How (~~normal~~ ordinary) people respond to (potentially deviant) sentences is not as simple as linguists thought to be

An initial report on Acceptability Rating Database of Japanese (ARDJ) project

Kow Kuroda (Medical School, Kyorin University)

at Joint Workshop of Linguistics and Language Processing-23, 2017-12-16, Waseda University

Introductory Note

- These slides are the slightly modified version of the slides I actually used in JWLLP-23.

Acknowledgements

- I thank for my collaborators
  - Keiga Abe (Gifu Shotoku gakuin), Yoshihiko Asao (NICT), Toshiyuki Kanamaru (Kyoto University), Yuichiro Kobayashi (Nihon University), Takumi Tagawa (Tsukuba University), Tomoyuki Tsuchiya (Kyushu University)

- and
  - Shunji Awazu (Toyo University), who kindly helped us to conduct a survey in Tokyo area.

Motivating Questions 1/2

- What is acceptability?
  - (Too) many people say (too) many different things about it.
    - but they tend to confuse definitions with descriptions.
- What factors affect acceptability?
  - Some people say there are (too) many factors that affect acceptability, thereby suggesting that there aren't such things like "unacceptable sentences".

Motivating Questions 2/2

- As far as I can see, nobody has ever tried to observe acceptability using a large, unbiased set of stimuli in a realistic setting.
  - Virtually, all previous investigations into acceptability were confirmatory.
- How to get out of confirmation bias and obtain a general (enough) theory of acceptability?

Ultimate Motivation for ARDJ

- To answer the last question, we need a reference behavioral data, which should allow us to pose questions like:
  - "How is acceptability "shaped" in a multi-dimensional possibility space?"
- What do we need?
  - We first need to try to describe acceptabilities as they are, before trying to explain them.
  - Our work is done in the case of Japanese, but the adopted methodology is expected to be applicable to any other languages without drastic modification.
- This talk is an initial report of a pilot study we ran this summer.

Design features of ARDJ

- Acceptability Rating Data of Japanese (ARDJ) is a work in progress to provide such a reference data, with the following features:
  - Requirement 1. using as unbiased stimuli as possible
  - Requirement 2. finer-grained, rather than simple acceptable or not
  - Requirement 3. larger-scale
- They are explained briefly below, in turn.

Requirement 1: Unbiasedness

- Motto: Go exploratory to avoid the two modes of confirmation bias.
- We need to do away with two kinds of bias that affected virtually all previous studies of acceptability.
  - First, acceptability ratings by linguists are biased and unreliable.
    - All categorizations by professionals, i.e., stake-holders, are theory-laden, and therefore inevitabley biased.
  - Second, used stimuli are unsystematically designed and too opportunistic.

Requirement 2: Finer-grainedness

- Motto: Go finer-grained and more realistic to overcome underanalysis.
  - Simple acceptable/unacceptable dichotomy isn't enough to see what acceptability really is.
- ARDJ uses 4-point scale for resolution:
  - (0) perfectly fine, (1) deviant but easy to understand, (2) deviant and hard to understand, (3) incomprehensible
- This follows the format of forced choice task and is a finer-grained version of YES/NO choice.

Requirement 3: Large(r)-scaledness

- Motto: Go better-grounded to overcome small number effects.
- Judgements by a few people are just too unreliable to generalize properly, even if they are experts.
- Desirably, we could have 100's of responses to every single stimulus.

