

容認度判定の実態調査の報告

その正体は不均一な反応からなる、バイ
アスのかかった心理評定である

黒田 航 / 杏林大学 医学部 Medical School, Kyorin University

仲村 哲明 / 京都大学 情報学研究科

河原 大輔 / 科学技術振興機構さきがけ

言語処理学会第23回年時大会, 2017/3/15,

筑波大学

問題設定

■ 動機

- 刺激として与えられる文 $E = \{e1, e2, \dots, en\}$ の容認度判定 (容認可能か不可能か) の結果を、より実態に合うようにモデル化したい
- 必要な事
 - 文脈がどういう効果をもつのかはっきりさせる
 - 評定者の個性 (= 反応タイプ) をモデルに取り入れる
 - 評定者が不均一反応をしているのは経験的に明らか

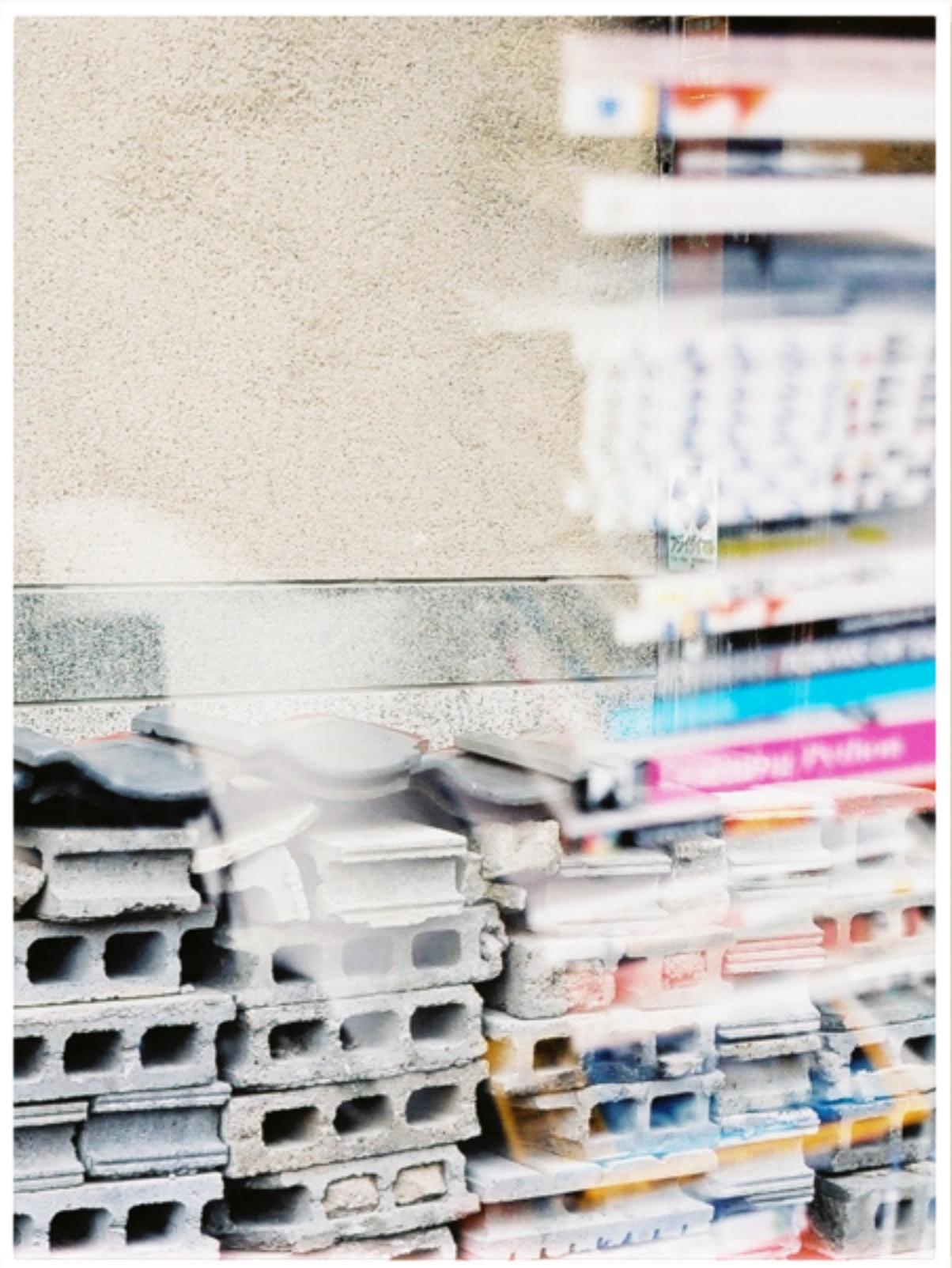
発表の流れ

■ 導入

- 容認度判定 => 多段階の容認度評定に一般化
- 刺激 E の容認度判定 A を, E と言語文脈 C と評定者の反応特性 R の相互作用 $A = E \times C \times R$ として記述
 - C と R の追加が本質的
- 検証
 - $A = E \times C \times R$ モデルを仲村・河原(2016)で実施された容認度判定課題の結果の分析に適用して、有効性を確認
 - 計測結果は $E \times R$ マトリックス
 - E を R を属性として分類 (階層クラスタリング)
 - R を E を属性として分類 (階層クラスタリング)

Part I: Theoretical Stuff

容認度判定のモ デルの精緻化



もっとも単純なモデル 1/3

- 表現 e の、評定者 r による容認度判定 $a(e, r)$ の結果は
 - 文脈の影響を受けない
- 想定 1
 - 想定 2
- どんな文脈 c に e があっても結果が同じ
 - どの判定者も、文脈が同じなら、結果が同じ
 - すべての判定者が一様な反応をする

もっとも単純なモデル 2/3

- 想定 I が妥当でない事は周知で、その影響には文脈効果という名前が付いている
- けれども、文脈(効果)の実態がわかっている訳ではない

もっとも単純なモデル 3/3

- 想定 2の妥当性は言語学内で真剣に検討された事がない
- 以前から判定結果が(何らかの理由で)収束しないという事実は知られていた。
- ただ、それは広い意味での“文脈効果”に押し込められていた

発展させたモデル I/3

- 二つの拡張
 - 二値判定 (judgment) を多値評定 (rating) に拡張
 - これは自明と言えば自明
 - 厳密な意味での文脈効果 C と判定者の反応特性 R を区別
 - 従来の広い意味での文脈 C^* は、言語文脈 C と反応特性 R との相互作用 ($C^* = C \times R$)

発展させたモデル 2/3

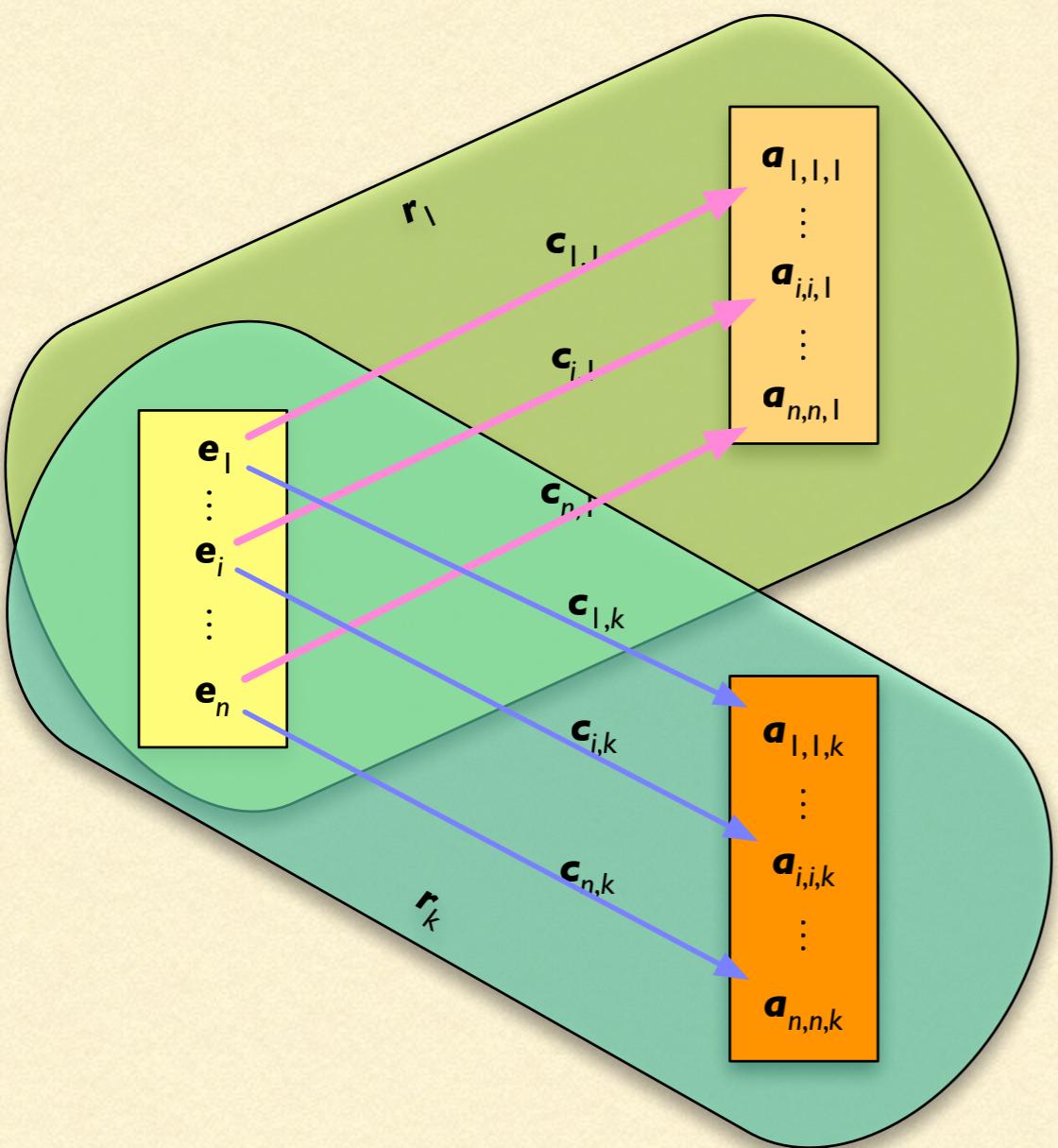
- 以下の手順
- まず R は一様だとして、 C の効果を表現
- その後に R の効果を追加

WHAT REALLY ARE CONTEXTUAL EFFECTS?

変異を取り込む
ための $A = E \times C$
 $\times R$ への拡張



“文脈”概念の精緻化



- まず二種類の文脈 C, C^* を区別
 - 明示的な、言語的文脈 = 狹義の文脈 C
 - 非明示的な、内面化された文脈 = 広義の文脈 C^*
- その上で、 C^* は C の効果と R の効果の混合
 - $C^* = C \times R$
 - だと考える

$A = E \times C$ の概要

	c	...	c	...	c
e	a	...	a	...	a
⋮	⋮	⋮⋮	⋮	⋮⋮	⋮
e	a	...	a	...	a
⋮	⋮	⋮⋮	⋮	⋮⋮	⋮
e	a	...	a	...	a

狭義の文脈効果 C ($\neq C^*$)

- 文脈効果の再規定
 - 表現 e_i は文脈 $C = (c_1, c_2, \dots, c_m)$ の異なりごとに、異なる容認度を有する $A = (a_1, a_2, \dots, a_m)$
 - つまり、 A_i は独自の分布構造を有する
- 直観的な解釈
 - e_i の c_j での容認度は、 e_i の文脈 c_j に対する適応度

