Seeing ‘water’ in ‘desert’:

Semantic radical activation in visual Japanese compound recognition

Koji Miwa, Gary Libben & R. Harald Baayen

University of Alberta

Canadian Linguistics Association Conference

Vancouver, June 1st 2008

Semantic Radicals

- What is a semantic radical?

海 液 池 涙 漠 法

‘sea’ ‘liquid’ ‘pond’ ‘tear’ ‘broad’ ‘law’

 : *Sanzui-hen* (‘water’ radical)

- Semantic radicals are orthographic morphemes that have no phonological counterpart.

Research Question

Do these anomalous morphemes play a detectable role during reading of two-character Japanese compounds ?

Past Studies

- **Decompositional models**: Multi-level lexical representation

Stroke features

Radical

Character

Compound

‘sand - broad’ = ‘desert’

Past Studies

- Evidences for radical processing in visual word recognition
  - Flores d'Arcais & Saito (1993)
  - Flores d'Arcais, Saito, Kawakami (1995)
  - Feldman & Siok (1997)
  - Feldman & Siok (1999)
  - Saito, Yamazaki, & Masuda (2002)
  - Yencken & Baldwin (2006)

- Feldman & Siok (1997) and Feldman & Siok (1999)
  - Radicals with large families elicit shorter RTs

Issues to be Considered

The present study extends previous research using lexical decision by:

- Presenting two-character compounds
  - No isolated presentation of radicals
  - Most Japanese words are two-character compounds
  - Less likely to induce strategic radical decomposition
- Using better non-words
  - NOT illegal combinations of radicals
  - BUT illegal combinations of characters

* Use of illegal combinations of radicals as non-words may encourage participants to decompose characters in an unnatural way