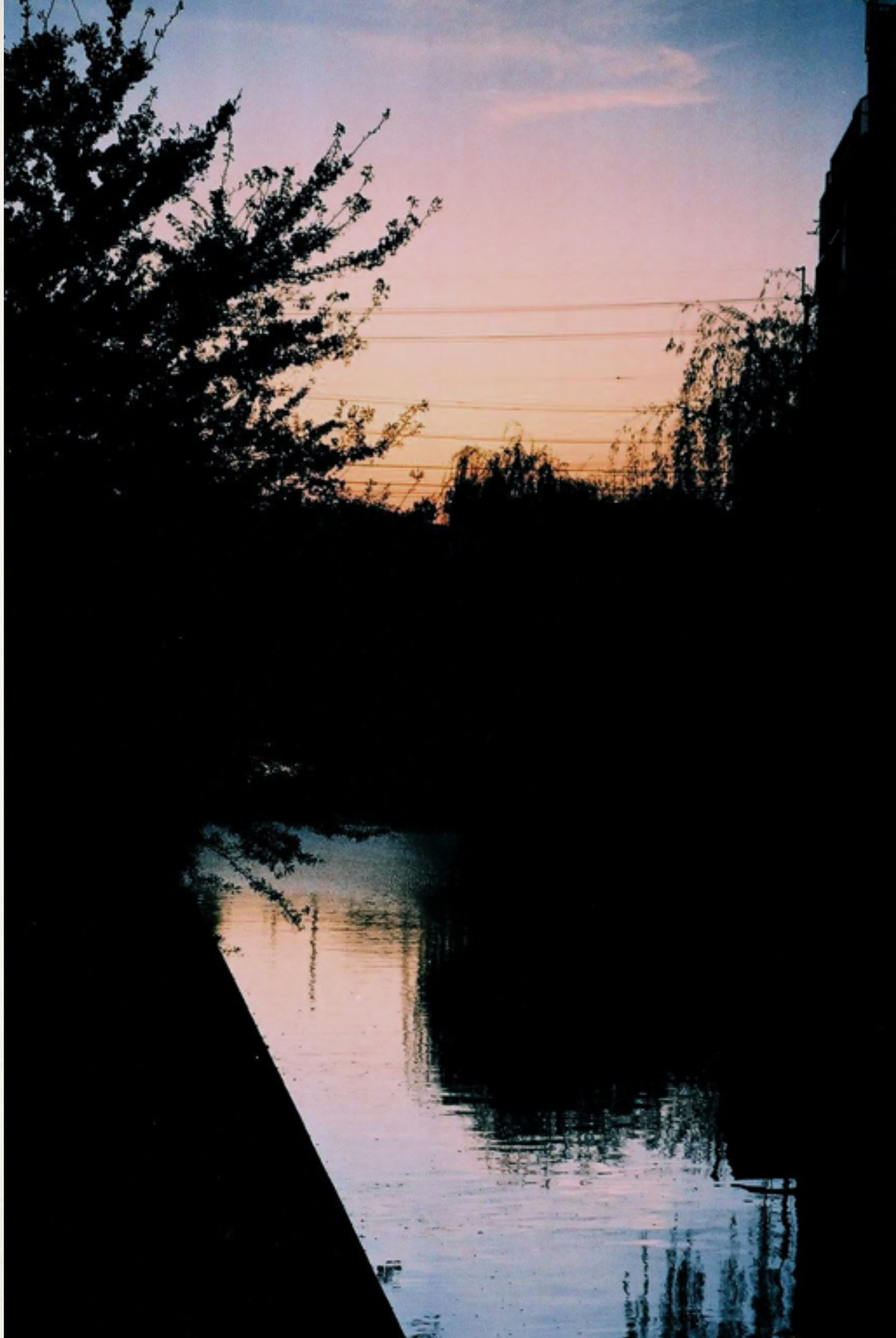
What to talk about today

- Data preparation
  - especially on how (potentially) deviant sentences were generated nearly automatically
- Experiment
- Analysis/Results and Discussion
- Future plans

Findings or take-home messages

- Main claim: Normal people do not simply classify stimuli into dichotomy: deviants or non-deviants.
- Rather, their responses have three classes:
  - normal stimuli (= without deviance) and two subtypes of deviant stimuli, though we are not yet able to characterze their diffrence.
- To my suprise, simple classification of responses by mutation types revealed no differentiation.

Data preparation



How to systematically generate deviant sentences near-automatically? 1/2

- For stimuli for acceptability rating, we need:
  - as many sentences as possible of as many kinds as possible,
  - and to generate them in least biased ways.
- Consequences
  - human-generation is a misfit for systematic generation of stimuli.

How to systematically generate deviant sentences near-automatically? 2/2

- Solution to this
  - Semi-automatic generation of (potentially) deviant sentences from normal instances, i.e., seeds
- Implementation via mutation-like random modifications with the following steps
  - 1. Human (linguists or not) construct normal or subnormal sentences (originals) *O*, for "seed" sentences.
  - 2. Randomly apply mutation of specific kinds to instances of *O* and get their modifications *M*.
  - 3. Filter *M*, if necessary.

Procedure

- *O* were constructed by:
  - 1. selecting 9 verbs from NINJAL-LWP (for BCCWJ), and
  - 2. lexcially filling 4 pre-specifed patterns/constructions whose verbs are one of the 9 verbs above
- *M* were constructed from *O* by applying
  - one of the 3 types of lexical replacement, or
  - phrasal displacement of a randomly selected pair of nominal phrases (bunsetsu)

9 Verbs Selected

- ID correspond to the frequencies of NINJAL-LWP data.
  - 22: *iku* (行く) [go]
  - 26: *shiru* (知る) [know]
  - 44: *kan-jiru* (感じる) [feel]
  - 116: *kotae-ru* (答える) [answer]
  - 326: *damaru* (黙る) [be(come) silent]
  - 338: *makeru* (負ける) [lose]
  - 377: *tutawaru* (伝わる) [carry, propagate, get through]
  - 1147: *shiri+au* (知り+合う: VV-compound) [know each other]
  - 1197: *kansen+suru* (感染+する: NV-compound) [acquire, contract, catch, develop (a disease)]
- I have very much to say about how I selected these verbs, but I don't have enough time.

4 Constructions/Patterns used

- P1: _-*ga* _-*de* _-*ni* _-*to* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Goal/Place + Committive/Manner + V
  - example: s111: Douryoo-ga shitumon-de aite-ni ina-to kotae-ta.
- P2: _-*ga* _-*de* _-*ni* _-*wo* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Goal/Place + Object/Result + V
  - example: s151: Kazokudure-ga shiohigari-de umi-ni kai-wo sagashi-ta.
- P3: _-*ga* _-*de* _-*wo* _-*ni* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Object/Result + Goal/Place + V
  - example: s197: Kanojo-ga tegami-de shinjitu-wo fui-ni shit-ta.
- P4: _-*ga* _-*de* _-*kara* _-*wo* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Source/Material + Object/Result + V
  - example: s71: Horyo-ga jinmon-de chuuseishin-kara himitu-wo damat-ta.
- I have much to say about how I decided on these 4 patterns, but I don't have enough time.

Types of Mutation

- Lexical
  - replace a (randomly) selected word/ morpheme of a targeted POS for a randomly selected similar word of the same POS, and do so recursively
  - Subtypes:
    - n-type: nouns (including adjectival nouns and bases of certain adverbs) are replaced
    - v-type: verbs (excluding auxiliaries) are replaced
    - p-type: case-markers are replaced.
- Positional
  - s-type: exchange the positions of a randomly selected pair of NP (or PP), and do so recursively.
- Note
  - Python scripts for these modification were developed and will be available at GitHub.