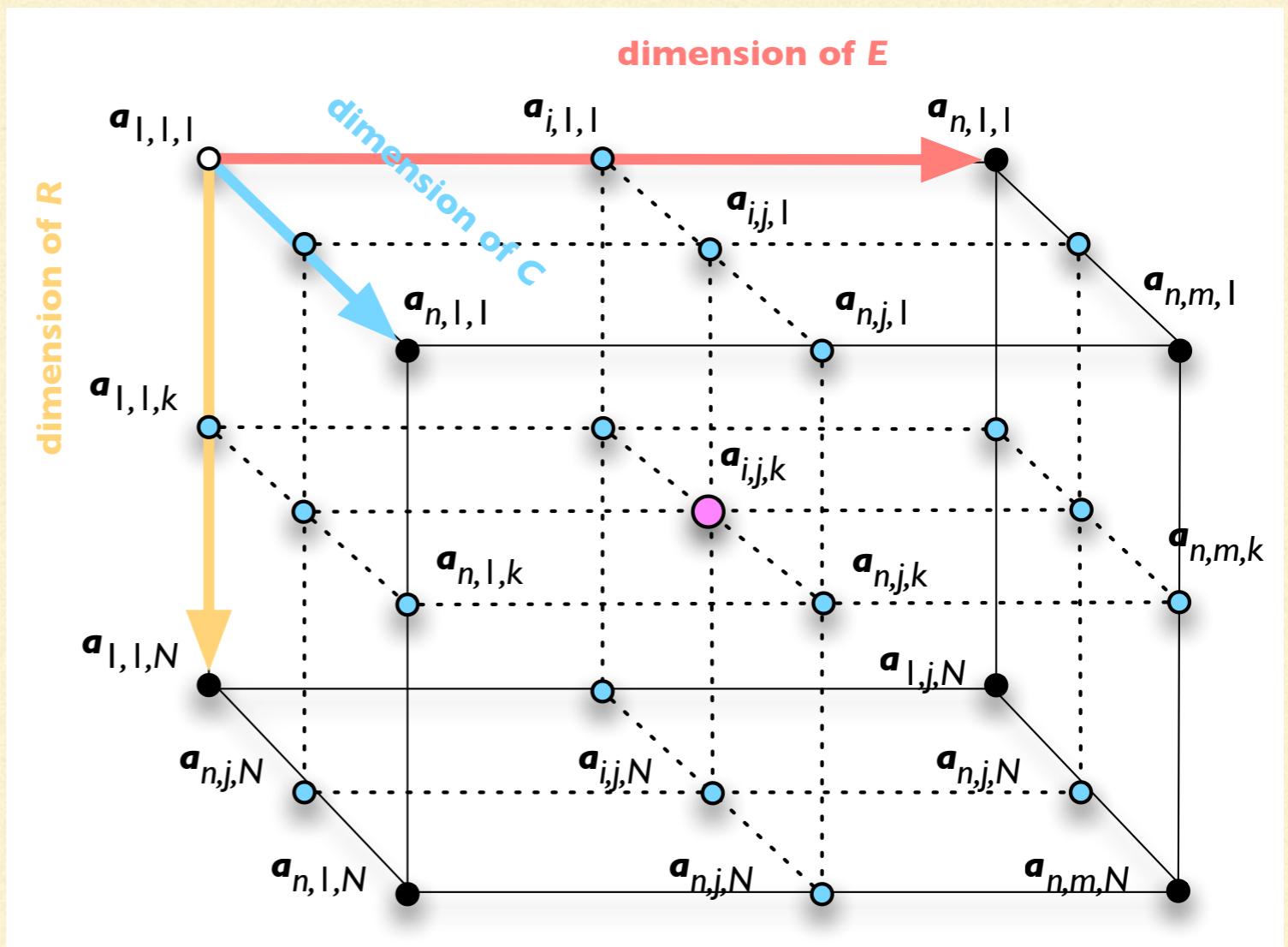
均一反応の仮定の放棄

- 事実
 - 同一の言語文脈 C にある同一の文 e に対しても、評定結果が評定者によって異なる
 - だが、言語学は逆を想定
 - 言語学の想定は均一反応の仮定 (homogeneous response hypothesis)
 - 必要な修正
- 不均一反応を許容
 - ただし
 - 評定者の反応はてんでバラバラという訳でなく、反応パターン P_1, P_2, \dots, P_n がある
 - P_1, P_2, \dots, P_n は解釈戦略に対応
 - Item Response Theory: IRT (Embretson & Reise 2000) を使ってモデル化が可能

Tensor $A = E \times C \times R$

- $A = E \times C$ を拡張し、テンソル $A = E \times C \times R$ を考える

- (赤色の) 次元 E は刺激文のベクトル
- (空色の) 次元 C は明示的文脈のベクトル
- (橙色の) 次元 R は評定者のベクトル



Conflation of $A = R \times E \times C$ (Case I)

$$A_{k,j} = [a_{i,1,k}, a_{i,2,k}, \dots, a_{i,m,k}] \quad (m \text{ is the size } c)$$

	e	...	e	...	e
r	A	...	A	...	A
⋮	⋮	⋮⋮	⋮	⋮⋮	⋮
r	A	...	A	...	A
⋮	⋮	⋮⋮	⋮	⋮⋮	⋮
r	A	...	A	...	A

Conflation of $A = R \times E \times C$ (Case 2)

$$A^{**}_{k,j} = [a_{1,j,k}, a_{2,j,k}, \dots, a_{n,j,k}] \quad (n \text{ is the size of } e)$$

	C	...	C	...	C
r	A	...	A	...	A
⋮	⋮	⋮⋮	⋮	⋮⋮	⋮
r	A	...	A	...	A
⋮	⋮	⋮⋮	⋮	⋮⋮	⋮
r	A	...	A	...	A

Part 2: Empirical Stuff

実データを使つ
た $A = E \times C \times R$
の実用性の検討



見取り図

- 提示の流れ
 - R を属性に使った $E (= S)$ の階層クラスタリング
 - E を属性に使った R の階層クラスタリング
 - いずれも欠損値ありクラスタリング (Maechler, et al. 2015) を利用
- 得られた知見
 - E の4クラスターは逸脱のタイプに対応
 - R の5クラスターは評定者の応答の癖に対応

データの説明 1/2

入稿したタスクの画面プレビュー

プレビュー表示切り替え：パソコン用 スマートフォン用

前の設問へ 次の設問へ

設定した設問ID：132

この文は自然な文ですか？

私が弁当を作る

はい
いいえ

表示された文の内容が自然に（簡単に）わかる場合だけ「はい」を選んで下さい。
内容が簡単にわからない場合または判断に迷う場合は「いいえ」を選んで下さい。

（例1）：「母が電車に乗る」や「車が道路を走る」のような文では「はい」を選んで下さい。
（例2）：「学校が電車に乗る」や「コーヒーが道路を走る」のような文では「いいえ」を選んで下さい。

■ 利用したデータ

- 仲村・河原 (2016) が格フレーム辞書を使って自動生成した文を Yahoo! クラウドソーシングで容認判定させた結果
- 10文が検査用, 1935が本番用 probes

■ 課題の概要が左の図

- 期待される反応, 自然な文か否か
- 厳密には言語学が念頭に置いている容認度を聞いているとは言えないが, 操作化としては妥当

データの説明 2/2

- 検査用の刺激文 $D = \{d1, d2, \dots, d10\}$
 - $\{d1, \dots, d5\}$ が誰もが容認すると想定される文
 - $\{d6, \dots, d10\}$ が誰もが容認しないと想定される文
- 基本統計
 - 最小値 10 で最大値 99
 - 平均して 88.9 文を評価
- 1 文が平均して 9.13 名の評定を受けた
 - フィルタリング
 - 244 名が参加したが、逸脱反応をしている 59 名を除いて 185 人の反応を採用
 - 基準
 - $0.15 < \text{平均容認率 } r < 0.85$
 - かつ
 - $80 < \text{評価分数 } n$

DI-DI0

R.COUNT	D-ID	S	EXPECTED PATTERN	GLOSS
190	1	太郎が学校に行く	1	Taro goes to school.
250	2	太郎がコーヒーを飲む	1	Taro drinks coffee.
190	3	太郎が本を読む	1	Taro reads books/a book.
300	4	太郎が自転車に乗る	1	Taro rides on bicycles.
230	5	太郎がボールを投げる	1	Taro throws a ball/balls.
200	6	太郎が花を読む	0	Taro reads flowers.
160	7	太郎が電車を食べる	0	Taro eats trains.
230	8	太郎が恐竜を寝る	0	Taro sleeps a dinosaur.
160	9	太郎がパソコンを泣く	0	Taro cries a computer.
240	10	太郎が黒板を運転する	0	Taro drives a blackboard.

Stimuli for testing (10 random samples)

S.COUNT	S-ID	S	STATUS
10	5	人が出会いを頂く	Odd
10	227	人が下りるを待つ	Odd
10	442	太陽が顔を出す	Fine
10	617	対戦がレースを楽しめる	Odd
10	722	署員が煎るを見付ける	Odd
10	761	グループが結果を表に纏める	Awkward
10	1203	人が予備軍を含める	Awkward
10	1695	人が日祝を除く	Awkward
10	1757	人が沢を詰める	Awkward
10	1857	ペンギンが空を実際に飛ぶ	Fine, but factually odd

What can we know about S by
clustering it?

E の階層クラス
タリング



解析法について

■ 難点

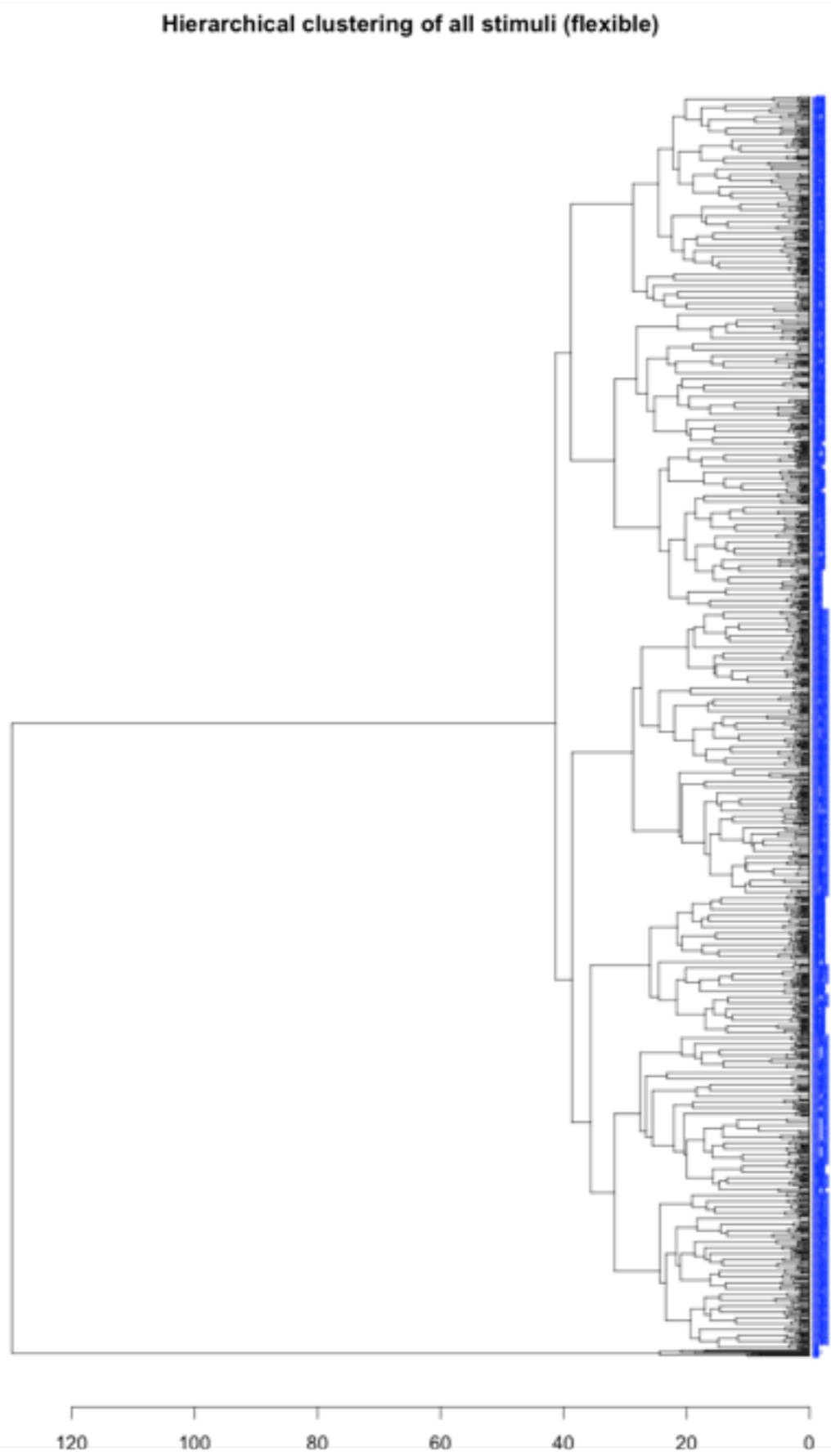
- 生データは欠損値 (missing values) だらけ
- 185名に限定しても、D の他に全員が評定している刺激文がない

■ 対処

- 欠損値ありクラスタリング
(clustering with missing values)
(Maechler, et al. 2015) を実行

■ 補足

- 別の人工データを使った実験で、ランダムに70%, 80%, 90% のデータを欠損させた場合、90%までは原型を保存される事を確認
- ただし枝葉末節の位置は搅乱される