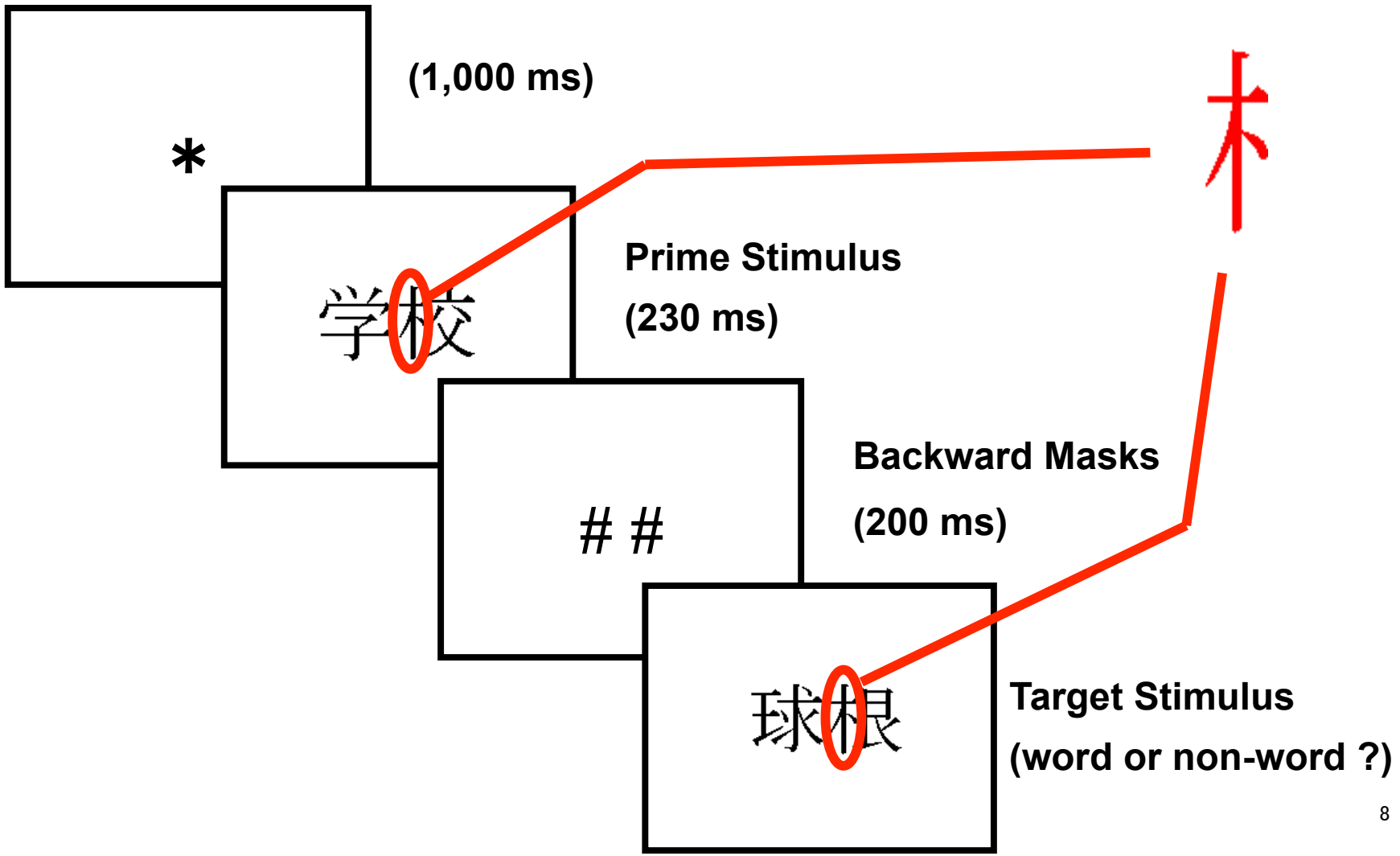
Experiment: LD with Priming

Original Hypothesis:

Repetition of semantic radicals induces a facilitatory priming effect.

- **Participants**
  - 30 native speakers of Japanese
- **Material**
  - 46 prime-target pairs of two-character compounds
  - 46 prime-nonword pairs

Experiment: LD with Priming



Interim Result

- Facilitatory priming effect (25 ms) was not significant ($p=0.09$).
- But, fortunately, we can make use of analyses of covariance using a wide range of lexical predictors.

Lexical Predictors

: *Sanzui-hen* (‘water’ radical)

- Radical type frequency: 103
- Radical token frequency: 503,833
- Radical stroke: 3
- Radical transparency: 5.14 (7-point ratings about how meaningful the radical is)

液 : *eki* ‘liquid’

- Character AoA: 5
- Character type frequency: 20
- Character token frequency: 22,792
- Character stroke 11