	A	B	C	D	E	F	G	H
1	S-ID	Gro	Source	Typ	S	Type2	Note	
2	1	1	0131-p3-3	n	刑事が捜査で手がかりを退職たり所存に探した。	[n-changed]		
3	2	2	1147-p1-3	o	友人が旅先で未来の恋人にちゃっかりと知り合った。	[original]		
4	3	3	0022-p4-1	n	伊勢神宮がバイクで正面から本殿を行った。	[n-changed]		
5	4	4	0377-p1-3	s	電話でメッセージが声と相手に伝わった。	[swapped]		
6	5	5	1147-p1-1	s	職場で夫が妻と真夜中に知り合った。	[swapped]		
7	6	6	1147-p1-4	s	夏に恋人と娘が旅行先で知り合った。	[swapped]		
8	7	7	0338-p2-1	o	その候補が大会で格下相手に楽な試合を負けた。	[original]		
9	8	8	0044-p3-2	v	高校生がデートの場でしらじらしさを恋人に感じ取った。	[v-changed]	p-changedを訂正	
10	9	9	1147-p4-1	s	双子が子供の頃からお互いの好みを生まれつきで知り合	[swapped]		
11	10	0	1147-p1-1	p	夫が職場で真夜中に妻へ知り合った。	[p-changed]		
192	191	1	1197-p1-4	v	コンピュータウイルスが不注意でネットワークに段々と感染	[v-changed]	p-changedを訂正	
193	192	2	0116-p3-3	n	老人が詐欺で通帳番号を迂闊に答えた。	[n-changed]		
194	193	3	1197-p1-3	o	私が遊園地でインフルエンザに家族と感染した。	[original]		
195	194	4	0326-p1-1	s	店員にその売り場で客が文句を言うと黙った。	[swapped]		
196	195	5	0377-p1-3	o	メッセージが電話で相手に声と伝わった。	[original]		
197	196	6	1147-p1-1	n	夫が職場で毎晩に妻と知り合った。	[n-changed]		
198	197	7	0026-p3-2	p	彼女が手紙で真実が不意に知った。	[p-changed]		
199	198	8	0040-p4-2	s	生活の必要から書道を字のうまい青年が外国で教えた。	[swapped]		
200	199	9	1197-p1-4	o	コンピュータウイルスが不注意でネットワークに段々と感染	[original]		
201	200	0	0326-p1-1	o	客がその売り場で店員に文句を言うと黙った。	[original]		
202								

Samples of 200 stimuli

Data preparation in summary

- 33 originals were selected.
- 167 mutations were generted via either n-, v-, p- or s-type modification, so that we had 200 sentences in total.
  - The size of stimuli, 200, was determined on the number of available participants.

Type	Count	Ratio
o	33	0.165
n	49	0.245
v	29	0.145
p	36	0.180
s	53	0.265
sum	200	1.00

Experiment

Partition and Randomization

- 200 stimlulus sentences were divided into 10 groups of 20 sentences via random partitioning, thereby giving gr0, gr1, …, gr9.

- 4 different versions, A, B, C and D, were constructed from each group, so that the stimuli appeared in four different orders
  - This was done to reduce order effects.

- We thus had 40 cases of materials: gr0-A, gr0-B, gr0-C, gr0-D, gr1-A, …, gr9-A, gr9-B, gr9-C, gr9-D. They were randomly assigned to participants.

Task

- Part 1: 10 questions for social attributes
- Part 2: Rating task (main)
  - Instructions: “This is a task to rate the naturalness of sentences below. To each of the 20 sentences given, give a rating for comprehensibility on the following scale from 1.”
    - 1. perfectly fine (degree 0 deviance)
    - 2. feel somewhat awkward but find no trouble understanding (degree 1 deviance)
    - 3. feel rather awkward and find trouble understanding (degree 2 deviance)
    - 4. completely incomprehensible (degree 3 deviance)
- Note
  - We didn’t ask if a particular sentence is acceptable or not, which turned out to be both question-begging and under-specified.

2　文の自然さを評価する課題

以下に示す 2.1–2.20 のそれぞれの文をどう感じるかを答えて下さい．1, 2, 3, 4 のいずかの数字を選んで○で囲んで下さい．

2.1　その候補が大会で格下相手で楽な試合を負けた。

1. 違和感なし　**2.** 違和感を感じるが理解可能　**3.** 違和感を感じて理解困難　**4.** 不自然な文で理解不能

2.2　伊勢神宮がバイクで正面から本殿を行った。

1. 違和感なし　**2.** 違和感を感じるが理解可能　**3.** 違和感を感じて理解困難　**4.** 不自然な文で理解不能

2.3　私が遊園地でインフルエンザに家族と感染した。

1. 違和感なし　**2.** 違和感を感じるが理解可能　**3.** 違和感を感じて理解困難　**4.** 不自然な文で理解不能

2.4　波紋が投石で落下点から水面を見て取れた。

1. 違和感なし　**2.** 違和感を感じるが理解可能　**3.** 違和感を感じて理解困難　**4.** 不自然な文で理解不能

2.5　得意種目を有名選手が大事な試合で無名選手に負けた。

1. 違和感なし　**2.** 違和感を感じるが理解可能　**3.** 違和感を感じて理解困難　**4.** 不自然な文で理解不能

2.6　優勝候補がトーナメントで初戦を突然に敗れた。

1. 違和感なし　**2.** 違和感を感じるが理解可能　**3.** 違和感を感じて理解困難　**4.** 不自然な文で理解不能

2.7　恋人と娘が旅行先で夏に知り合った。

10 attribute collected as social attriutes (to be used in future)

- 1) age [number]
- 2) sex/gender [f/m/other]
- 3) native place [number encoding prefecture]
- 4) if Japanese is the rater's mother tongue [Yes/No]
- 5) if the rather has lived in other country or countries for more than one year [Yes/No]
- 6) the number of foreign languages the rater has learned [number]
- 7) the total length of foreign language learning [number]
- 8) if the rater has a frequent contact with non-Japanese speaking foreigners [Yes/No/Don't know]
- 9) the number of books read in a month [number]
- 10) years of education [number].

1.4　母語

あなたの母語は日本語ですか？ 右の空欄に 1, 2, 3 のいずかの数字を記入して下さい.

1. はい　2. いいえ　3. わからない

1.5　海外生活経験

過去に 1 年以上の非日本語圏での生活経験が過去にありましたか？ 右の空欄に 1, 2, 3 のいずかの数字を記入して下さい.

1. はい　2. いいえ　3. どちらとも言えない

1.6　外国語学習経験 (言語数)

学んだ事のある外国語の個数を，おおよその数で答えて下さい．右の箱に 2 桁の数字を記入して下さい．100 個以上の場合は 99 として下さい.