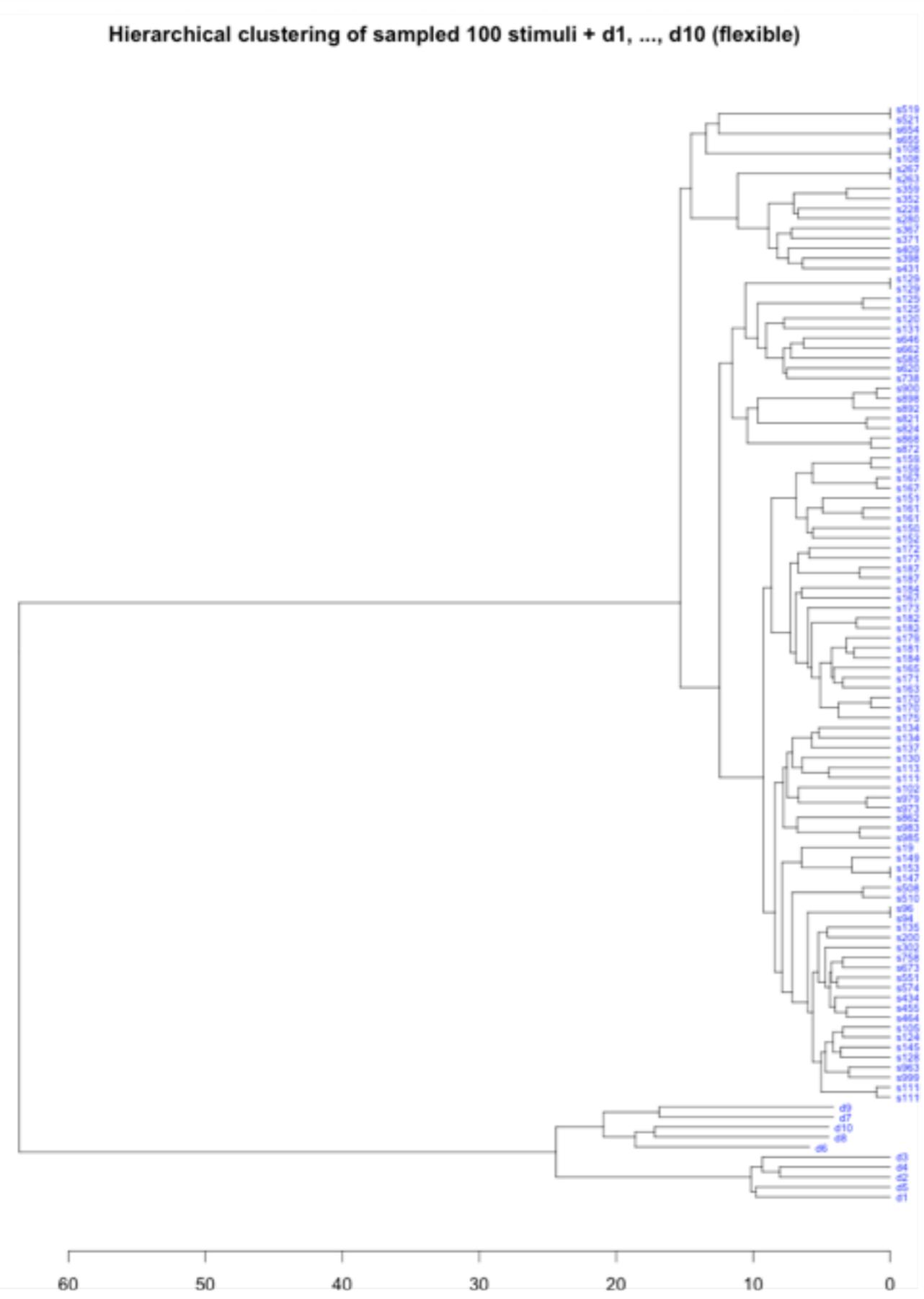
全1945事例の階層クラスタリングの結果

- 解析法
 - 欠損値ありクラスタリング (Maechler, et al. 2015)
 - 対象 $E = S + D$
 - $S = \{s1, s2, \dots, s1935\}$, $D = \{d1, d2, \dots, d10\}$
- 概要
 - $D = \{d1, d2, \dots, d10\}$ が S から完全に分離している

解釈

- Sへの反応(の平均)は Dへの反応(の平均)と異質
- つまり, S と D は質の異なる刺激
- これは言語学の一様反応の仮定 (assumption of uniform responses) に明らかに矛盾しており, $A = E \times C \times R$ の必要性を実証

Hierarchical clustering of sampled 100 stimuli + d1, ..., d10 (flexible)



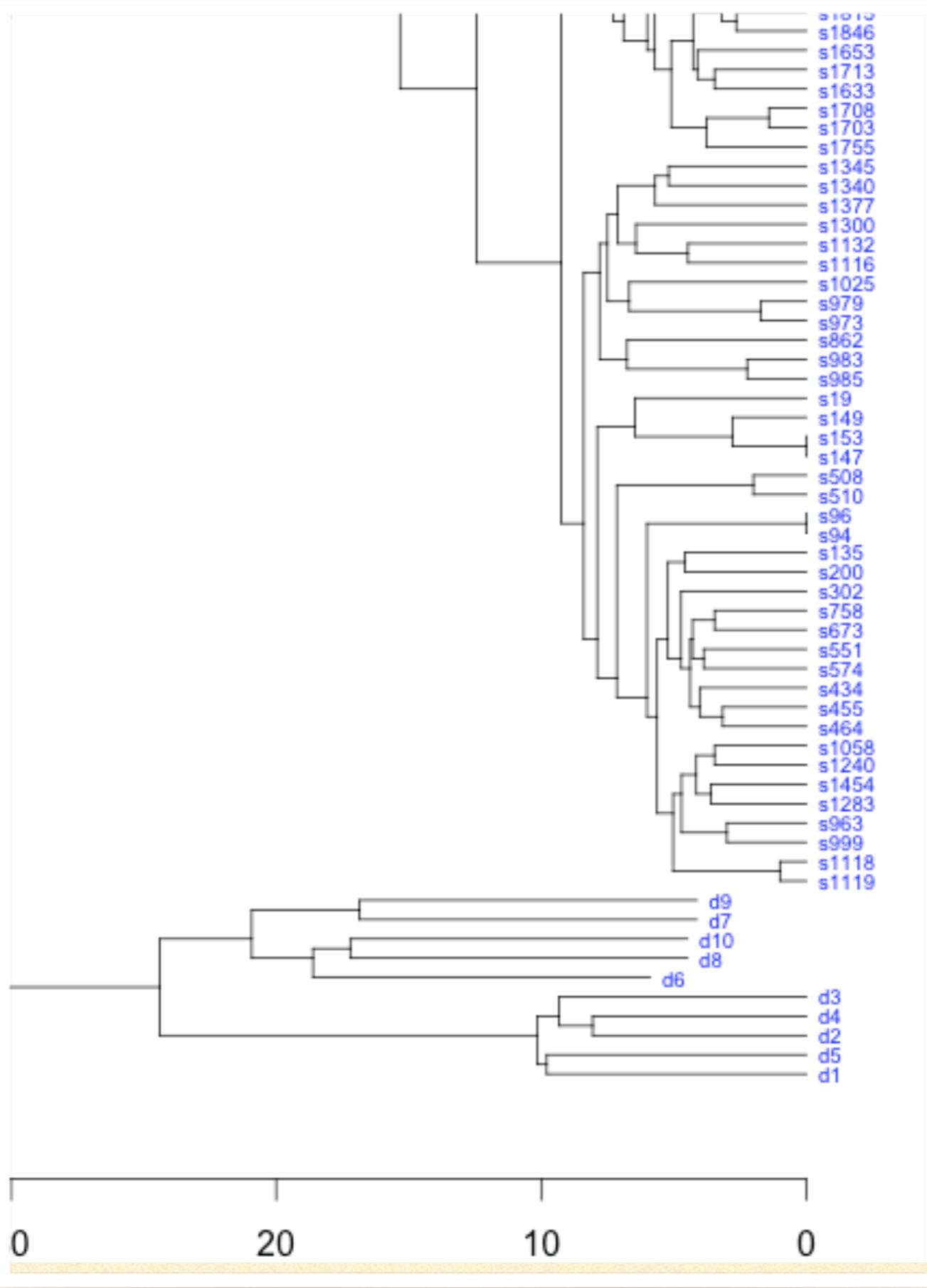
S+D 混合の100例を無作為サンプリングして解析

■ 解析法

- 欠損値ありクラスタリング (Maechler, et al. 2015)
- $E = S = \{s1, s2, \dots, s1935\} + D = \{d1, d2, \dots, d10\}$ の100例のランダムサンプリング

■ 概要

- $D = \{d1, d2, \dots, d10\}$ が S から完全に分離しているのが確認できる
- $D1 = \{d1, d2, \dots, d5\}$ と $D2 = \{d6, d7, \dots, d10\}$ が分離しているのが確認できている



S+D 混合の100例を無作為サンプリングして解析 (Dの拡大図)

■ 概要

- $D_1 = \{d_1, d_2, \dots, d_5\}$ と $D_2 = \{d_6, d_7, \dots, d_{10}\}$ が分離しているが
- D_1 (理想的に容認可能な文の集合) と D_2 (理想的に容認可能な文の集合) では平均相互類似度が異なる
 - D_1 は相互類似度が高いが、 D_2 は相互類似度が D_1 に較べて低い

■ 含意

- 容認度の低い刺激に対する反応はばらつきが大きい

Sのみの階層クラスタリング

■ 解析法

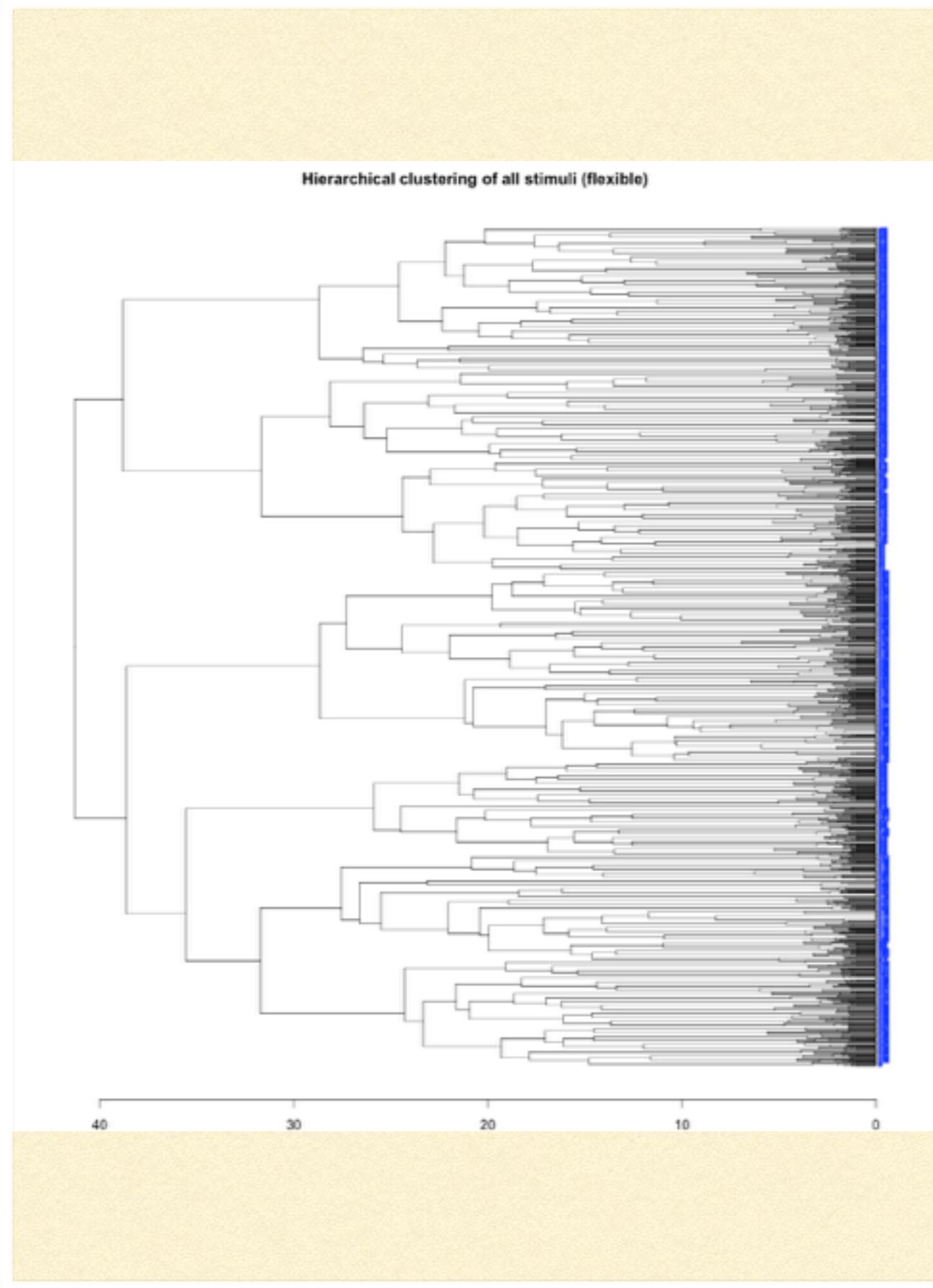
- 欠損値ありクラスタリング (Maechler, et al. 2015)

