

音韻対立点の左側環境による
検索範囲の縮小を考慮した
機能負担量の指標

北原 真冬

早稲田大学

音韻論学会@法政大学

2008-06-20

はじめに

- 対立 (opposition): 意味を区別する音素の違い。
- 最小対 (minimal pair): ひとつだけ音素が違う単語の対
 - /t/-/d/: 「退学」 - 「大学」, 「書いた」 - 「嗅いだ」
 - /p/-/k/: 「パン」 - 「缶」, 「きっぱり」 - 「きっかり」
- Kitahara(2008): アクセントの違いを無視 → 今回

可能な最小対の総数

m : 最大単語長

n : 音素数

C : ある音素体系における最小対の総数

$$C = \sum_{i=1}^m \frac{mn^m(n-1)}{2}$$

$$(m, n) = (1, 3); C = 3$$

$$(m, n) = (3, 3); C = 111$$

$$(m, n) = (10, 20); C = 2,037,221,052,631,580 \text{ (2037兆...)}$$

問題

なぜ音韻体系は膨大な可能性の『一部』しか利用しないのか？

構造主義音韻論

| | | | | |
|---|---|---|---|-----|
| p | — | t | — | k |
| | | | | |
| b | — | d | — | () |

- 音素体系は均衡と整合性を持つ
- 空き間のある音素体系 → 埋めるような構造的圧力 (Hockett, 1966)

機能負担量 (Trubetzkoy, 1939; Martinet, 1955)

- /t/-/d/ ≫ /p/-/b/
- 多くの単語の弁別に関わる → 高い機能負担量
- 機能負担量の低い対立をつぶすような構造的圧力 (Hockett, 1967)

日本語における機能負担量(1)

近傍 (neighborhood) ある単語から音素を1つだけ削除/置換/添加

base 「言語/gengo/」

削除 「援護/engo/」

置換 「弁護/bengo/」

添加 「元号/gengoo/」

共通群 (cohort) ある単語と初頭部分を共通にする単語群

「雨/ame/」

/a/-cohort: 「朝/asa/」, 「青/ao/」, 「汗/ase/」...

/am/-cohort: 「海女/ama/」, 「網/ami/」...

日本語における機能負担量(2)

- 近傍の計算 \approx 最小対の数え上げ
- しかし「対立はどこでも平等」か?
- 共通群によって検索範囲は絞られる

対立点の左側音素列とその生起数(抜粋)

| 最小対 | 左側環境 | 左側環境を共有する単語数 |
|-------------------|-------|--------------|
| toriQpu - toriQku | toriQ | 2 |
| toraQpu - toraQku | toraQ | 3 |
| : | : | : |
| supaato - sukaato | su | 1611 |
| ope - oke | o | 3155 |

日本語における機能負担量 (3)

対立点の左側環境による検索範囲の絞りこみを考慮

機能負担量の指標

$$f(p) = \sum_c \sum_s \frac{s(c)}{L}$$

- $f(p)$: 音素 p の機能負担量の指標
 c : p を含む対立
 $s(c)$: c における左側音素列の共有数
 L : レキシコンの総項目数

対立 /p/-/k/, $L = 88574 (F \geq 1)$

| 左側環境 | 左側共有数 $s(c)$ |
|-------|--------------|
| toriQ | 2 |
| toraQ | 3 |
| ⋮ | ⋮ |
| su | 1611 |
| o | 3155 |

$$\begin{aligned}
 & \sum_s \frac{s(c)}{L} \\
 &= \frac{2}{88574} + \frac{3}{88574} \cdots \\
 &= 6.0738
 \end{aligned}$$

対立 /t/-/k/: 61.2414

対立 /b/-/k/: 41.6202

$$f(/k/) = \sum_c$$

$$\begin{aligned}
 &= 6.0738 + 61.2411 + \cdots \\
 &= 1728.9011
 \end{aligned}$$