1.7　外国語学習経験 (学習期間)

上の外国語を学んだ期間の合計を，おおよその年数で答えて下さい．右の箱に 2 桁の数字を記入して下さい.

1.8　現在の日常的に異国語を話す人との接触の程度

異国語を話す人と日常的に接しますか？ 右の空欄に 1, 2, 3 のいずかの数字を記入して下さい.

1. はい　2. いいえ　3. どちらとも言えない

Participants

- Experiments were conducted in three places: Tokyo, Gifu and Fukuoka/Kyushu.
- We had 251 participants (93 participants in Tokyo, 109 in Gifu, and 49 in Fukuoka/Kyushu).
  - All participants were students of university or college.
- We had effective 216 responses in total.
  - 35 responses were removed after validating all the responses.
- Note: some corrections to our subumission to LREC 2018.

Some statistics of raters/participants

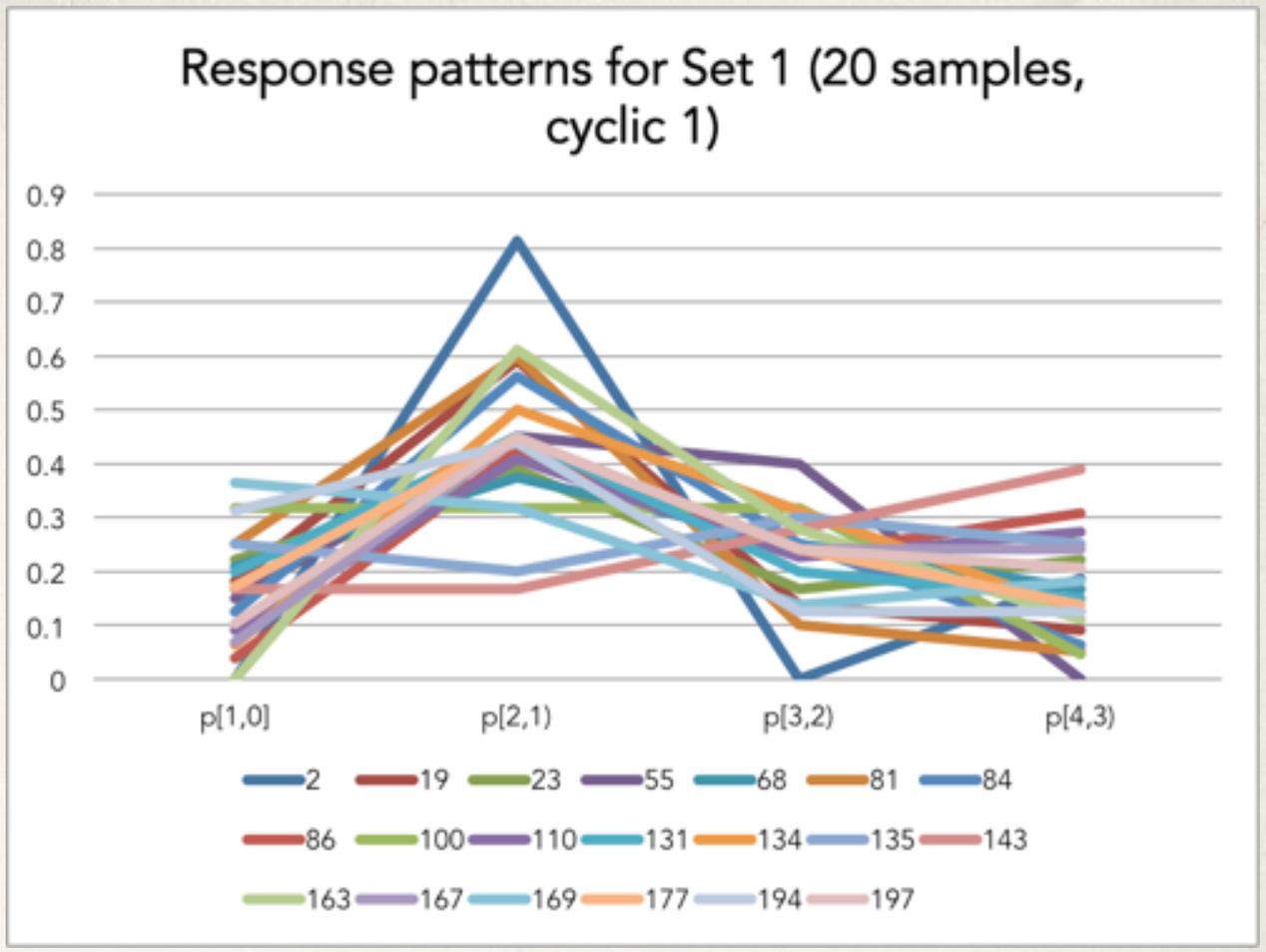
gr.ID	ratio	sum	tokyo	gifu	fukuoka
0	0.11	24	8	12	4
1	0.09	20	4	11	5
2	0.07	16	3	9	4
3	0.08	18	3	10	5
4	0.07	16	1	12	3
5	0.09	20	5	10	5
6	0.12	26	13	8	5
7	0.13	29	15	9	5
8	0.12	25	9	11	5
9	0.10	22	10	9	3
sum	1.00	216	71	101	44