$$\text{■ } E = S = \{s1, s2, \dots, s1935\}$$

■ 結果

- Major division: $G^*1 = \{C^*1, C^*2\}$ and $G^*2 = \{C^*3, C^*4\}$

- Further divisions: C^*1, C^*2, C^*3, C^*4 , and C^*5



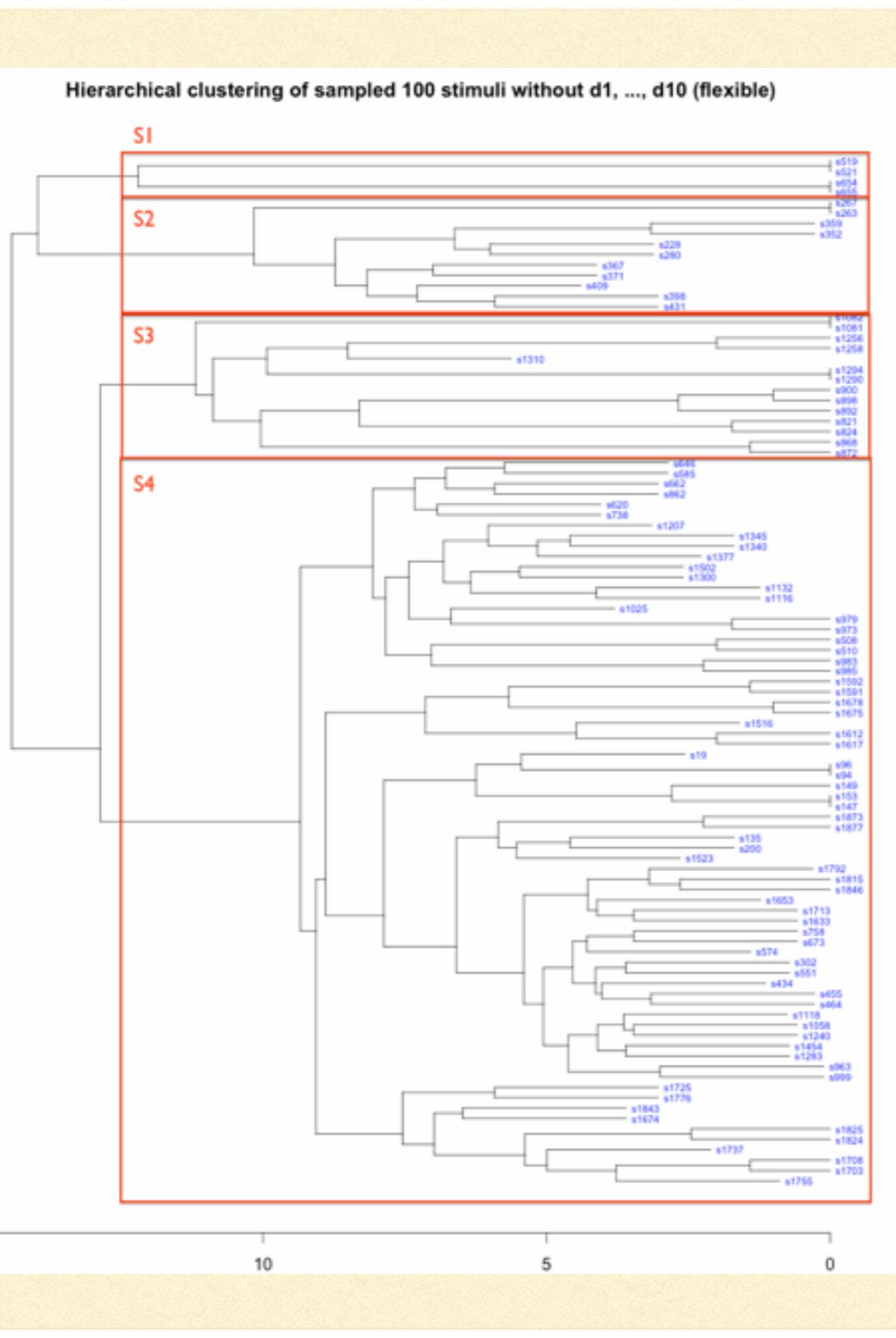
S の100事例の無作為抽出 結果の階層クラスタリング

■ 結果

■ 大分類: {S1, S2} と {S3, S4}

■ 中分類: S1, S2, S3 と S4

■ 中分類を赤い枠で明示



S から無作為抽出された100例の分布

- S1
 - s519, s521, s654, s655
- S2
 - s267, s263, s359, s352, s228, s280, s367, s371, s409, s398, s431
- S3
 - s1082, s1081, s1256, s1258, s1310, s1294, s1290, s900, s898, s892, s821, s824, s868, s872
- S4
 - s646, s585, s662, s862, s620, s738, s1207, s1345, s1340, s1377, s1502, s1300, s1132, s1116, s1025, s979, s973, s508, ..., s1755

Examples of SI

SID	S	Rating Average	Rating Stdev
s519	選手が監督を有るにやる	0.00	0
s521	プロが試合をやるにやる	0.00	0
s654	数々が両方を楽しめる	0.00	0
s655	引きが展開を楽しめる	0.00	0

Examples of S2

SID	S	Rating Average	Rating Stdev
s267	情報が改札を画面に出る	0.00	0
s263	記事が出口を上に出る	0.00	0
s359	人がモーツァルトを聞く	1.00	0
s352	人が期間を入れる	0.00	0
s228	人が伸びるを待つ	0.10	0.316227766
s280	行動が域を裏目に出る	0.00	0
s367	人がシリーズを含む	0.00	0
s371	人がカードを含む	0.00	0
s409	人がフォームを含む	0.00	0
s398	人が効果を含む	0.00	0
s431	人が円を掛ける	0.10	0.316227766

Example of S3

SID	S	Rating Average	Rating Stdev
s1082	金が頭をところに使える	0.00	0
s1081	金が頭を使える	0.00	0
s1256	道路が下を真っ直ぐに走る	0.50	0.534522484
s1258	私がトップを主義に走る	0.00	0
s1310	激震が世界を業界に走る	0.00	0
s1294	風が海岸を走りに走る	0.00	0
s1290	ランナーがメートルをトイレに走る	0.00	0
s900	比率が%を超える	0.13	0.353553391
s898	者数が人を超える	0.00	0
s892	人がイヴを過ごす	0.75	0.46291005
s821	人が出汁汁を加える	0.38	0.51754917
s824	客が同店を店に訪れる	0.00	0
s868	人がブラックを調べる	0.33	0.5
s872	人が含むを調べる	0.11	0.333333333

Example of S4 | /2

SID	S	Rating Average	Rating Stdev
s646	海水浴がサーフィンを楽しめる	0.11	0.333333333
s585	人が有るを学ぶ	0.00	0
s662	コントラストが効果を楽しめる	0.33	0.5
s862	センサーが無礼を働く	0.29	0.487950036
s620	こなしがスタイルを楽しめる	0.11	0.333333333
s738	雪解けが距離を進む	0.00	0
:	:		
s1025	人が臨席を賜る	0.00	0
s979	人が数式を用いる	1.00	0
s973	人が計算を用いる	0.70	0.483045892
s508	社員が仕事をやる	0.78	0.440958552
s510	人が役をやる	0.78	0.440958552
:	:		
s19	人が土日を使う	0.14	0.377964473
s96	妻が小言を親に言う	1.00	0
s94	妻が小言を言う	1.00	0
:	:		

Example of S4 2/2

SID	S	Rating Average	Rating Stdev
s1725	人が食べるを見掛ける	0.11	0.333333333
s1776	アルバムがCD作れる	0.25	0.46291005
s1843	コントラストが効果を楽しめる	0.14	0.377964473
s1674	ヒントが回答を貰える	0.50	0.527046277
s1825	資格が元気を貰える	0.25	0.46291005
s1824	人が金を貰える	1.00	0
s1737	人が気付きを貰う	0.44	0.527046277
s1708	人がリストを眺める	1.00	0
s1703	人が短縮化を図る	0.71	0.487950036
s1755	私があなたを殺す	1.00	0

Findings | /3

- 知見 I
- $S4 \Rightarrow S3 \Rightarrow S2 \Rightarrow SI$ の外心構造がある
- $S4$ が“普通の文”で,
- 外に行くにつれて容認度が下がる. Acceptability gets lower along the path of $S4 \Rightarrow S3 \Rightarrow S2 \Rightarrow SI$,

Findings 2/3

- 知見 2
- S3, S2, SI にはそれぞれ対応する逸脱の質がある
- S3 は必要な要素が欠けていて、容認度が下がっている
- S2 は要素の過不足はないが、選択制限の違反がある
- SI は不必要的要素があり、容認度を下げている

Findings 3/3

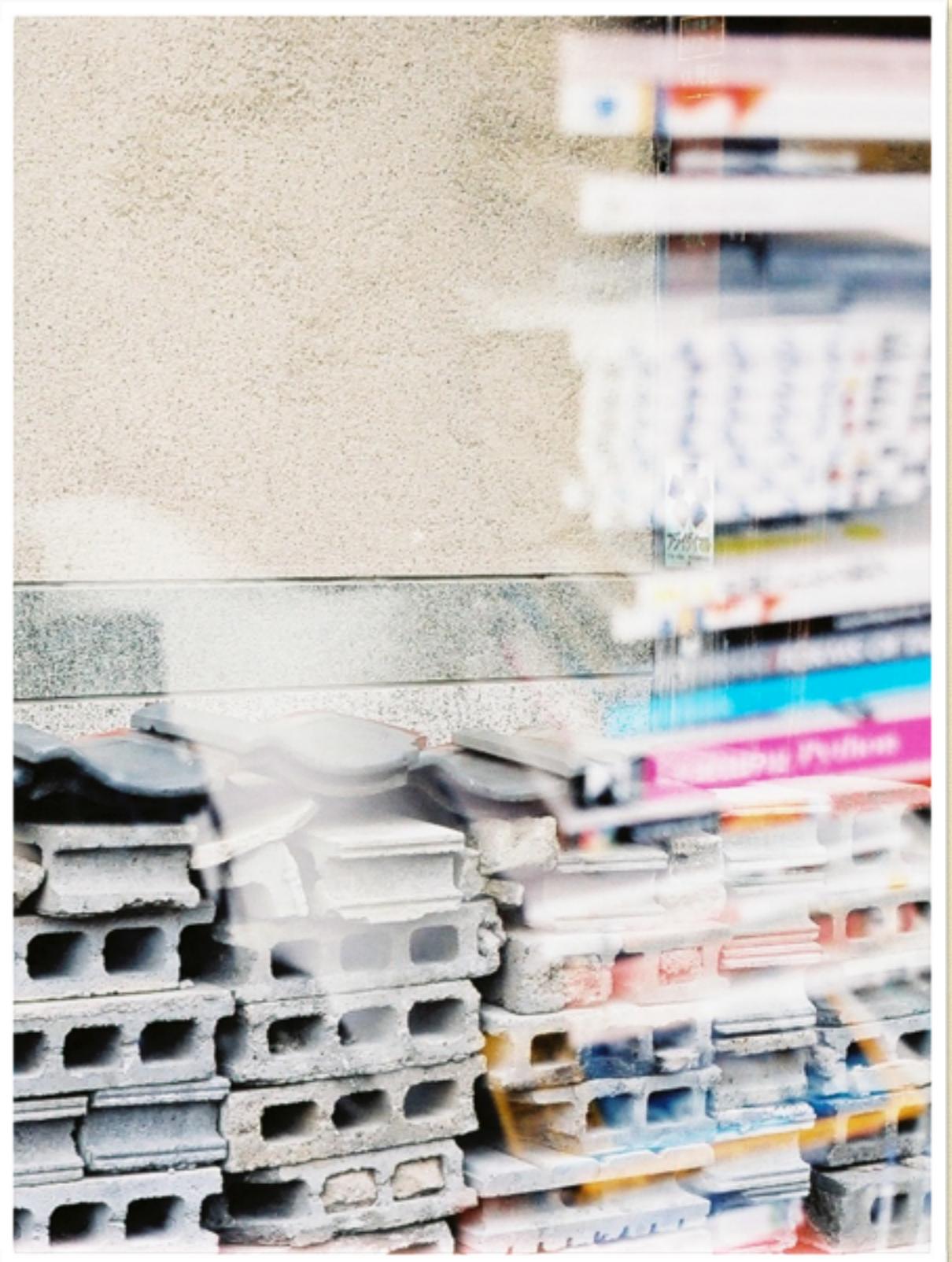
- 知見 3
 - 同一の述語が使われている文の容認度は近くなる
 - ただ、これは順序効果の現われなのかも知れない
 - 実際、この予想が事後確認された
- 含意
 - Yahoo! クラウドソーシングにランダマイズの必要な刺激を渡す時は、依頼者側がしないといけない

Discussion

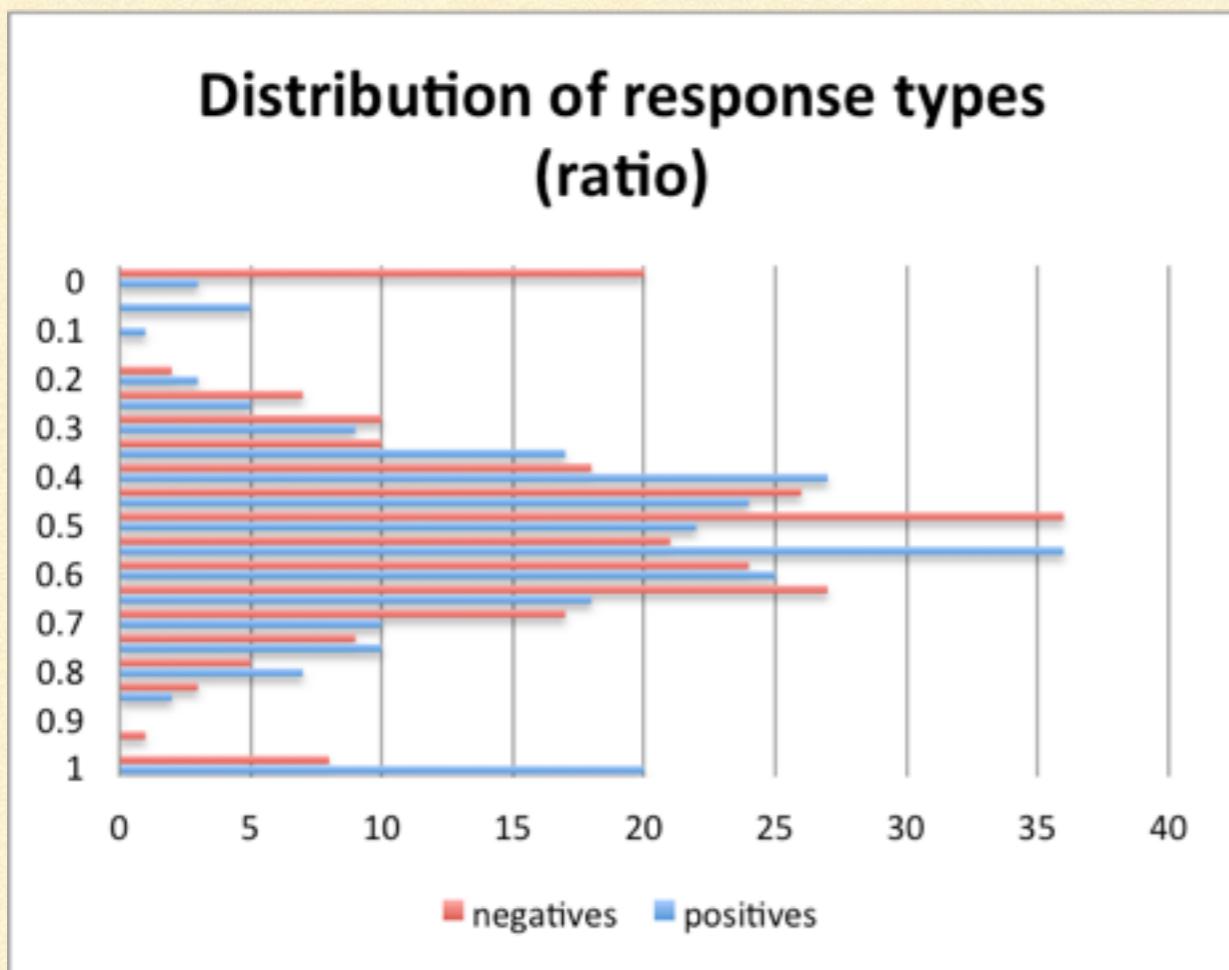
- 一般に刺激文の容認度評定は
 - D と S の質が明らかに異なる
 - S への反応は変異が大きいが、逸脱の質を反映するクラスターをなす
 - しかし、どうやら
- D の内部構造と S の内部構造は相似
- これらは従来の単純なモデルで説明し切れない
- 残された問題
 - R の効果は何なのか？

What can we know about R by
clustering it?