元データ

仮名:音素変換

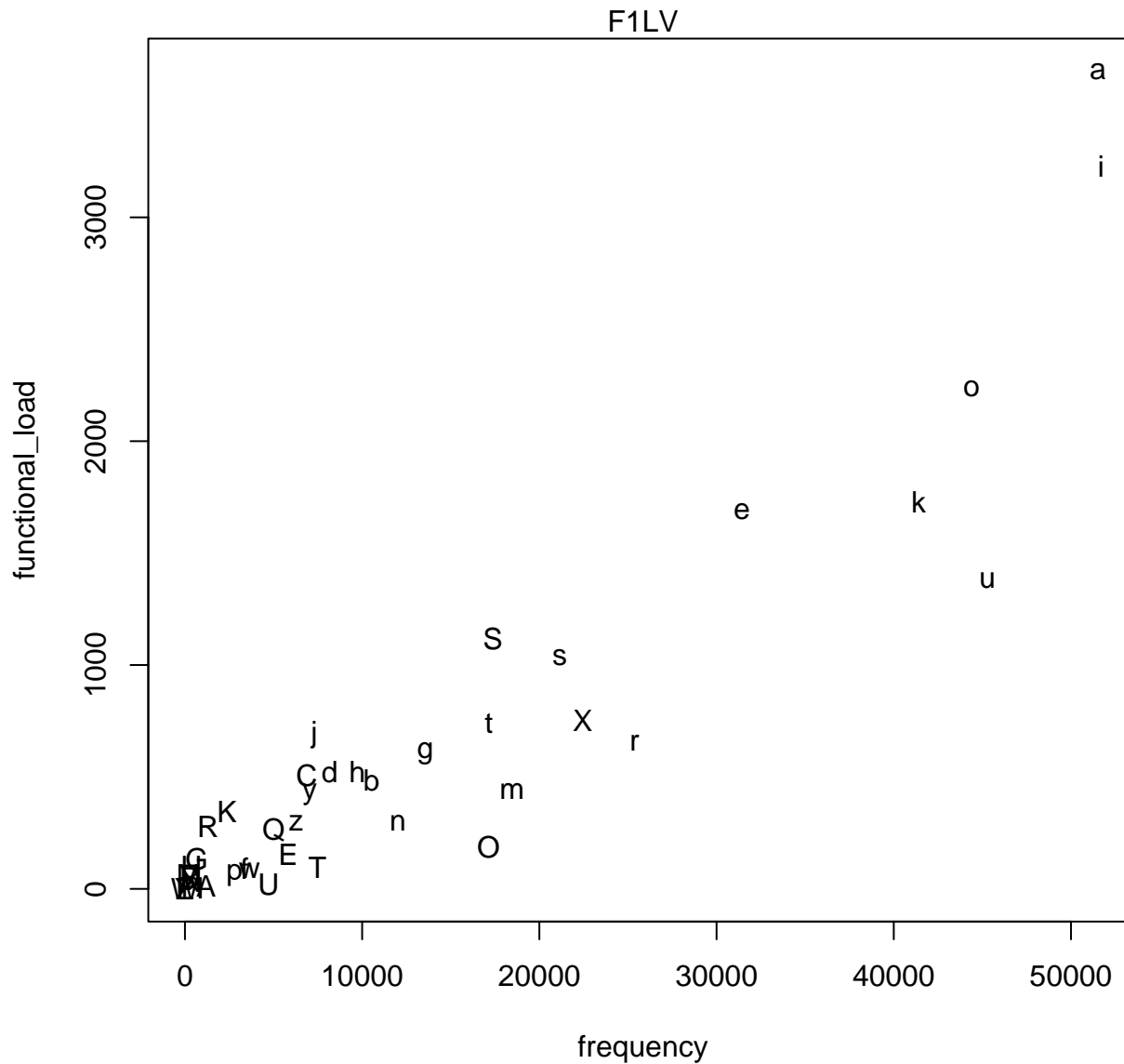
| | | | | | | | |
|----|----|----|----|----|----|----|----|
| ア | イ | ウ | エ | オ | ヤ | ユ | ヨ |
| a | i | u | e | o | ya | yu | yo |
| カ | キ | ク | ケ | コ | キャ | キュ | キョ |
| ka | ki | ku | ke | ko | Ka | Ku | Ko |
| ガ | ギ | グ | ゲ | ゴ | ギャ | ギユ | ギョ |
| ga | gi | gu | ge | go | Ga | Gu | Go |
| サ | スイ | ス | セ | ソ | | スユ | |
| sa | si | su | se | so | | Yu | |
| シャ | シ | シュ | シェ | シヨ | | | |
| Sa | Si | Su | Se | So | | | |
| ザ | ズイ | ズ | ゼ | ゾ | | | |
| za | zi | zu | ze | zo | | | |
| ジャ | ジ | ジュ | ジェ | ジョ | | | |
| ja | ji | ju | je | jo | | | |
| タ | ティ | トゥ | テ | ト | | テユ | |
| ta | ti | tu | te | to | | Wu | |
| チャ | チ | チュ | チェ | チョ | | | |
| Ca | Ci | Cu | Ce | Co | | | |
| ツア | ツイ | ツ | ツエ | ツオ | | | |
| Ta | Ti | Tu | Te | To | | | |
| ダ | ディ | ドゥ | デ | ド | | デュ | |
| da | di | du | de | do | | Zu | |
| ナ | ニ | ヌ | ネ | ノ | ニヤ | ニユ | ニョ |
| na | ni | nu | ne | no | Na | Nu | No |
| ハ | ヒ | | ヘ | ホ | ヒヤ | ヒユ | ヒョ |
| ha | hi | | he | ho | Ha | Hu | Ho |
| ファ | フィ | フ | フェ | フォ | | フユ | |
| fa | fi | fu | fe | fo | | Fu | |
| バ | ビ | ブ | ベ | ボ | ビヤ | ビユ | ビョ |
| ba | bi | bu | be | bo | Ba | Bu | Bo |
| パ | ピ | プ | ペ | ポ | ピヤ | ピユ | ピョ |
| pa | pi | pu | pe | po | Pa | Pu | Po |
| マ | ミ | ム | メ | モ | ミヤ | ミュ | ミョ |
| ma | mi | mu | me | mo | Ma | Mu | Mo |
| ラ | リ | ル | レ | ロ | リヤ | リュ | リョ |
| ra | ri | ru | re | ro | Ra | Ru | Ro |
| ワ | ウィ | | ウエ | ウォ | ン | ッ | |
| wa | wi | | we | wo | N | Q | |