Analysis and Results

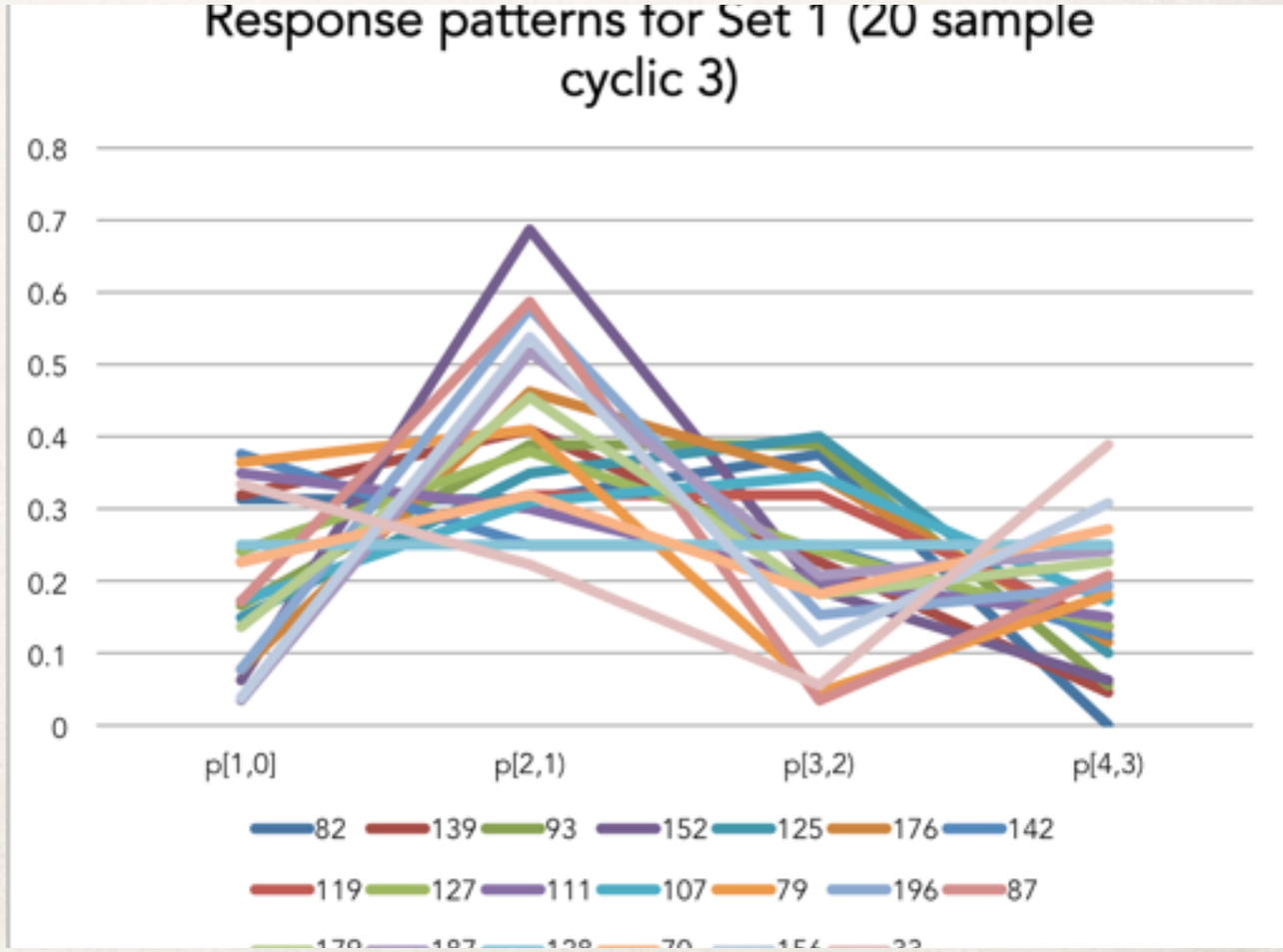
Encoding and Analysis

- Encoding
  - For each stimuli, probability distribution over four value ranges, [1,0), [2,1), [3,2) and [4,3), was calculated, and used as encoding vector.
- Tried analyses
  - Hierarchical clustering
  - *k*-means clustering
    - We tried cases of *k*=4, 5, 6 and 7.

Two instances of random sampling (n=20)