R の階層クラス タリング



平均容認度の分布(ハズレ値除去前)

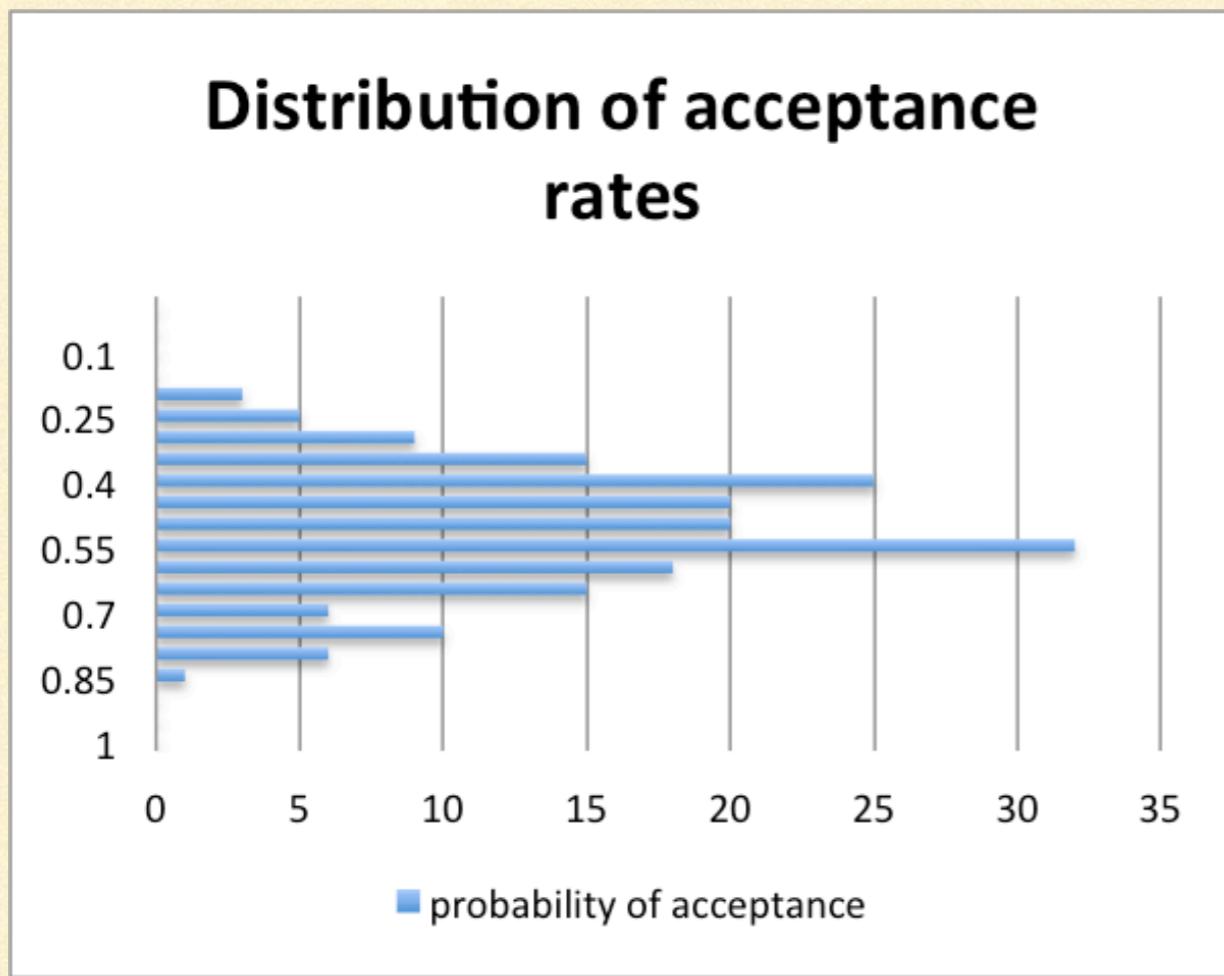


- 244 raters
 - Max: 1.00; Min: 0.00
 - Average: 0.51; Median: 0.51
 - Stdev: 0.22

ハズレ値の除去 (データ整形)

- Filter 1
 - 平均容認度が **0.86** より小さく **0.15** より大きい
- Filter 2
 - 80事例より多い評定をしている
 - 最大値は99
- このに条件で 244 名の評定者を 185名に
- 仲村・河原 (2016)はマクロ対策をしたが不十分だった
- クラウドソーシングは、手抜きをしたい *turkers* と手抜きをさせたくない *taskers* の間の“終わりのない軍拡競争”

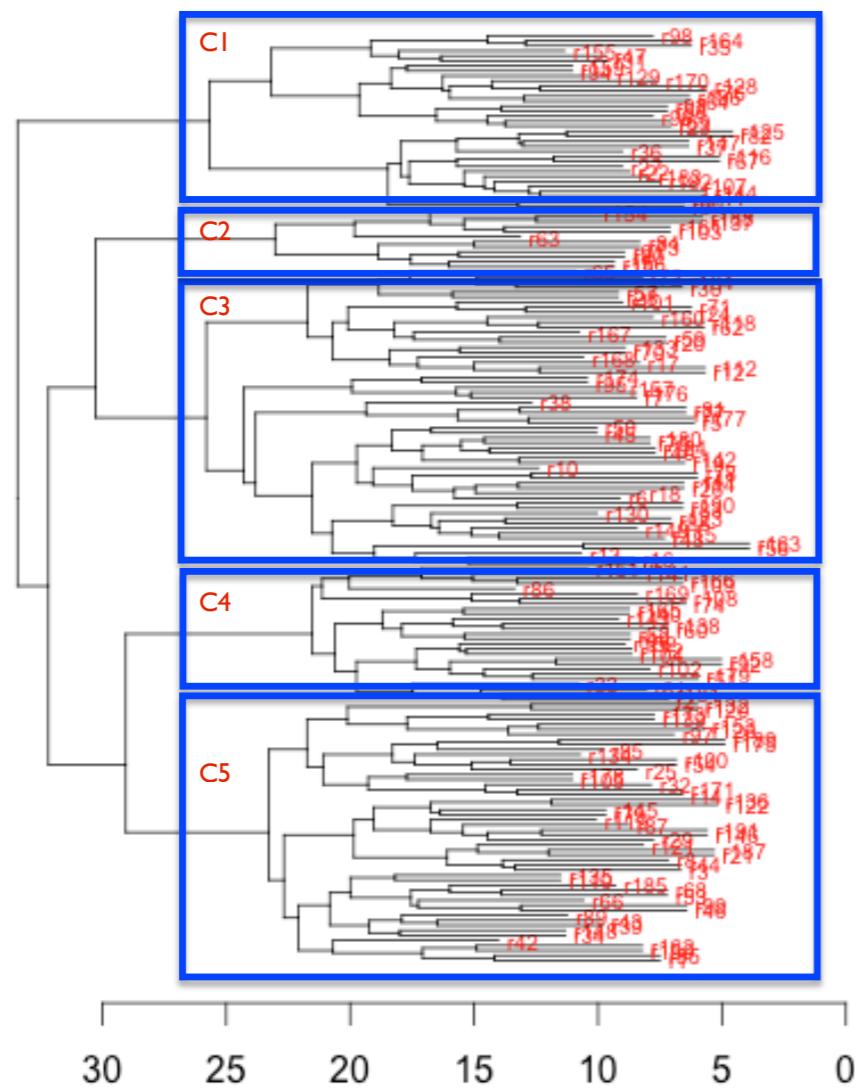
平均容認度の分布(ハズレ値除去後)



- 185 raters
 - Max: 0.84; Min: 0.16
 - Average: 0.49; Median: 0.48
 - Stdev: 0.14
- Remarks
 - 分布が Gauss 分布に見える
(けど、未検証).

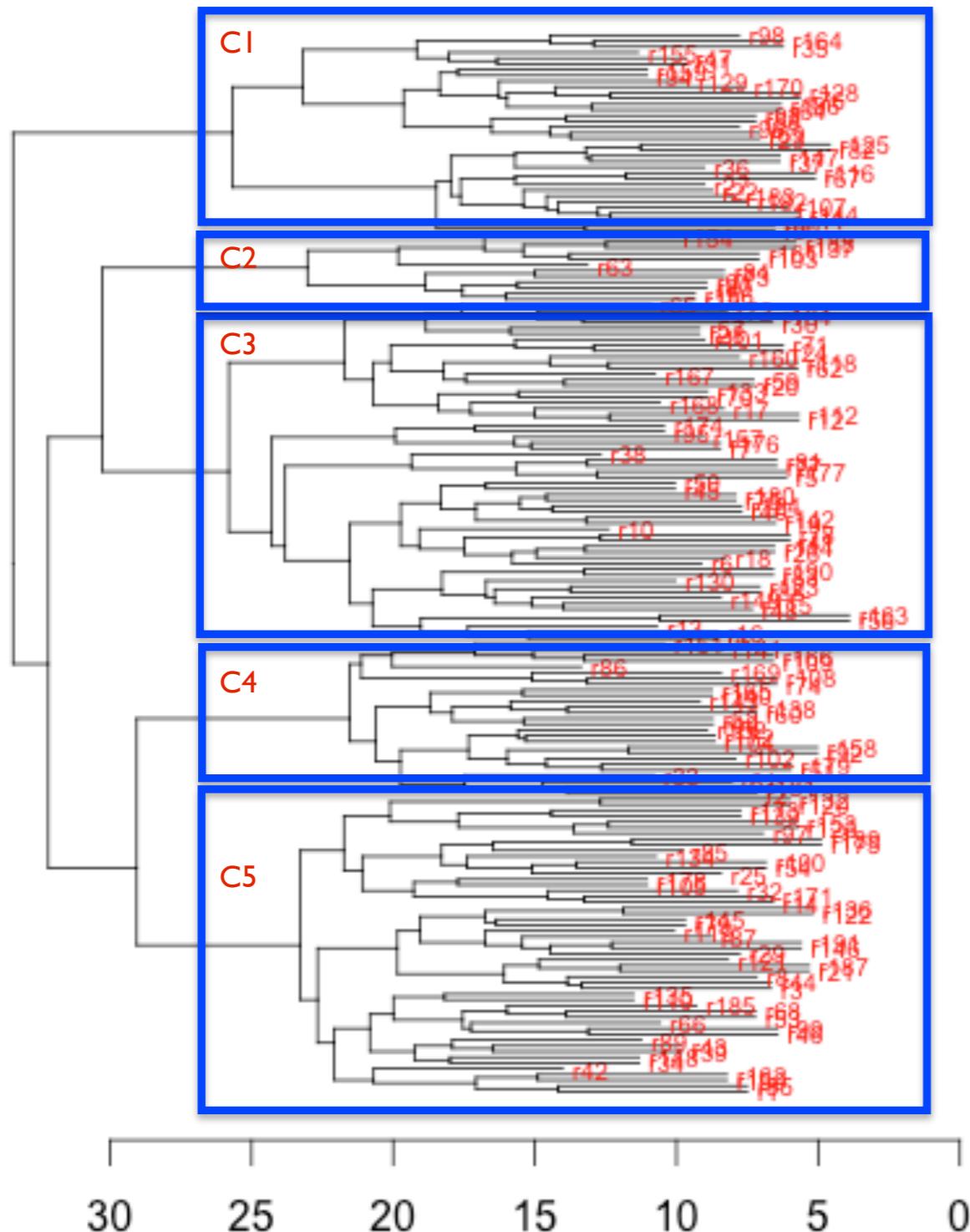
R の階層クラスタリング

Hierarchical clustering of raters (flexible)



- 185人分の反応を 0/1 ベクターにエンコード
 - $(s1, s2, \dots, s1935, d1, \dots, d10)$ のそれぞれに 0 か 1 の値がある
 - 結果
 - 細かく見るとクラスターが5つ
 - C_1, C_2, C_3, C_4, C_5
 - 粗く見るとクラスターが3つ
 - $C_1, \{C_2, C_3\}, \{C_4, C_5\}$

Hierachical clustering of raters (flexible)



拡大図

結果

- 細かく見るとクラスターが5つ
 - C_1, C_2, C_3, C_4, C_5
- 粗く見るとクラスターが3つ
 - $C_1, \{C_2, C_3\}, \{C_4, C_5\}$

クラスターの構成

■ C1 [35]:

- r98, r164, r35, r155, r47, r31, r159, r94, r129, r170, r128, r76, r186, r64, r93, r58, r96, r69, r23, r125, r82, r147, r37, r36, r116, r67, r27, r22, r183, r182, r107, r144, r11, r90, r9

■ C2 [12]

- r154, r188, r137, r161, r103, r63, r84, r73, r91, r80, r156, r15

■ C3 [59]

- r65, r72, r184, r30, r57, r28, r101, r71, r24, r160, r118, r62, r167, r59, r20, r133, r70, r168, r17, r112, r12, r174, r95, r157, r176, r7, r38, r81, r77, r177, r5, r50, r45, r180, r75, r124, r40, r142, r19, r10, r78, r41, r114, r26, r18, r6, r190, r83, r130, r123, r88, r149, r115, r48, r163, r56, r13, r16, r4

■ C4 [27]

- r151, r141, r166, r109, r86, r169, r108, r74, r165, r140, r143, r138, r60, r52, r49, r152, r172, r104, r158, r92, r102, r179, r51, r33, r61, r181, r2

■ C5 [52]

- r132, r120, r173, r139, r153, r126, r97, r189, r175, r85, r134, r100, r54, r25, r178, r105, r32, r171, r14, r136, r122, r145, r79, r119, r87, r191, r146, r29, r121, r187, r21, r8, r44, r3, r135, r110, r185, r68, r53, r66, r99, r46, r89, r43, r39, r148, r34, r42, r162, r150, r55, r1

■ 注意

- 欠損値ありクラスタリングでは枝葉末節の位置に搅乱あり➡ 帰属が正しくない事例がある

新しい問題

- Q1. クラスターの違いは何を反映しているのか？
 - どうも順序効果が発生しているのは明らか...
- Q2. 評定者の個性を特定/記述できるか？

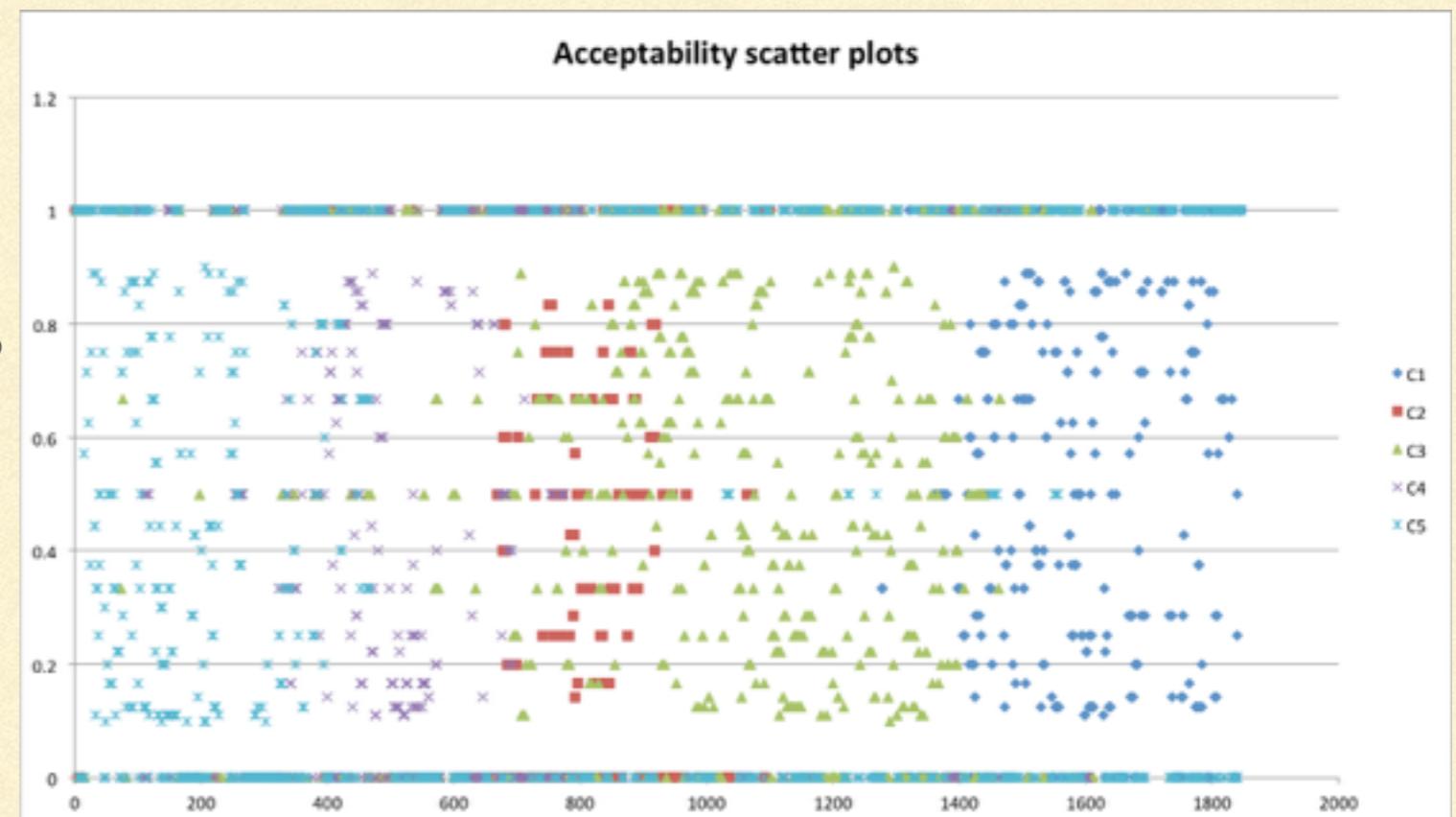
順序効果の解明



反応の平均値の散布図

手順

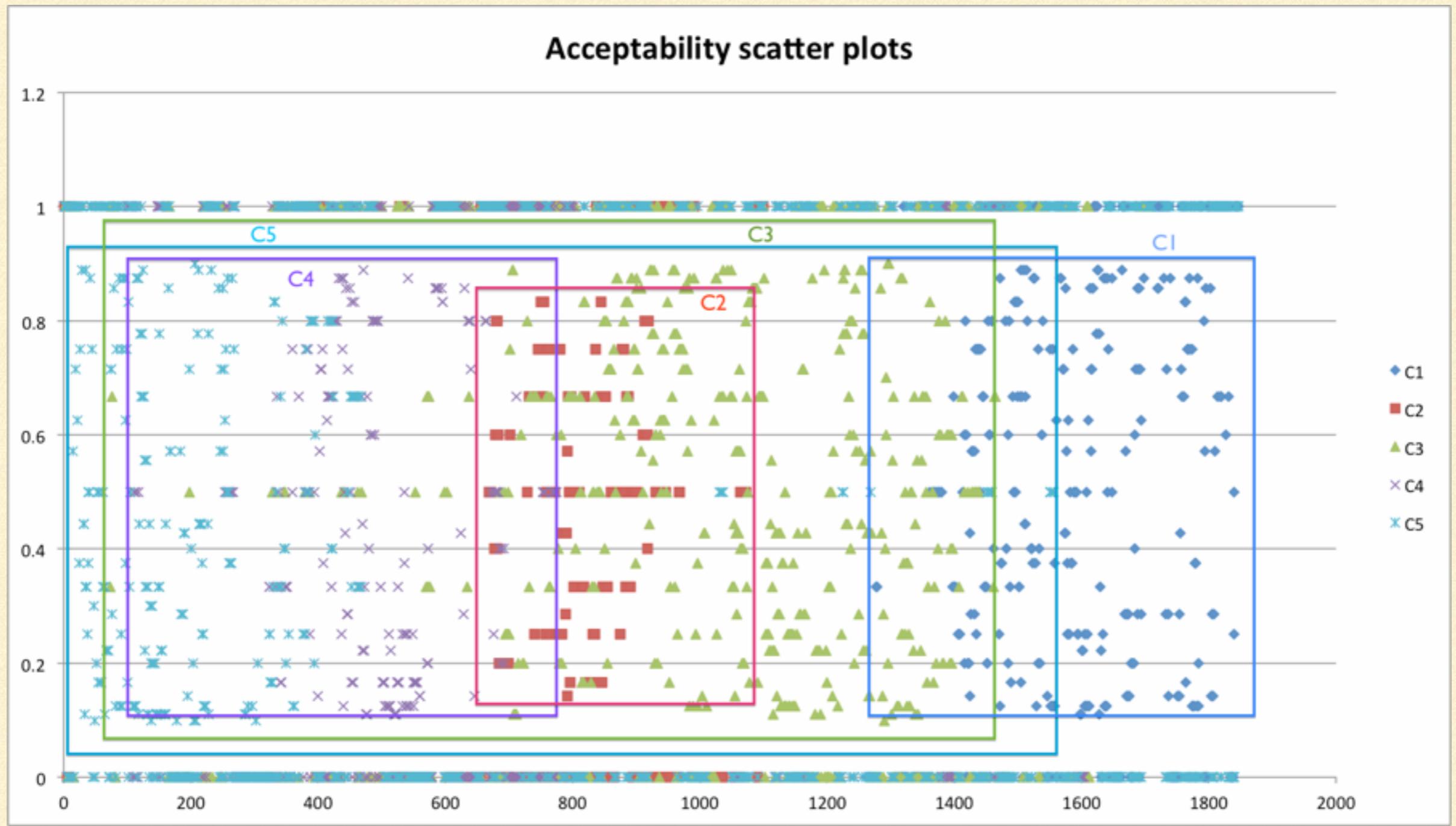
- I, s2, ..., s99, dI, d2, ..., d9, s100, ..., s999, d10, s1000, ..., s1935 に, C_1, \dots, C_5 のクラスターごと平均容認度を求めてプロット