- 天野・近藤(1999): 8万8千語の発音データ+親密度
- 同一モーラ数・同一アクセントの集合内でのみ計算
- 拗音と直音を別扱い:
 $c^y \rightarrow C$
- 長音の後部を別扱い:
 $CVV \rightarrow CV + V$
- 親密度別に分類

親密度と長さによる単語数分布

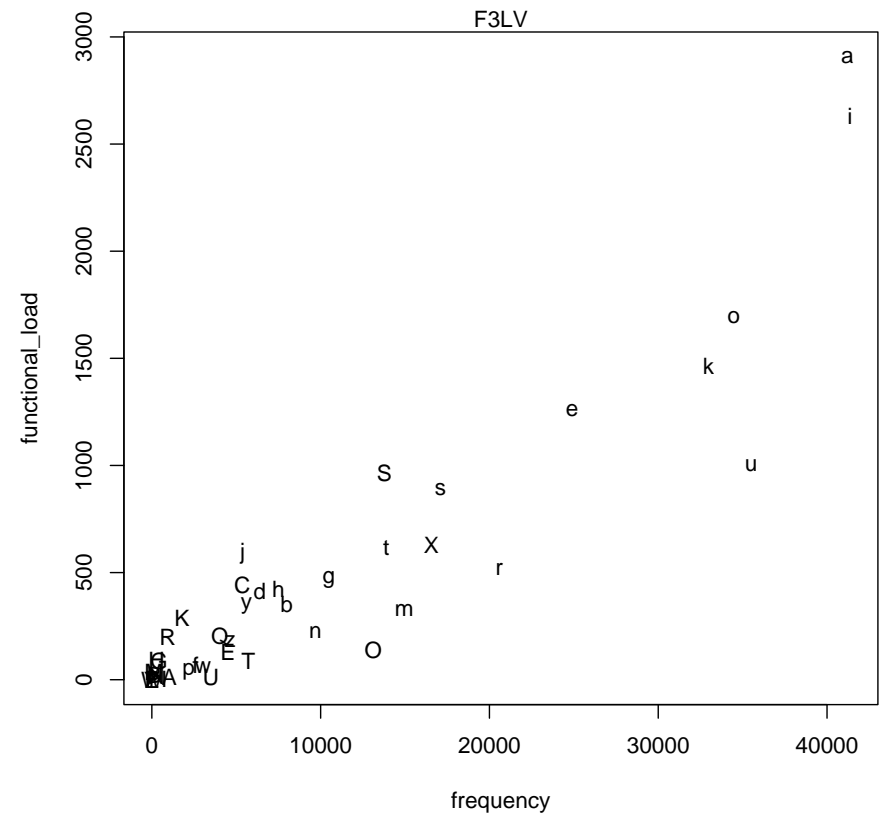
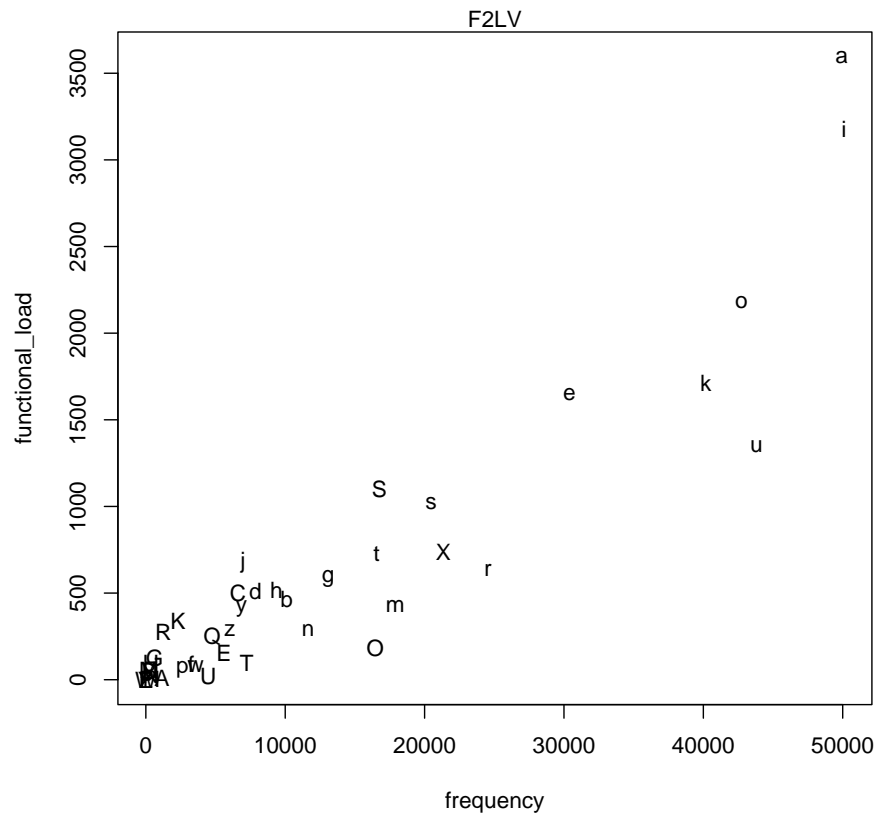
| | モーラ数 | | | | | | | | 計 |
|-----|------|------|-------|-------|-------|------|------|------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8+ | |
| F>1 | 346 | 4870 | 20878 | 37225 | 13731 | 7296 | 2620 | 1609 | 88574 |
| F>2 | 345 | 4772 | 20148 | 36178 | 13318 | 6983 | 2419 | 1433 | 85595 |
| F>3 | 344 | 4279 | 16252 | 29070 | 11261 | 5837 | 1896 | 1106 | 70044 |
| F>4 | 310 | 3603 | 12866 | 22892 | 8937 | 4585 | 1455 | 791 | 55438 |
| F>5 | 173 | 2304 | 8390 | 14698 | 5698 | 2870 | 802 | 459 | 35393 |
| F>6 | 16 | 477 | 1372 | 1817 | 486 | 210 | 62 | 49 | 4488 |