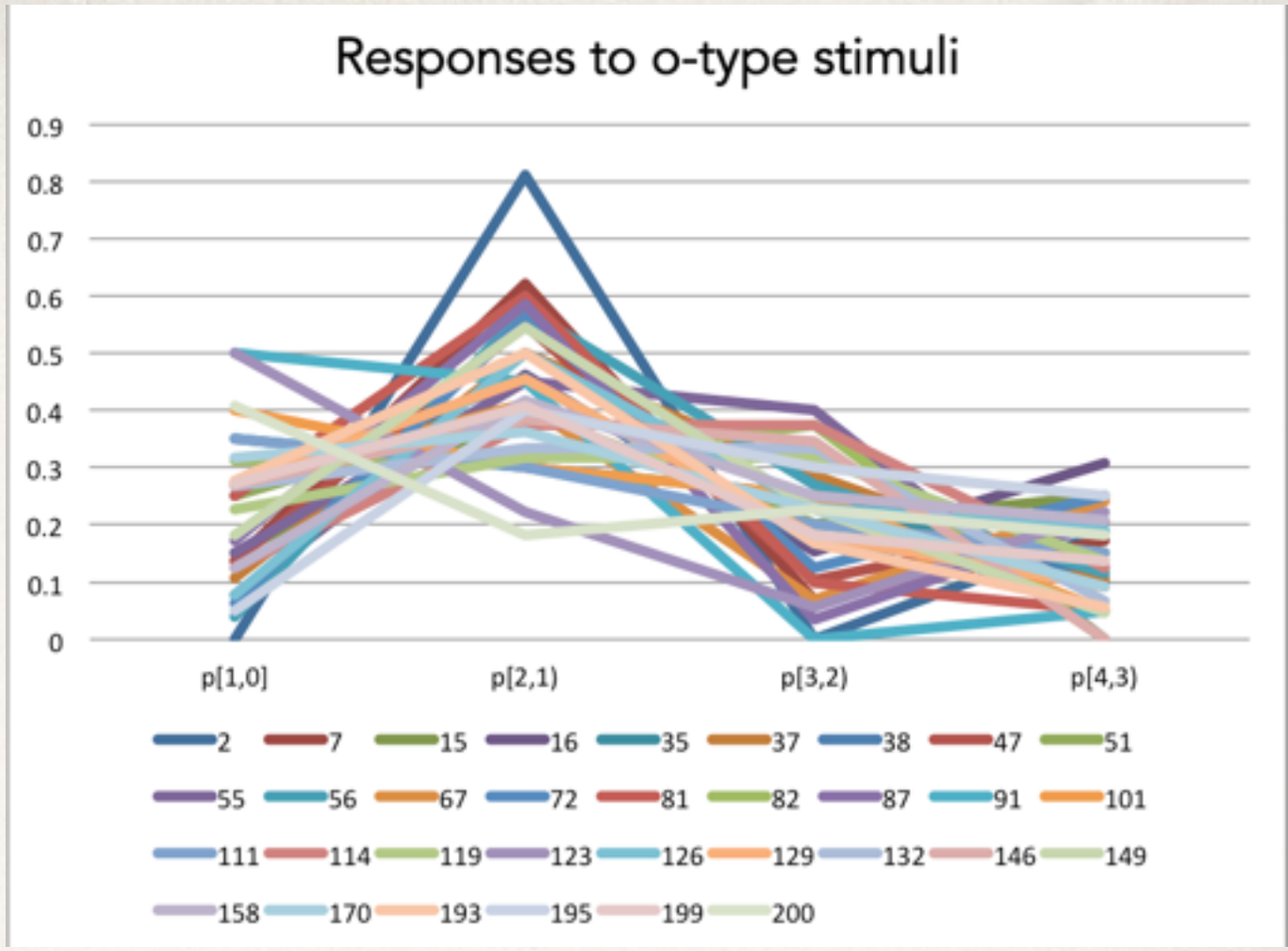


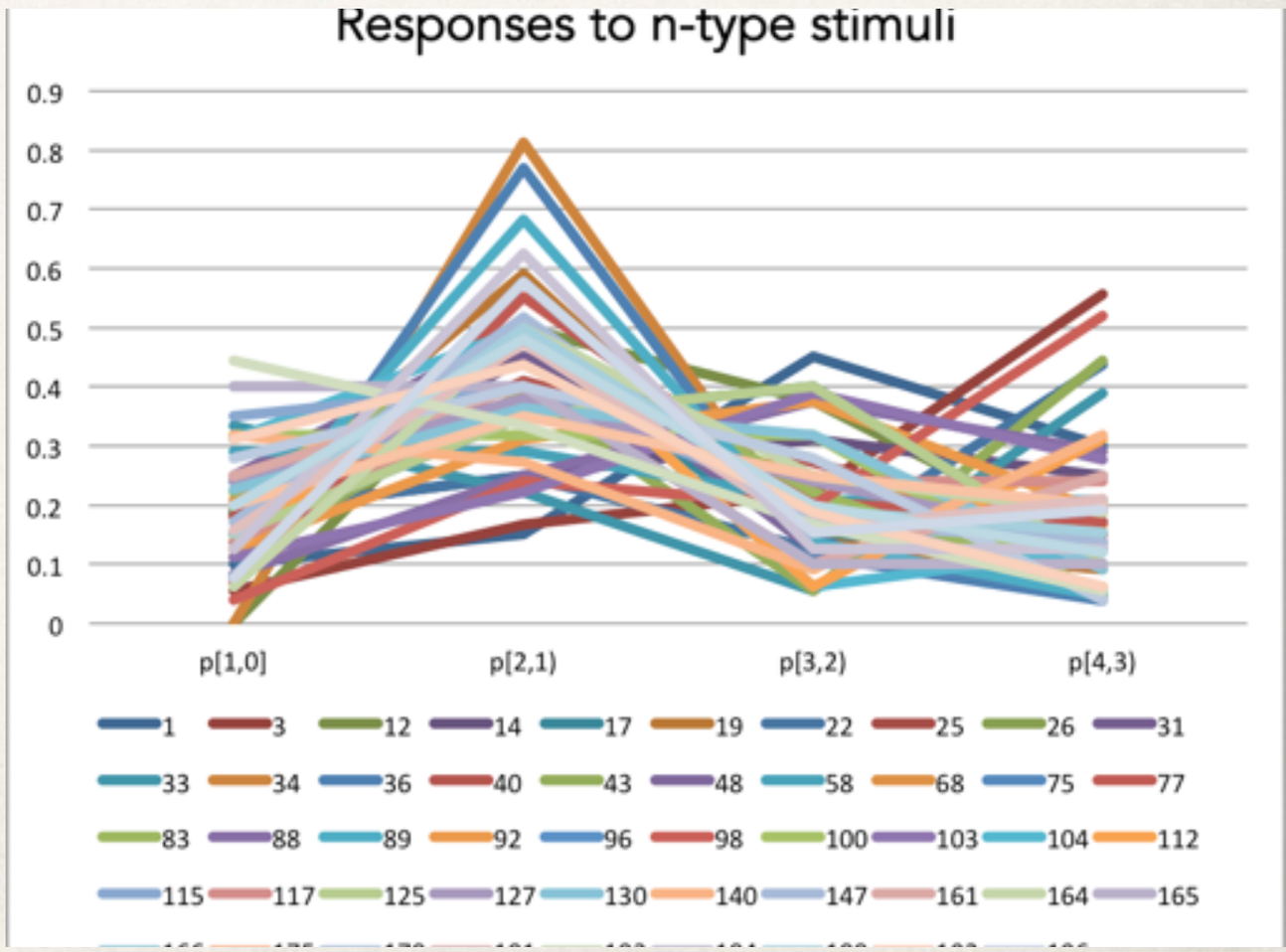
- They look the same.
  - This was what we expected. (Yes!)