Lexical Predictors

Compound

Compound Token Freq

Character

(Log) Right/ Left Characters Type Freq

(Log) Right/ Left Characters Token Freq

(Log) Right/ Left Characters AoA

Radical

(Log) Right/ Left Characters Radical Type Freq

(Log) Right/ Left Characters Radical Token Freq

Right Character's Radical Transparency

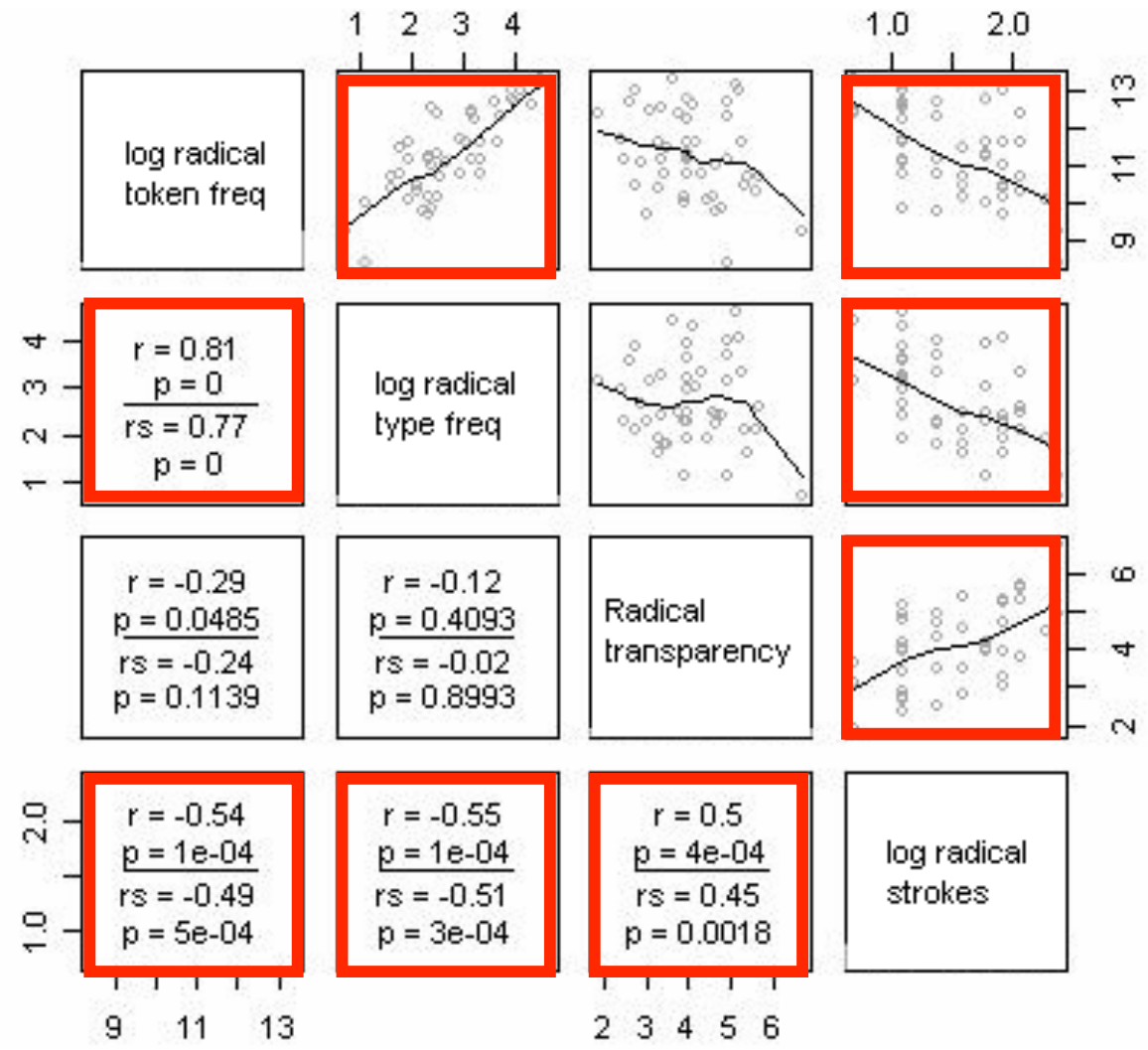
Stroke features

Right/ Left Characters Radical Stroke

Right/ Left Characters Stroke

Multi-collinearity

- Many of the 21 variables are highly correlated. (e.g. radical properties)



Principal Component Regression

1. We orthogonalized the predictors with principal component analysis (Uncorrelated new parameters were created).
2. We used the principal components as predictors.
3. We interpreted significant principal components by looking at the loadings of the original predictors.

Principal Component Regression

- In our principle component regression analysis, three principal components emerged as significant predictors (PC1, PC3, and PC5).

		Estimate	Std. Error	t value	p value
(Intercept)		6.59859	0.02960	222.9	0.000
PC1	inhibitory	0.03118	0.00862	3.62	0.000
PC3	facilitatory	-0.04017	0.01366	-2.94	0.003
PC5	facilitatory	-0.03333	0.01646	-2.03	0.043

Principal Component Regression: PC1

LogRadicalFreq	**-0.32595441**
LogRadicalTypeFreq	**-0.32461342**
LogTargetRightTypeFreq	**-0.25719829**
LogPrimeRightTokenFreq	**-0.24851236**
LogTargetRightTokenFreq	**-0.22849805**
LogPrimeWordFreq	**-0.20400900**
LogPrimeRightTypeFreq	-0.19793693
LogTargetLeftTokenFreq	-0.19694687
LogTargetLeftTypeFreq	-0.16765199
LogTargetWordFreq	-0.06674472
LeftRadicalStroke	-0.06558624
TargetLeftStroke	0.03508257
LogLeftRadicalFreq	0.05533828
RadicalTransparency	0.18117849
LogLeftRadicalTypeFreq	0.18936575
TargetRightStroke	**0.22199603**
TargetLeftAoA	**0.22712781**
PrimeRightAoA	**0.23350434**
RadicalStroke	**0.25461382**
PrimeRightStroke	**0.26371383**
TargetRightAoA	**0.29843572**

Type/ token frequencies of the right character, the right character's radical, and the prime.