分析(1)

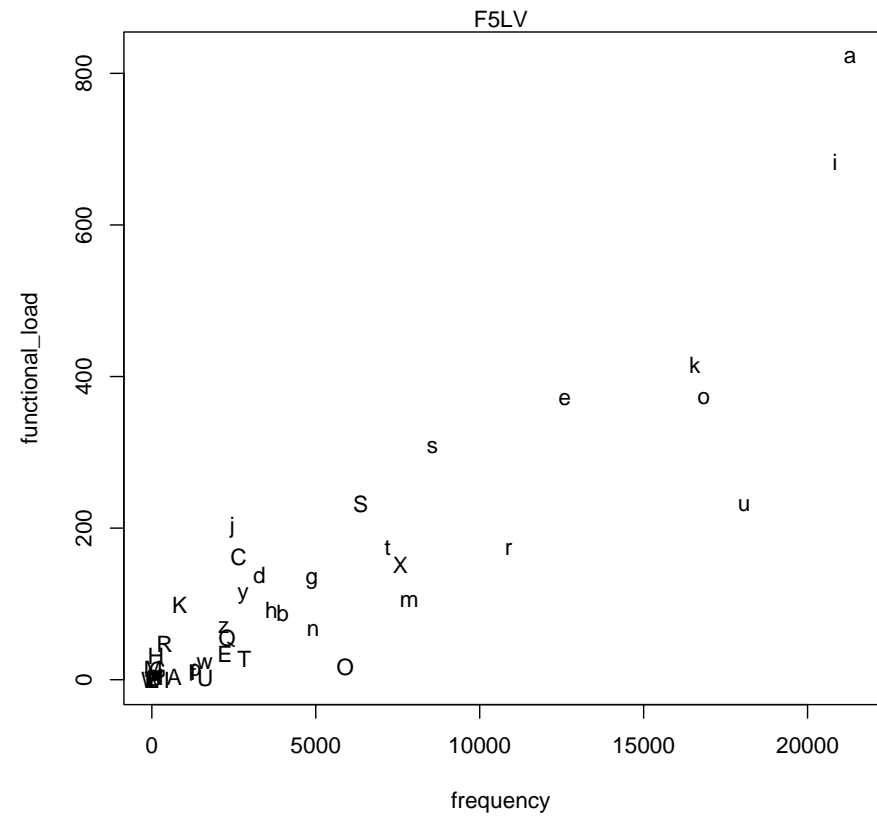
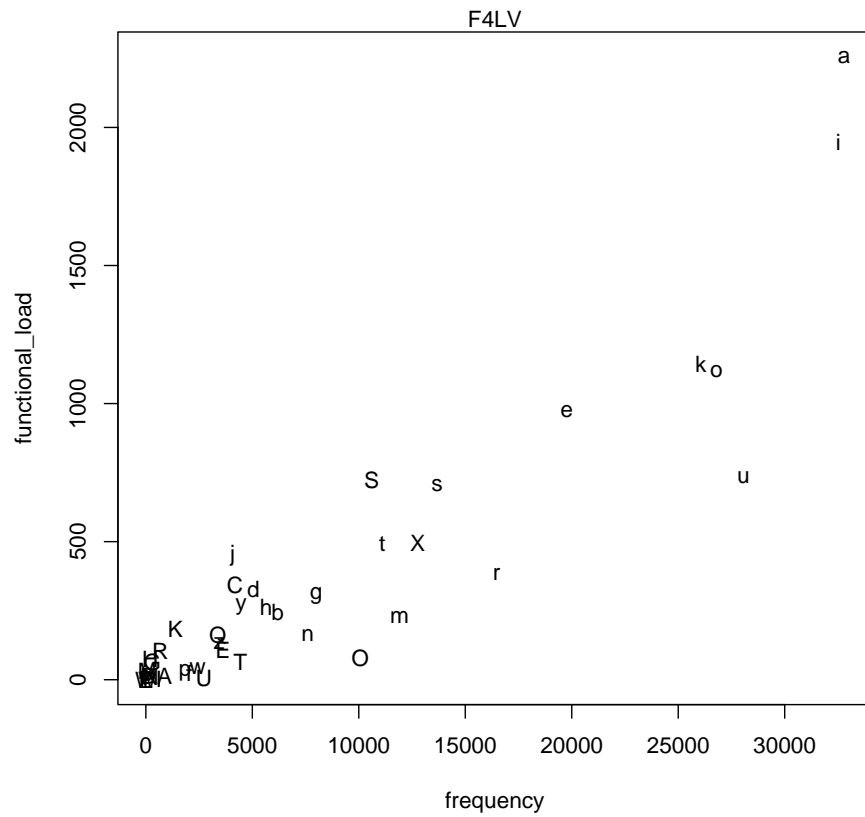