Comparison by edit type 1/5

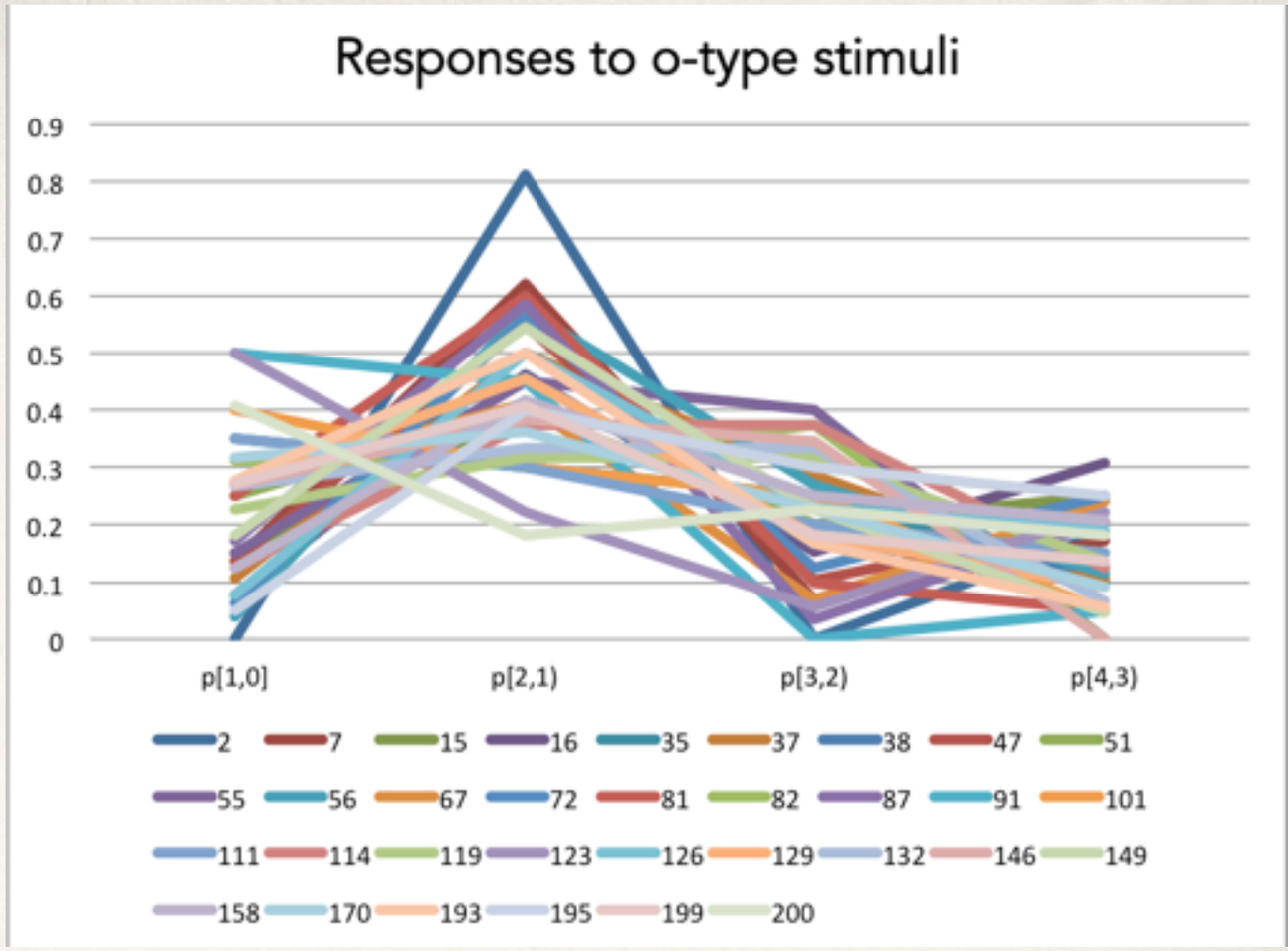
- Comparison of originals (above) and nominal changes.
- They look the same.
  - This was what we didn't expect.

