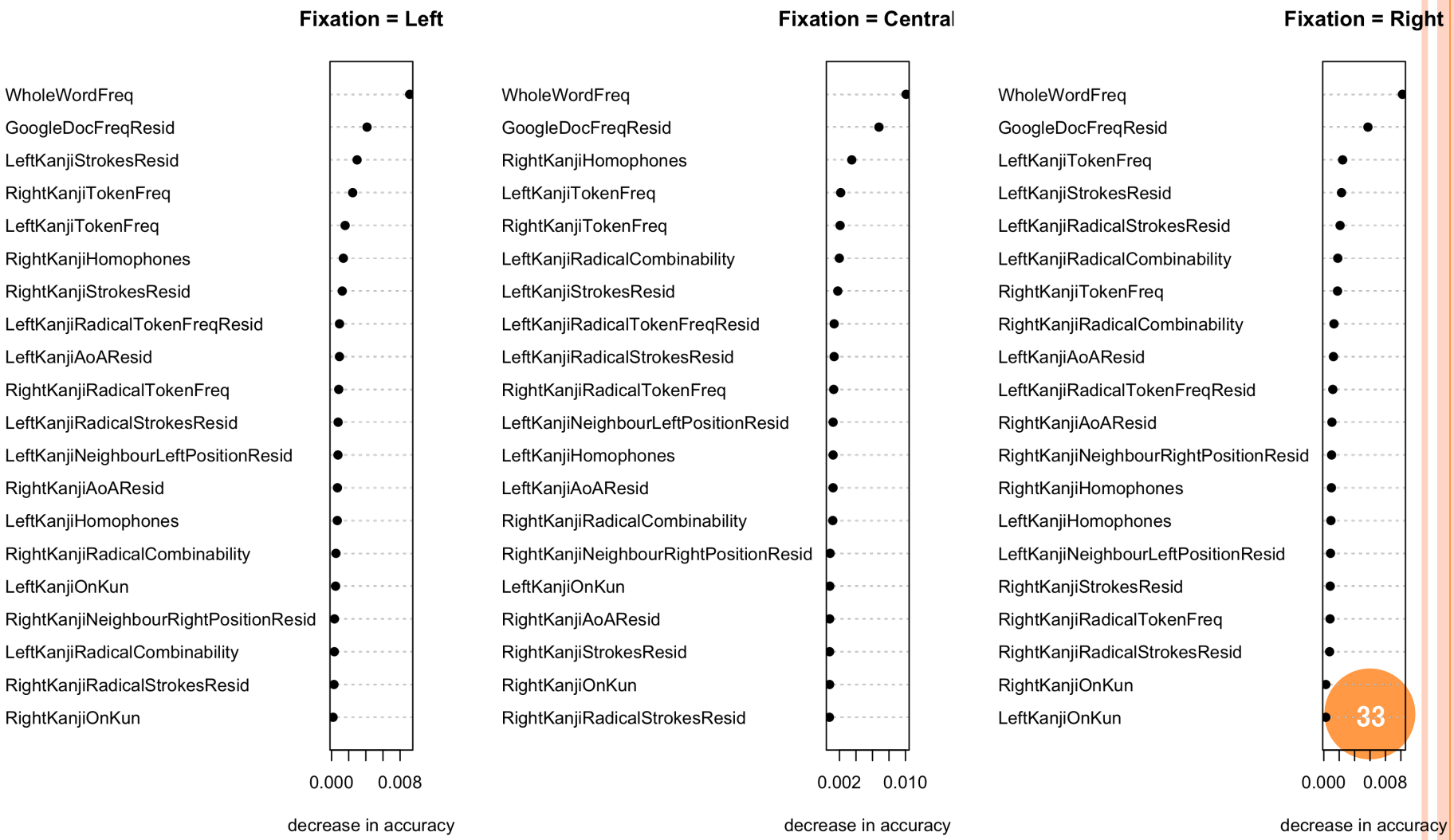
Conseil de recherches en sciences humaines du Canada