- x軸: 音素頻度
- y軸: 機能負担量
- 右下寄り: 頻度が高いのに弁別にあまり役立たない
- 左上寄り: 頻度が低いのに弁別に役立つ

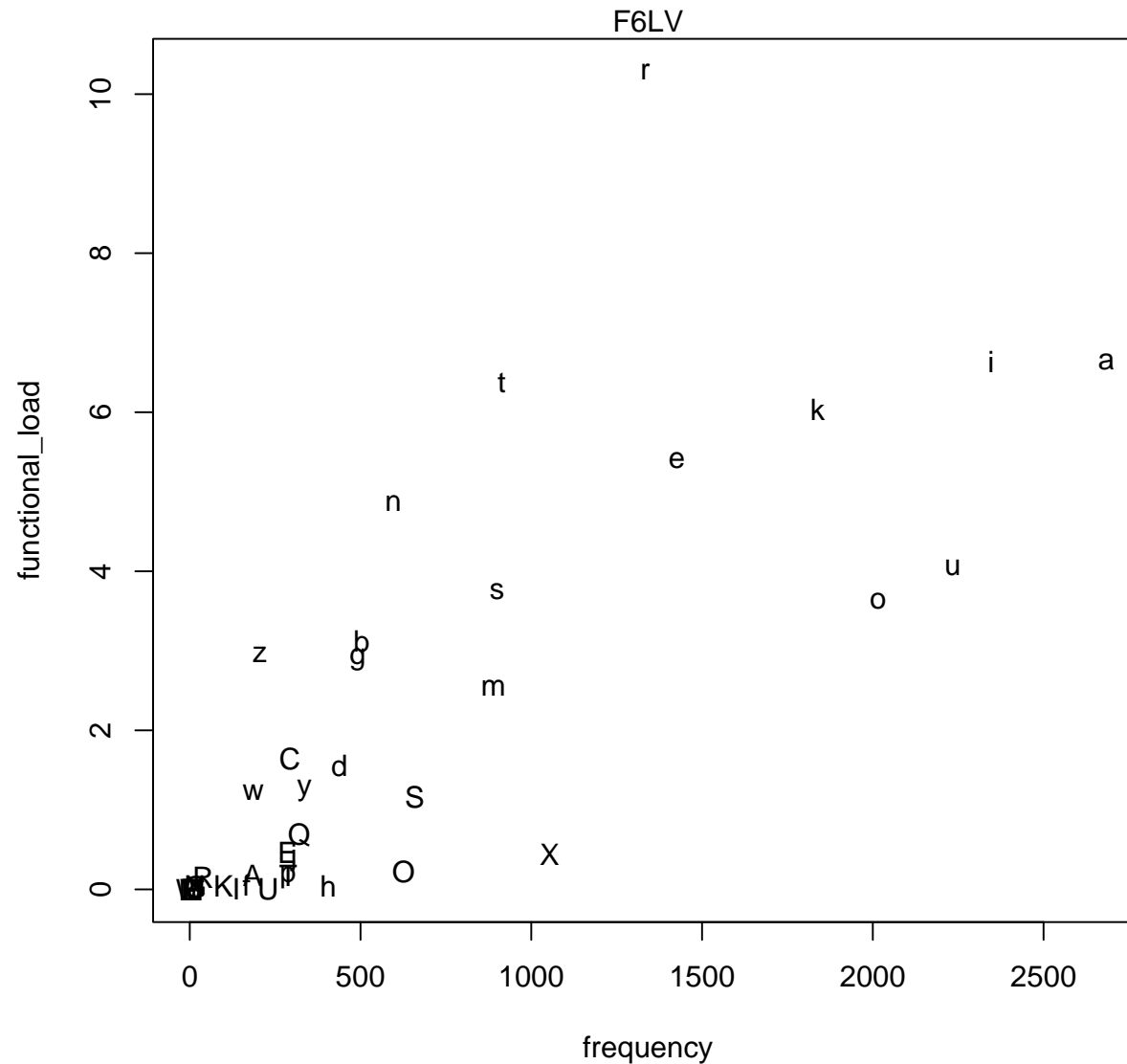
分析(2): 親密度2,3以上