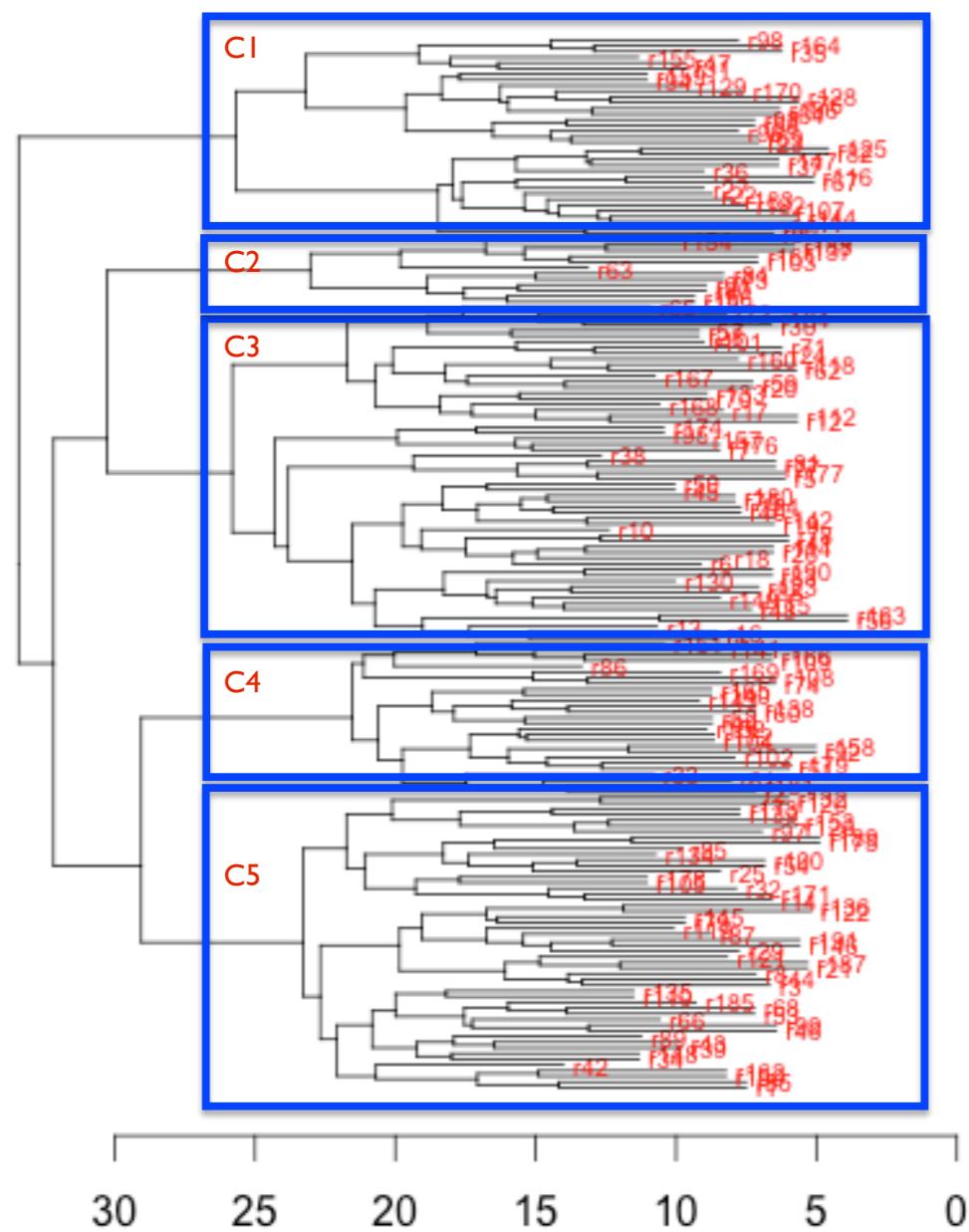
- a \Rightarrow blank

反応の平均値の散布図

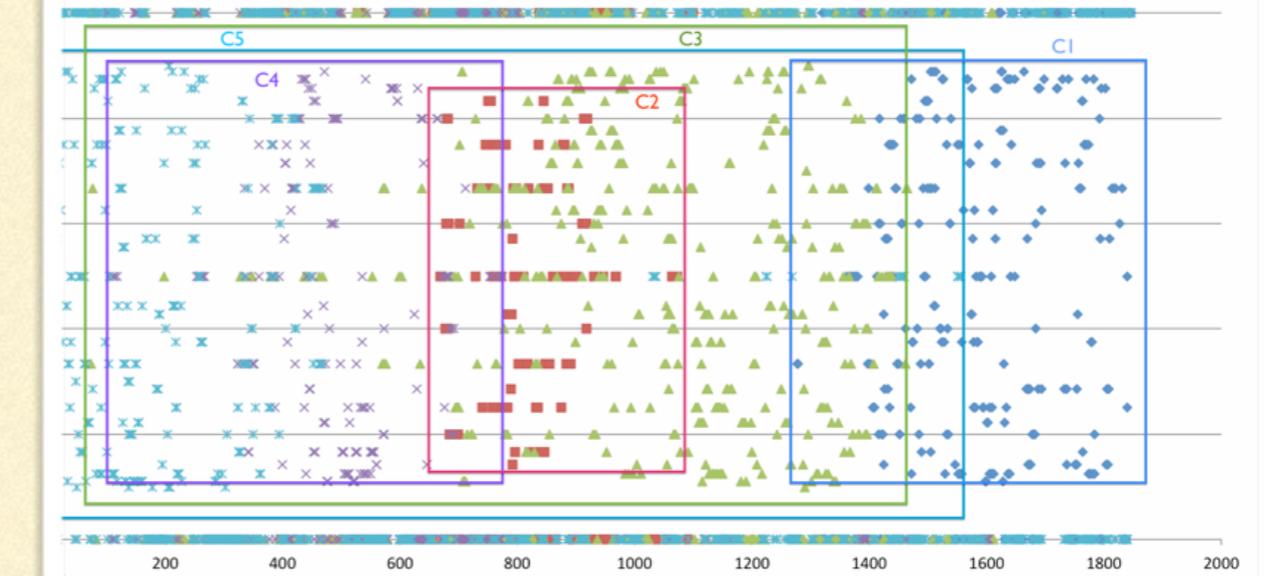
associated with C_1, C_2, \dots, C_5



Hierachical clustering of raters (flexible)



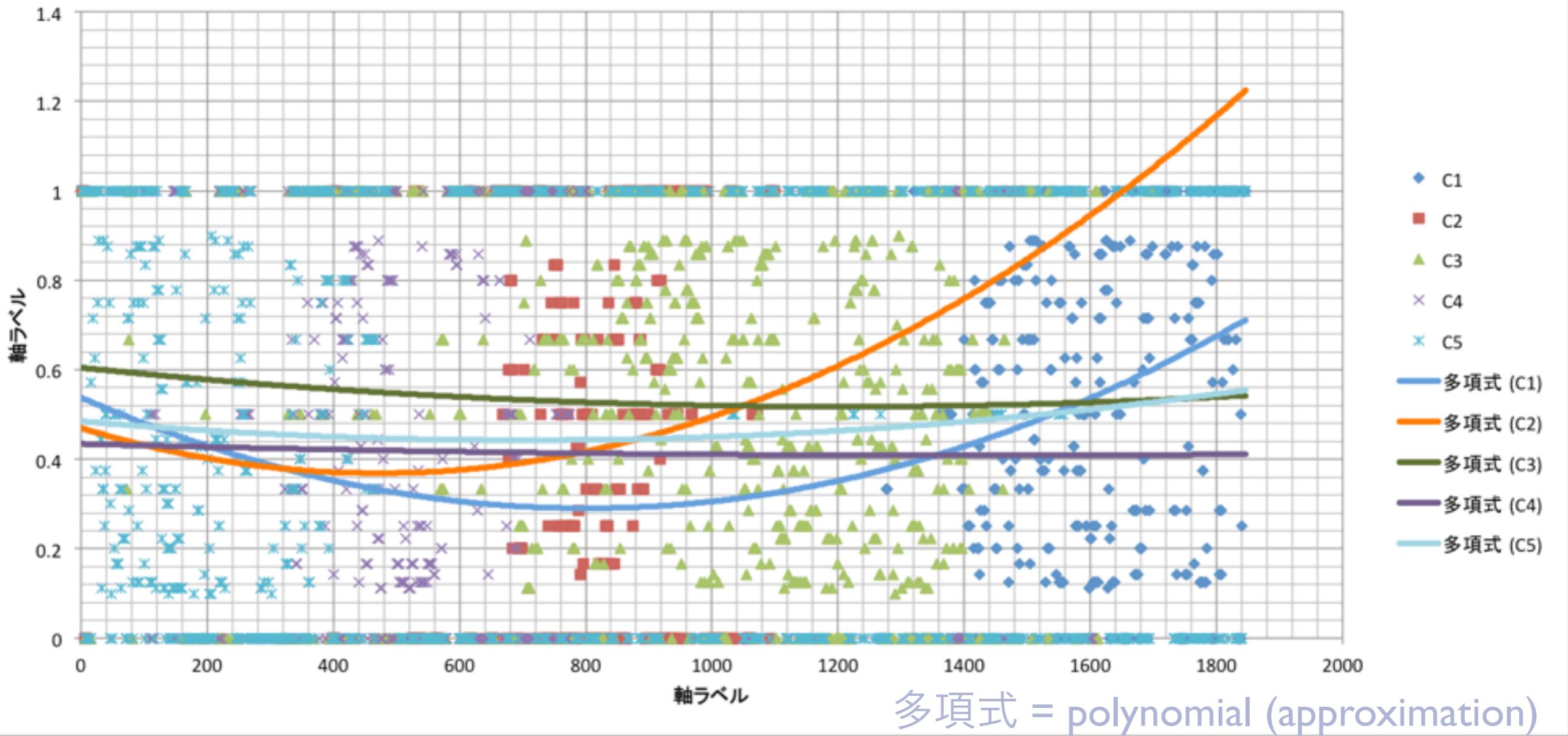
Acceptability scatter plots



- 階層クラスタリングのクラスターとの対応
- 順序効果は明らか

散布図データへのfitting (2次多項式)

X-axis: s1,...,s99,d1,d2,d3,d4,d5,d6,d7,d8,d9,s100,...,s999,d10,s1000,...



散布図データへのfitting (3次多項式)

X-axis: s1,...,s99,d1,d2,d3,d4,d5,d6,d7,d8,d9,s100,...,s999,d10,s1000,...



3次 fitting に基づく反応特性の解釈

- C4 (紫色)とC5 (空色)は一貫した評定を続けるが、 C4 は最後に評定が低くなる。
- CI (青色)とC2 (橙色)とC3 (緑色)は一貫した評定を続けない。
- CI は中庸な評価から始めるが途中で何も容認しなくなり、また元に戻る。
- C2 は中庸な評価から始めるが途中で容認度が低くなり、一旦元に戻るが、最後にすべてを容認しなくなる。
- C3 は中庸な評価から始めるが、途中で甘くなり、また元に戻る。
- これらの特性を持とに生データを事後補正できる可能性がある

まとめと展望

- 本発表の概要
 - 従来の容認度判定を、多値反応 $A = E \times C \times R$ としてモデル化し、
 - 提案の有効性を実データで確認
- 生データを事後補正できる可能性が示唆された
- 今後
 - 容認度評定を、社会調査として実施する

ARDJ の宣伝

- 日本語の容認度評定データベース
(Acceptability Rating Data for Japanese: ARDJ) を構築します
 - 言語研究者の容認度評定力の認証システムの試作 : 容認度評定データベースを基礎にして (挑戦的萌芽研究)
 - 研究代表者: 黒田 航; 期間: 2016–2018年度; 課題番号: 16K13223
 - 参加者 (7名)
 - 浅尾 仁彦 (NICT), 阿部 慶賀 (岐阜聖徳学園大学), 金丸 敏幸 (京都大学), 小林 雄一郎 (東洋大学), 田川 拓海 (筑波大学), 土屋 智行 (九州大学), 横野 光 (富士通研究所)
 - ARDJ は標準化のための調査で,
 - 容認度評定が集合知であると前提にし
 - 刺激文の評定値の平均値をばらつきの大きさを知る
 - 事を目的にしています

References

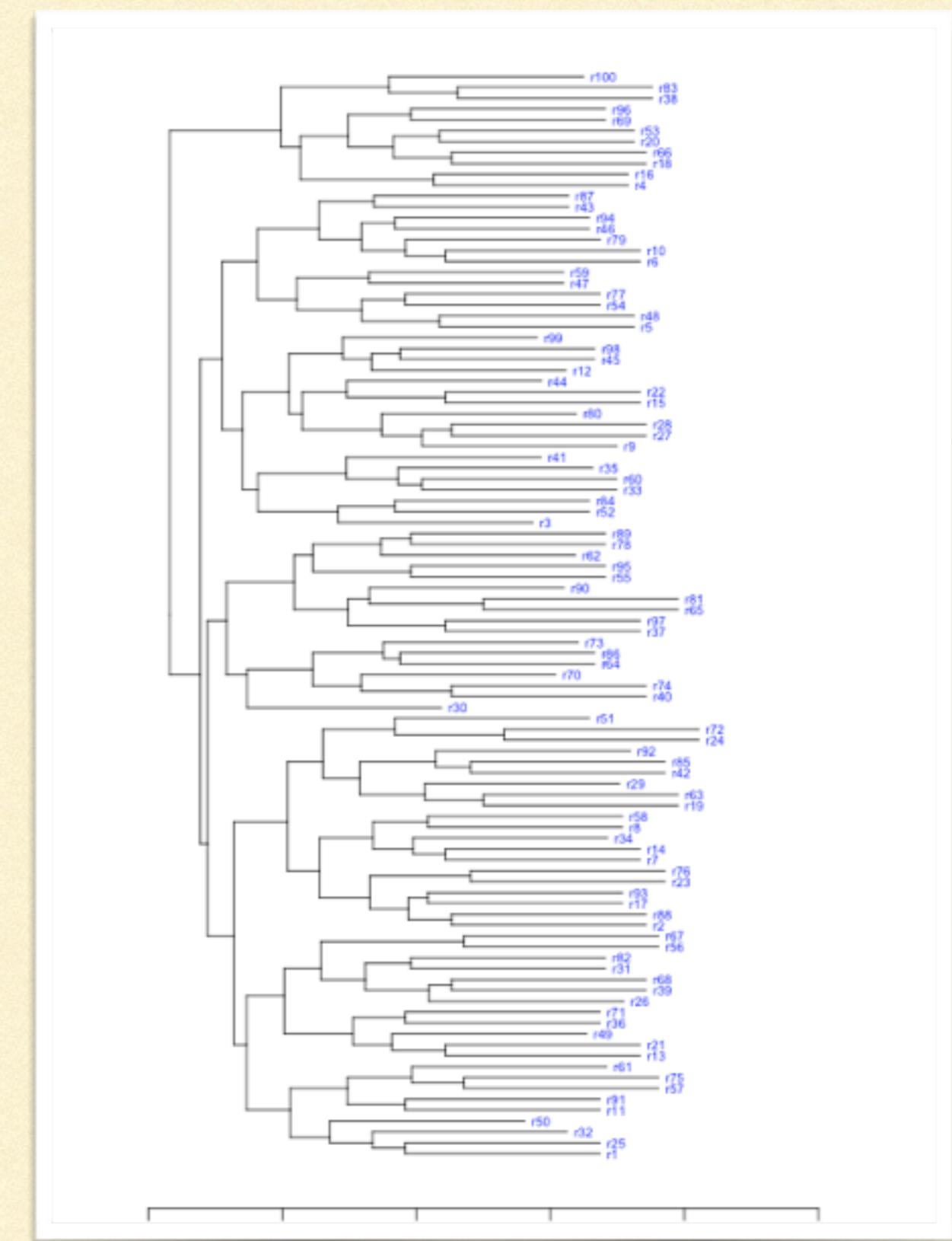
- 河原 大輔, 黒橋 祐夫 (2006). 高性能計算環境を用いたWebからの大規模格フレーム構築. In 情報処理学会研究報告, Vol. 2006-NL-171, pp. 67–73.
- 仲村 哲明, 河原 大輔 (2016). 集合知を用いた事態参与者の特徴変化に関する知識の獲得. In 言語処理学会第22回年次大会発表論文集, pp. 901–904.
- S. E. Embretson and S. P. Reise (2000). *Item Response Theory for Psychologists*, 1st edition, Psychology Press, 2000.
- M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2015). “cluster: Cluster analysis basics and extensions,” 2015. R package version: 2.0.4.
- R Core Team (2015). “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, 2015.
- <http://www.R-project.org/>