Canada

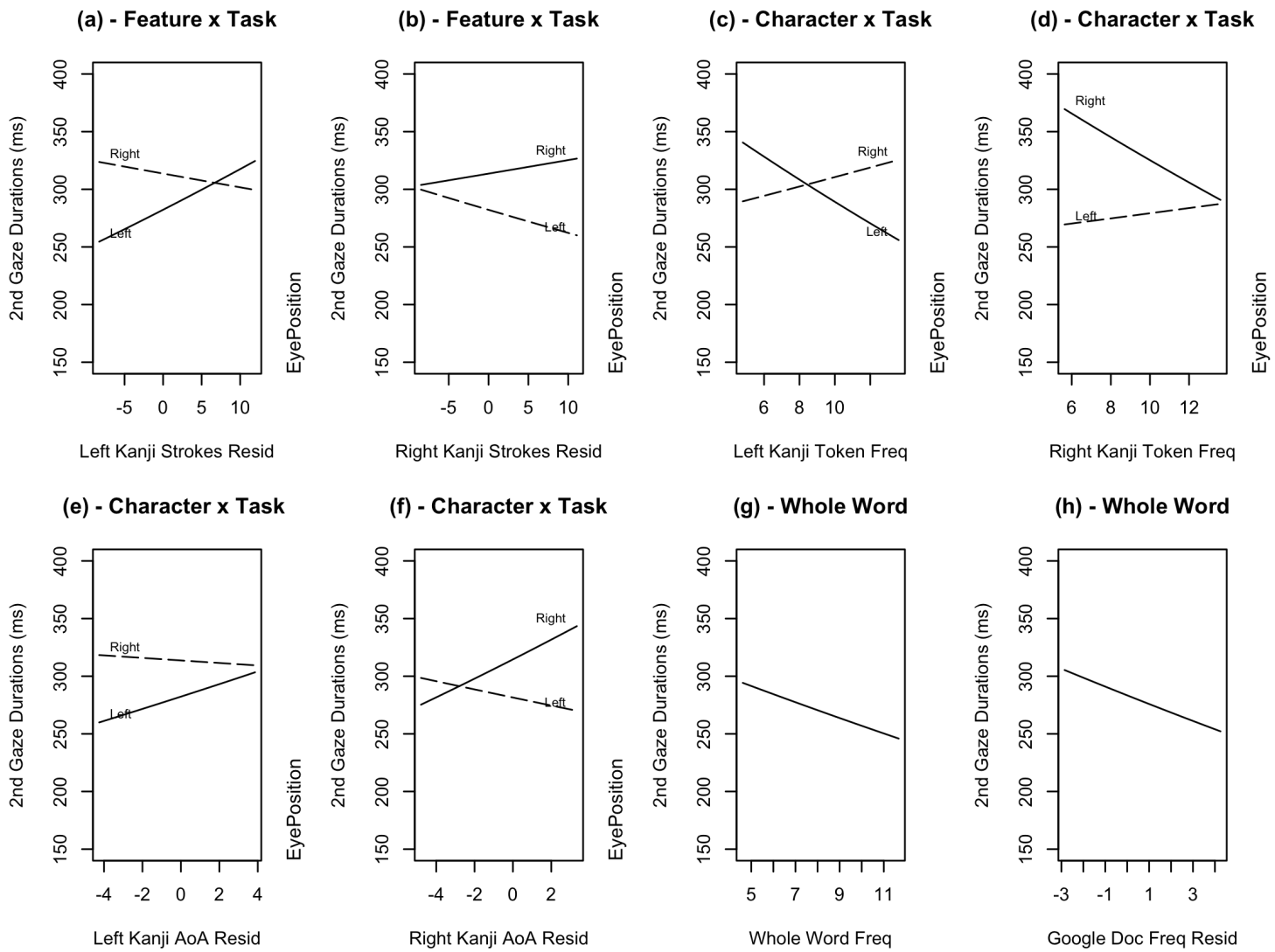
Density Plots for Different Fixation Positions in Sequence




-0.54
-0.56
-0.58
-0.60
-0.62
-0.64
RTinv
1
2
3
4
RightKanjiRadicalCombinability
-2
0
2
4
RightKanjiRadicalStrokesResid


Fixation = Left
WholeWordFreq
GoogleDocFreqResid
LeftKanjiStrokesResid
RightKanjiTokenFreq
LeftKanjiTokenFreq
RightKanjiHomophones
RightKanjiStrokesResid
LeftKanjiRadicalTokenFreqResid
LeftKanjiAoAResid
RightKanjiRadicalTokenFreq
LeftKanjiRadicalStrokesResid
LeftKanjiNeighbourLeftPositionResid
RightKanjiAoAResid
LeftKanjiHomophones
RightKanjiRadicalCombinability
LeftKanjiOnKun
RightKanjiNeighbourRightPositionResid
LeftKanjiRadicalCombinability
RightKanjiRadicalStrokesResid
RightKanjiOnKun
0.000
0.008
decrease in accuracy
Fixation = Central
WholeWordFreq
GoogleDocFreqResid
RightKanjiHomophones
LeftKanjiTokenFreq
RightKanjiTokenFreq
LeftKanjiRadicalCombinability
LeftKanjiStrokesResid
LeftKanjiRadicalTokenFreqResid
LeftKanjiRadicalStrokesResid
RightKanjiRadicalTokenFreq
LeftKanjiNeighbourLeftPositionResid
LeftKanjiHomophones
LeftKanjiAoAResid
RightKanjiRadicalCombinability
RightKanjiNeighbourRightPositionResid
LeftKanjiOnKun
RightKanjiAoAResid
RightKanjiStrokesResid
RightKanjiOnKun
RightKanjiRadicalStrokesResid
0.002
0.010
decrease in accuracy
Fixation = Right
WholeWordFreq
GoogleDocFreqResid
LeftKanjiTokenFreq
LeftKanjiStrokesResid
LeftKanjiRadicalStrokesResid
LeftKanjiRadicalCombinability
RightKanjiTokenFreq
RightKanjiRadicalCombinability
LeftKanjiAoAResid
LeftKanjiRadicalTokenFreqResid
RightKanjiAoAResid
RightKanjiNeighbourRightPositionResid
RightKanjiHomophones
LeftKanjiHomophones
LeftKanjiNeighbourLeftPositionResid
RightKanjiStrokesResid
RightKanjiRadicalTokenFreq
RightKanjiRadicalStrokesResid
RightKanjiOnKun
LeftKanjiOnKun
33
0.000
0.008
decrease in accuracy

2ND GAZE DURATION FOR TRIALS WITH 3 FIXATIONS



3RD GAZE DURATION FOR TRIALS WITH 3 FIXATIONS