分析(3): 親密度4,5以上

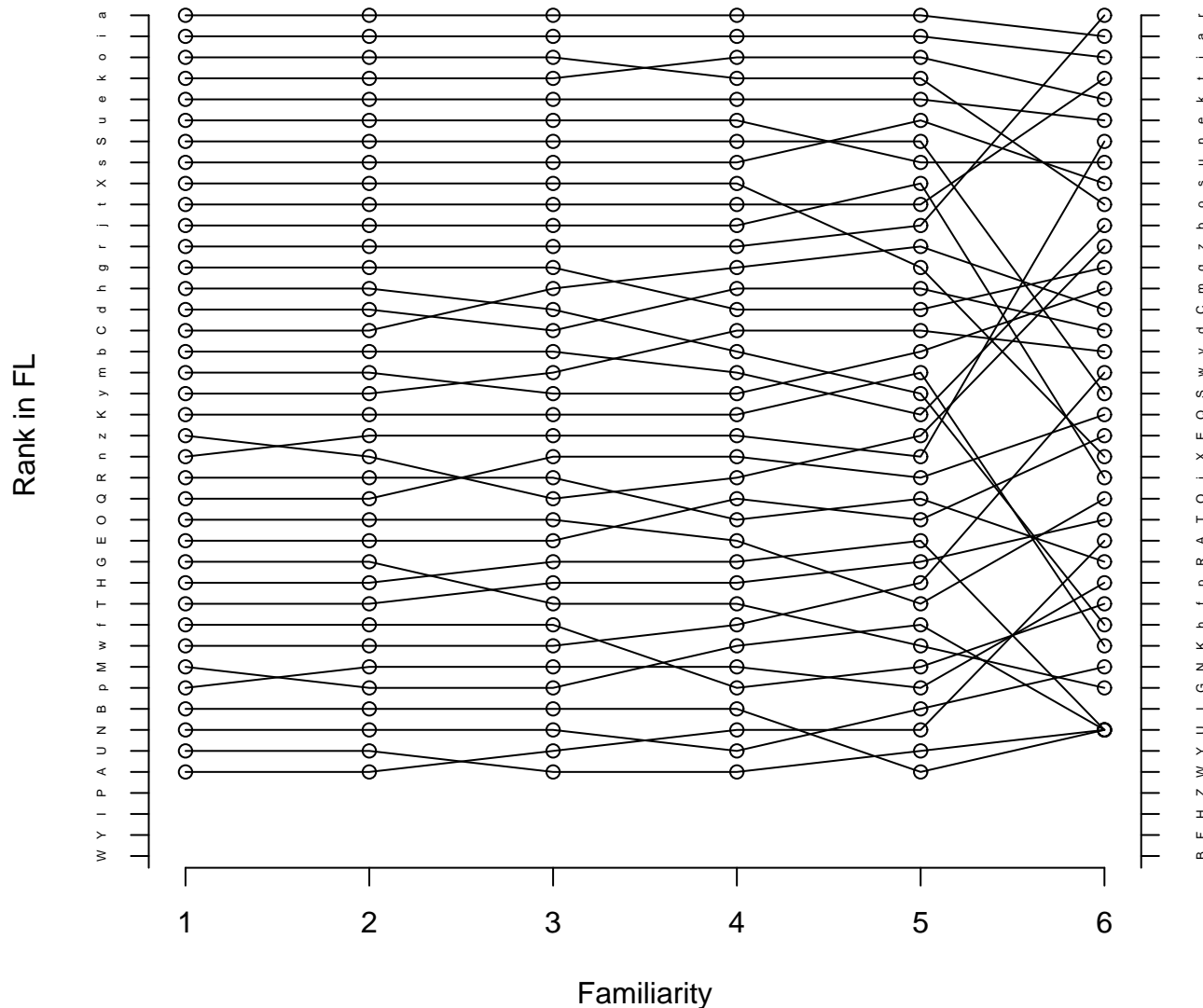


分析(4): 親密度6以上、全体の傾向



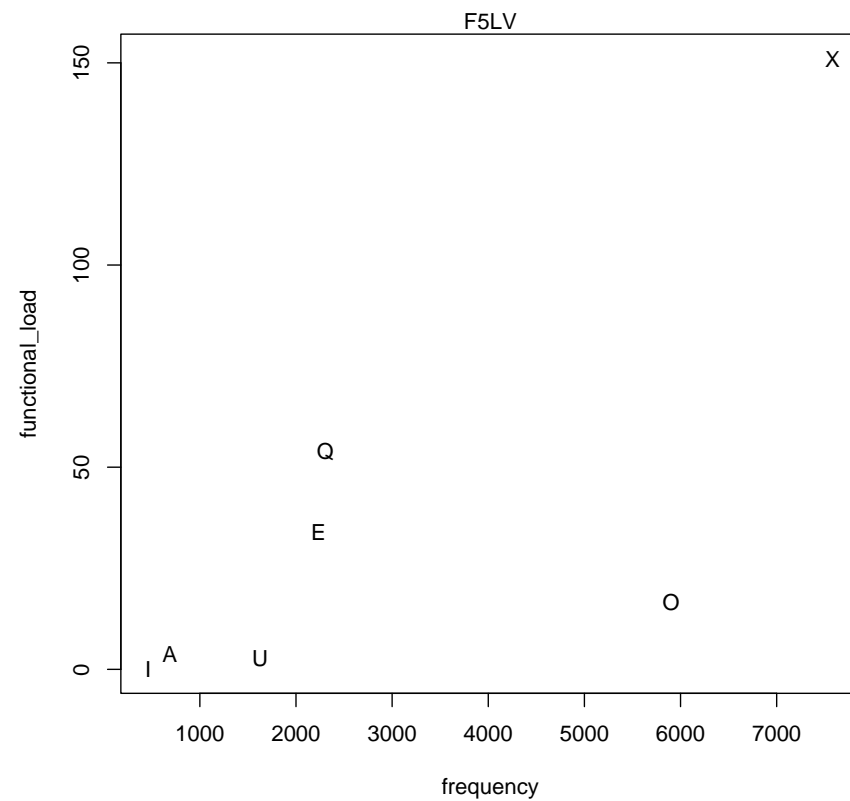