Number of strokes and AoA of the right characters and the right character's radical

Principal Component Regression: PC3

LeftRadicalStroke	−0.31890508
LogTargetLeftTypeFreq	−0.27377317
TargetRightAoA	−0.27015008
LogPrimeRightTokenFreq	−0.21593121
RadicalTransparency	−0.19802870
LogPrimeRightTypeFreq	−0.15852085
LogPrimeWordFreq	−0.15197259
TargetRightStroke	−0.13537410
LogTargetLeftTokenFreq	−0.06331644
RadicalStroke	−0.05751754
TargetLeftStroke	−0.05340011
LogRadicalTypeFreq	−0.03122143
LogLeftRadicalFreq	0.05025043
LogRadicalFreq	0.09020049
LogLeftRadicalTypeFreq	0.12549673
PrimeRightStroke	0.12576490
PrimeRightAoA	0.24687072
LogTargetRightTypeFreq	0.25513908
TargetLeftAoA	0.26796378
LogTargetWordFreq	0.39901548
LogTargetRightTokenFreq	0.43306385

- A number of strokes of the left character's radical
- Type frequencies of the left character
- Right character's AoA

- Token frequencies of right characters
- Token frequencies of the compound

Principal Component Regression: PC5

LogLeftRadicalTypeFreq	−0.590242019
LogLeftRadicalFreq	−0.526935150
TargetLeftStroke	−0.377893898
LogRadicalTypeFreq	−0.151640066
LogPrimeRightTypeFreq	−0.149656442
TargetLeftAoA	−0.090099401
RadicalTransparency	−0.063713697
LogRadicalFreq	−0.061873990
LogPrimeRightTokenFreq	−0.054440805
LogPrimeWordFreq	−0.030480989
LeftRadicalStroke	−0.014648391
PrimeRightStroke	−0.004027343
LogTargetRightTypeFreq	0.028660268
TargetRightAoA	0.038024481
LogTargetRightTokenFreq	0.045912640
LogTargetWordFreq	0.078954418
RadicalStroke	0.080092633
LogTargetLeftTokenFreq	0.096903705
TargetRightStroke	0.152684837
LogTargetLeftTypeFreq	0.157344826
PrimeRightAoA	0.301901441

- Type/ token frequencies of the left character's radical
- Number of the left character's strokes

— AoA of the prime's right character

Influential Lexical Predictors

Compound

Target Word Freq

Character

Log Right Character Type/ Token Freq,

Left Character AoA

Log Left Character Type Freq, Right Character AoA

Radical

Log Right Character Radical Type/ Token Freq

Log Left Character Radical Type/ Token Freq

Stroke features

Left/ Right Character Radical Stroke

Left Character Stroke

What We Have Learned

砂漠

Compound

Target Word Freq

- ✓ Word frequency effect

  Compounds that occur more frequently are recognized faster.

What We Have Learned

Character

Log Right Character Type/ Token Freq,

Log Left Character Type Freq

- ✓ When a compound has a right character with high frequencies of occurrence or with many type neighbors, the compound is recognized faster.
- ✓ When a compound has a left character with many type neighbors, the compound is recognized more slowly.

=> Modifier – head competition for head status

What We Have Learned

Character

Left Character AoA, **Right Character AoA**

- ✓ When a compound's right character was learned early in school, the compound is recognized faster.
- ✓ When a compound's left character was learned early in school, the compound is recognized more slowly.

=> Modifier – head competition for head status

What We Have Learned

Radical

Log Right Character Radical Type/ Token Freq

Log Left Character Radical Type/ Token Freq

- ✓ When a compound's right character contains a semantic radical with high frequencies of occurrence or with many type neighbors, the compound is recognized faster.
- ✓ When a compound's left character contains a semantic radical with high frequencies of occurrence or with many type neighbors, the compound is recognized more slowly.

=> Modifier – head competition for head status

What We Have Learned

Stroke features

Left/ Right Character Radical Stroke

Left/ Right Character Stroke

✓ When characters of a compound are made of many strokes, the compound is recognized slowly.

=> Orthographic complexity leads to slower word recognition.

What We Have Learned

空X => 砂漠

Prime Properties

Prime Right Character Token Freq

Prime Right Character Stroke

✓ Presentation of a prime compound with highly frequent right character speeds up the recognition of a target compound.

✓ Presentation of a prime compound with a orthographically complex right character slows down the recognition of a target compound.

=> The primes were processed.

Conclusion

- The priming manipulation failed, but regression analyses show that semantic radicals are at work.
- Inclusion of distributional properties of words provides a better picture. When distributional data are not available, priming is a good option. Otherwise, regression offers a more powerful methodology.

Thank You

Correspondence

Koji Miwa: kmiwa@ualberta.ca

Acknowledgment

I would like to express my special thanks to members of the Centre for Comparative Psycholinguistics for their feedbacks on the earlier version of this presentation and for their casual advices and insights.

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada

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).