THANK YOU FOR LISTENING

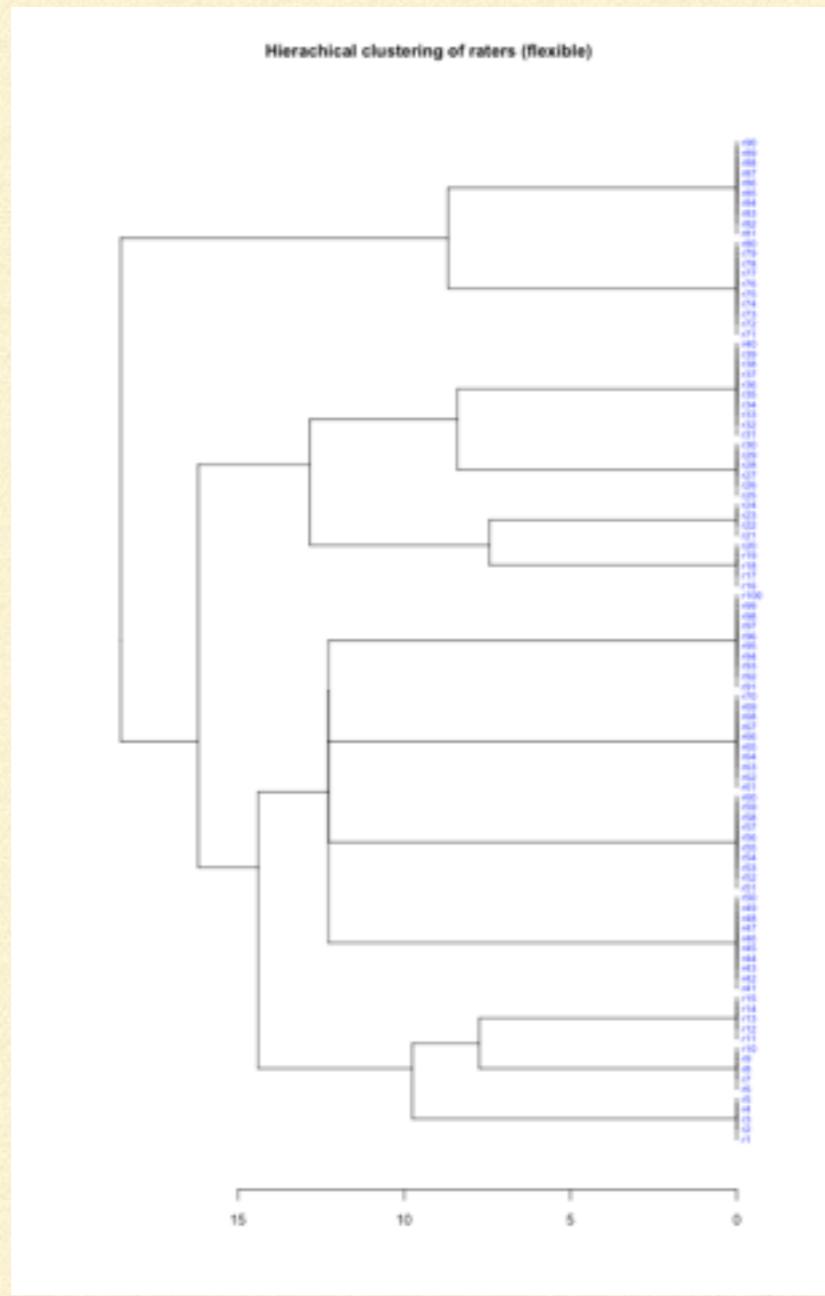
APPENDICES

HOW DAMAGING MISSING VALUES ARE

Appendix A



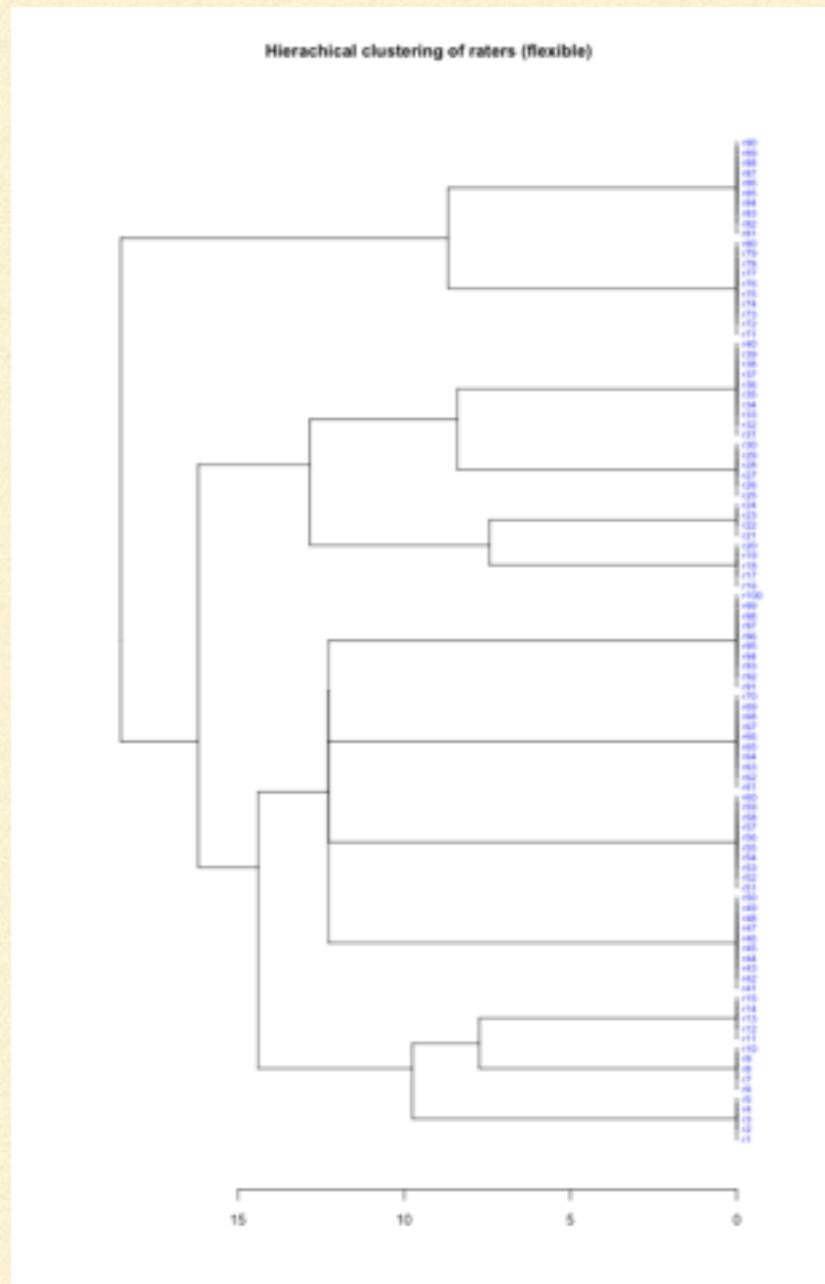
Missing values at 0%



- Artificial data
 - R: r1, r2, ..., r100
 - S: s1, s2, ..., s100
 - Generated C1, C2, ..., C13 manually

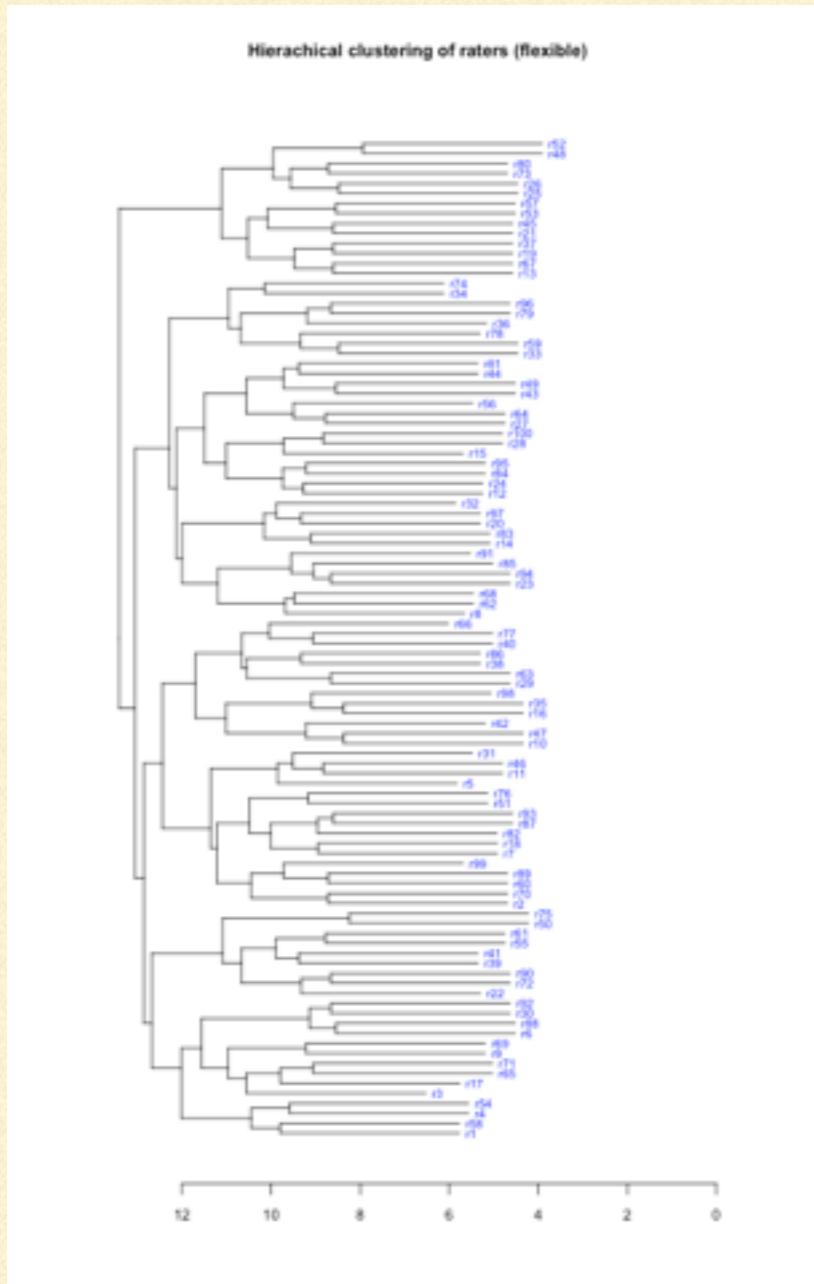
STAT	R.WISE VALUE
AVERAGE	12.60
MEDIAN	10.00
STDEV	4.41
MAX	20.00
MIN	10.00

Randomly generated missing values at 70%



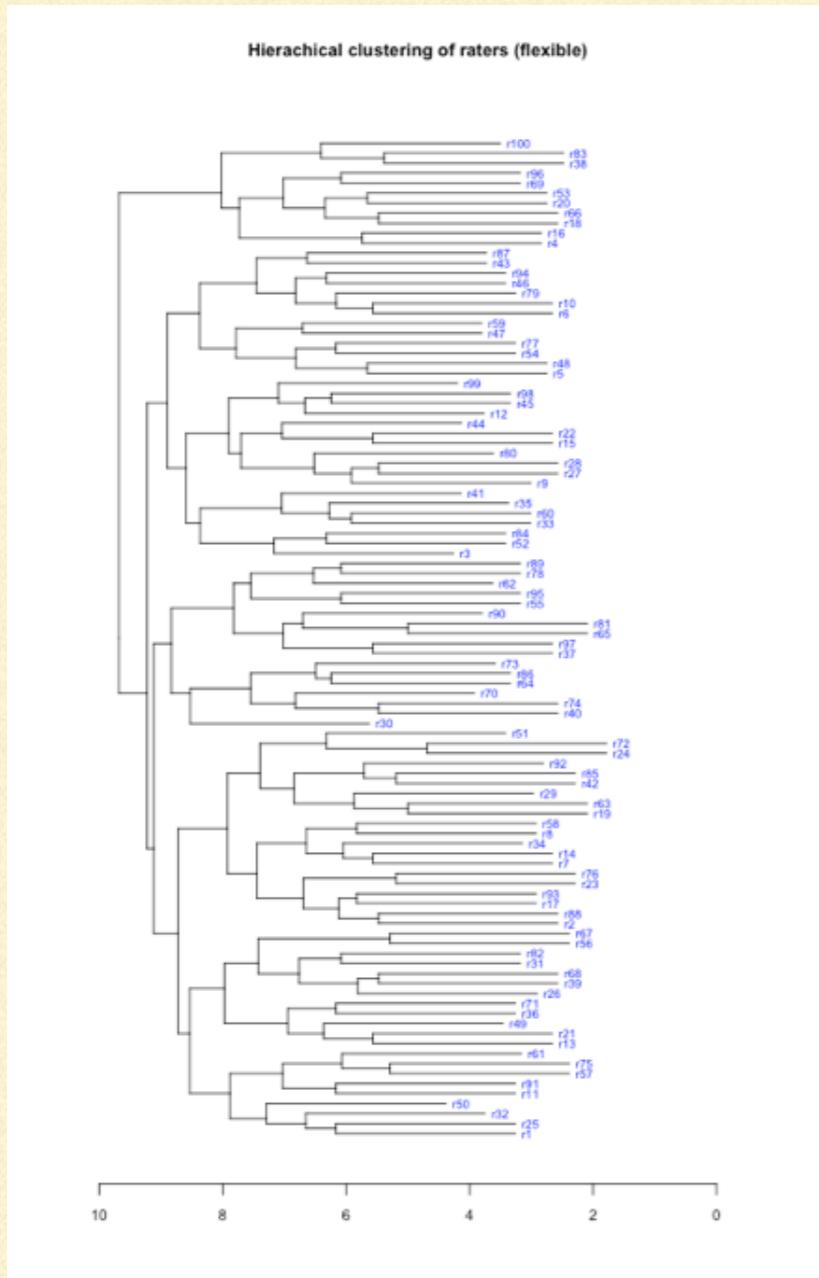
- No effects are observable

Randomly generated missing values at 80%



- Minor effects appear on leaves.
- But this retains grouping at larger scale well.

Randomly generated missing values at 90%



- More damages appear on leaves.
- But this still retains the overall grouping.