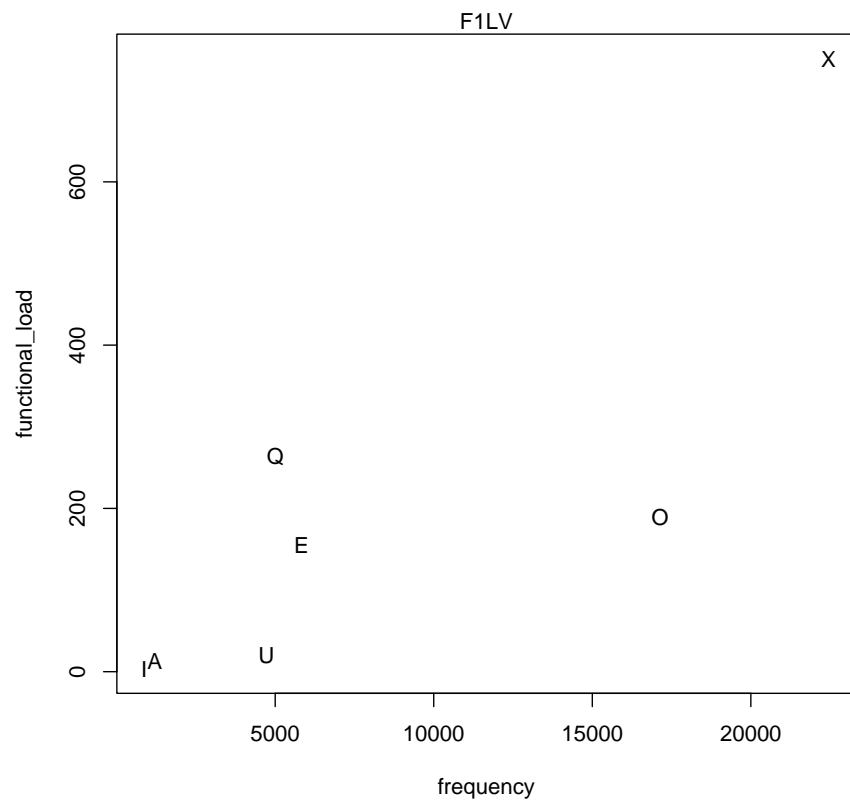
- 全体的に母音が上位
- 子音では/k/が飛び抜けている
- 頻度が高くても機能負担量が高いとは限らない
- 親密度4-5, 5-6の間に変化

分析(5): 機能負担量順位の遷移



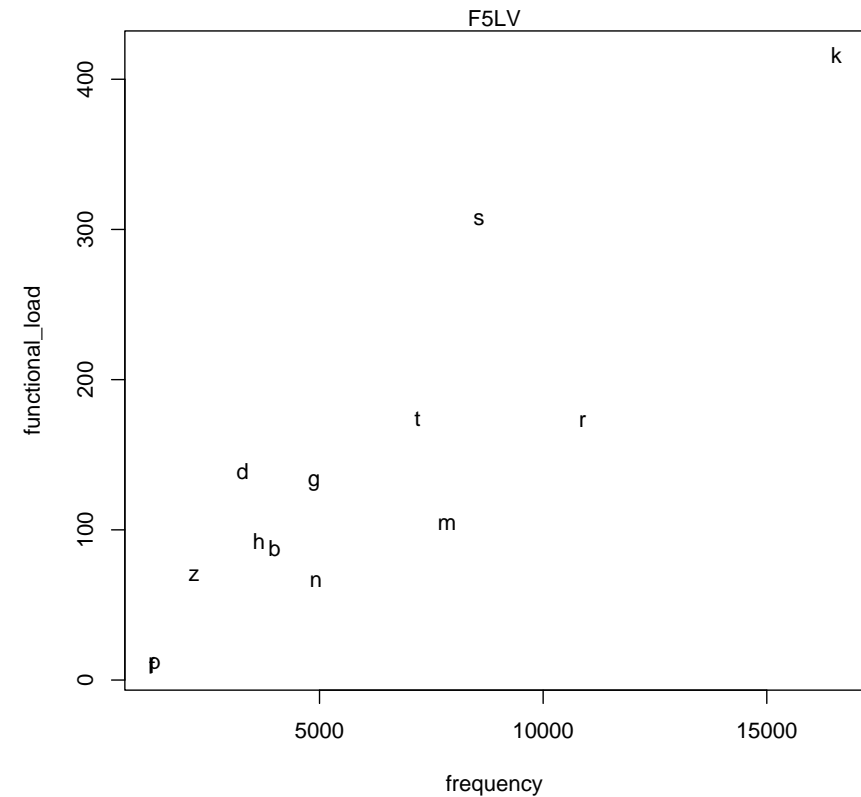
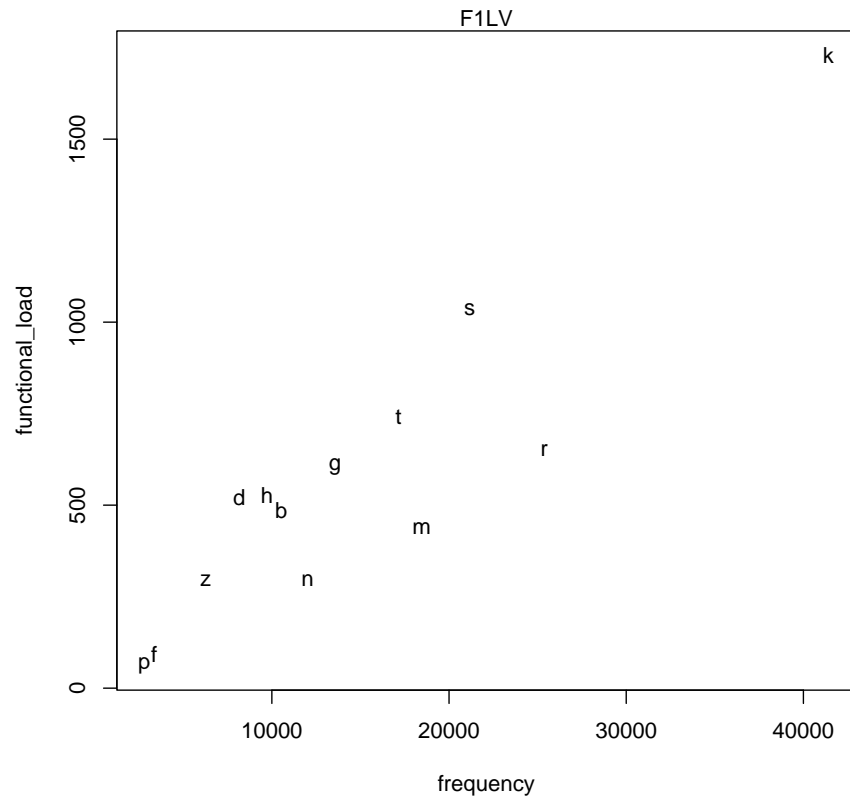
- F4~5で大きく上昇
/s,m,j,K,z,w,N/
- F5~6で大きく上昇
/r,t,n,b,z,w,O,A/
 - /z,w/: 用言の比率増大か?
- F4~5で大きく下降
/u,X,h,b,O,G,B/
- F5~6で大きく下降
/s,o,S,X,j,h,K,H/
 - /X,h/: 漢語の比率減少か?

分析(6): 特殊モーラ



- 撥音/X/の頻度・機能負担量が飛び抜けている
- 長音/O/は頻度の割りに弁別に役立たない

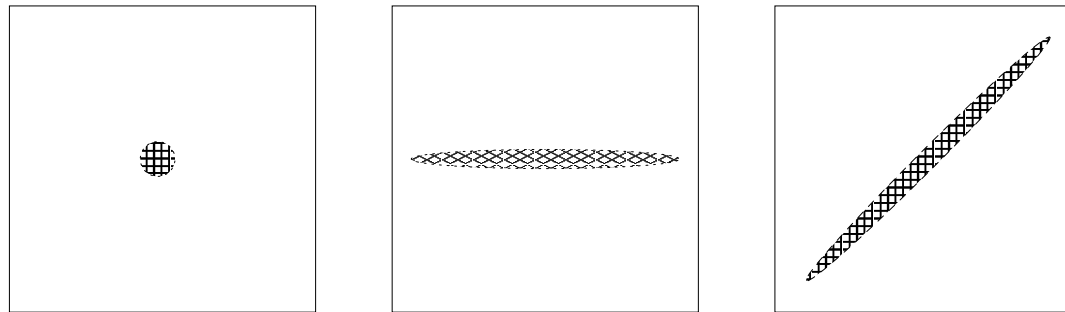
分析(7): 子音(直音のみ)



- /p/-/t/-/k/と/b/-/d/-/g/はまったく異なる分布
- 構造主義の言う「圧力」はどこにいつ働くのか?

おわりに

- もっとも均衡のとれた状態とは何であろうか？
- 等頻度・等機能負担量; 機能負担量のみ等しい; 機能負担量は頻度に比例？



次にやること

- F5~F6を0.2刻みで
- 品詞、語種の類別
- コーパスにおける出現頻度 (token freq) とかけあわせる
- こんなことを考えて一体何が楽しいのか考える