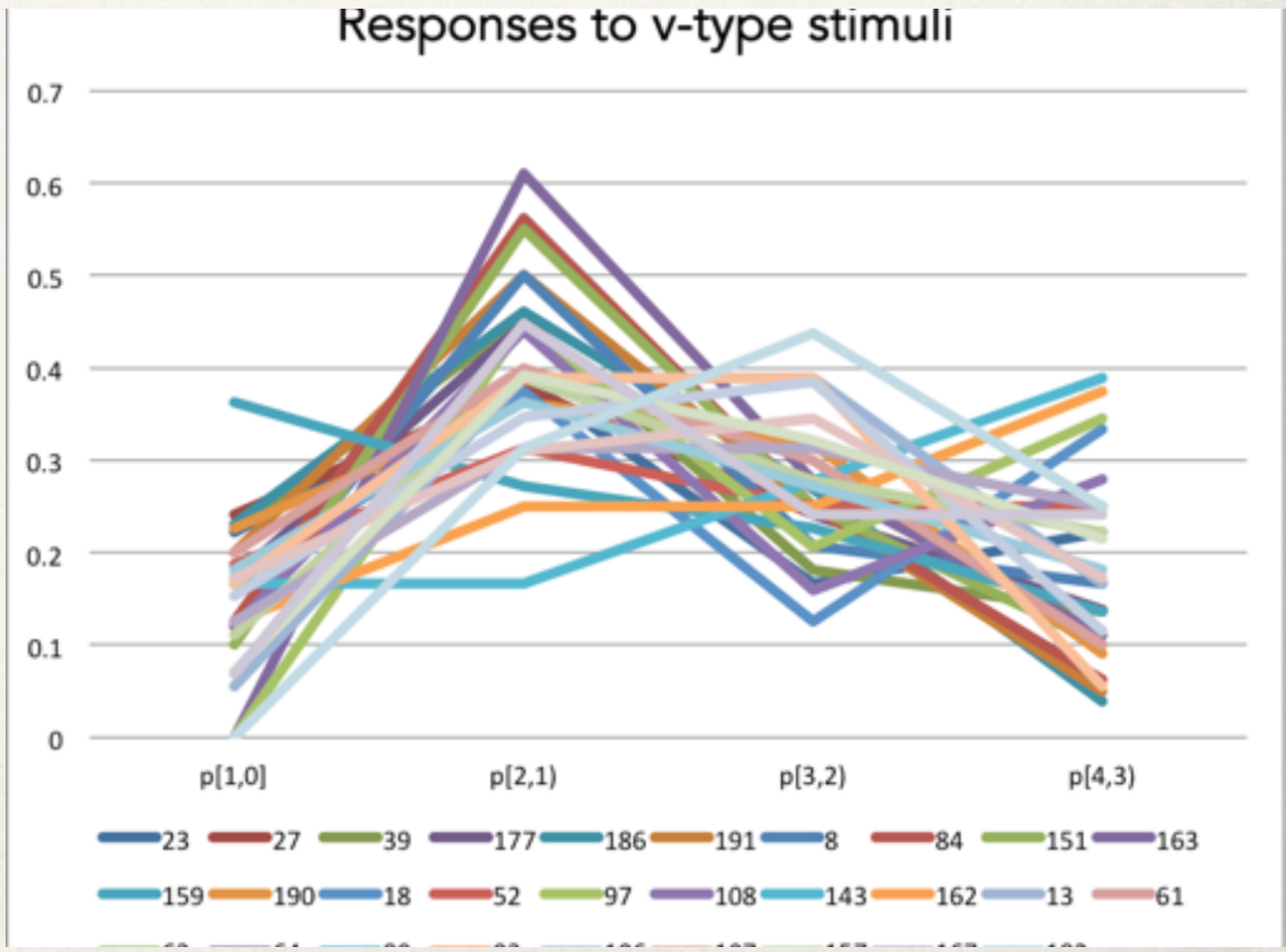




Comparison by edit type 2/5

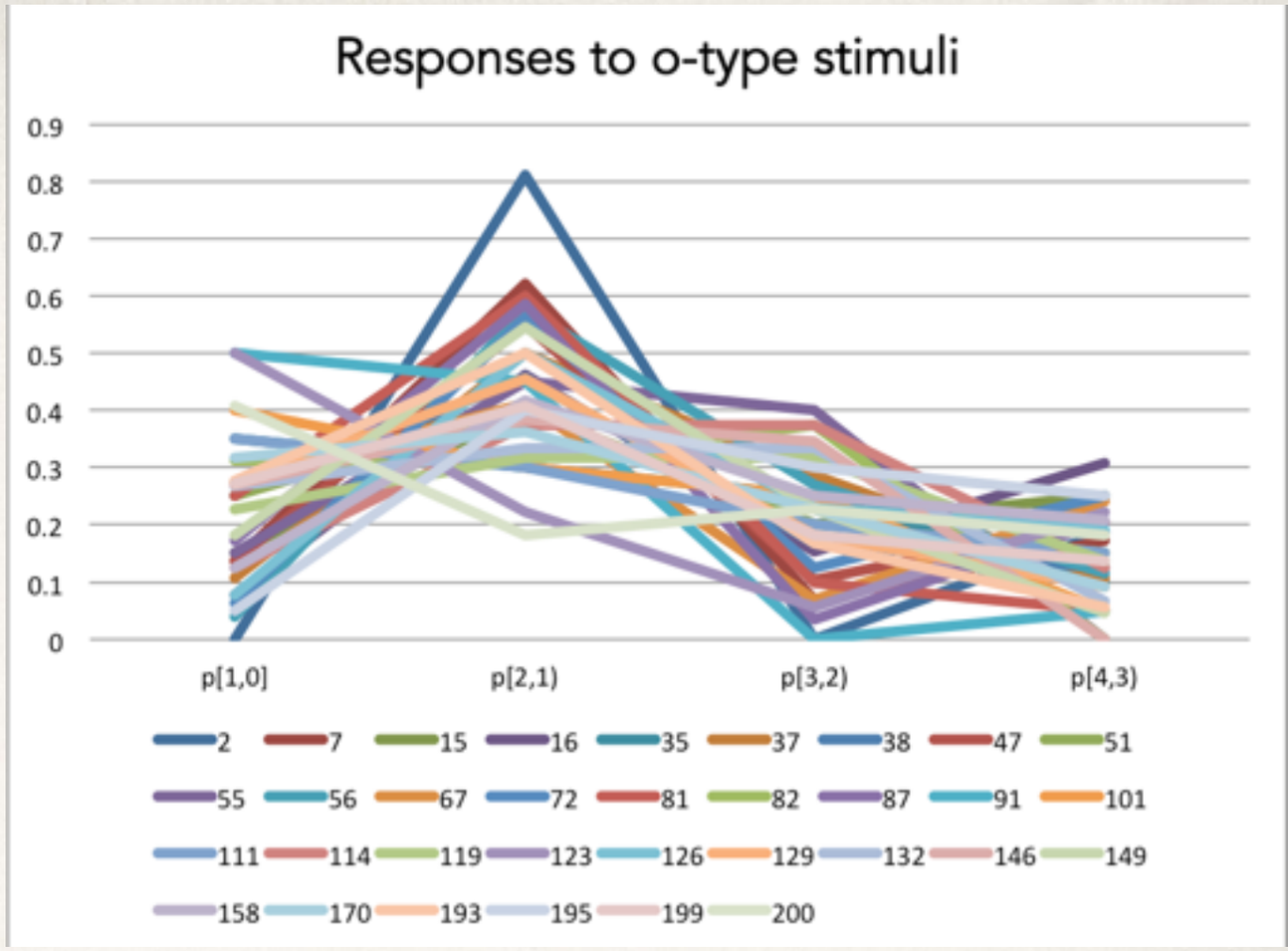
- Comparision of originals (above) and verbal changes
- Again, they look the same.
  - This was what we didn't expect.

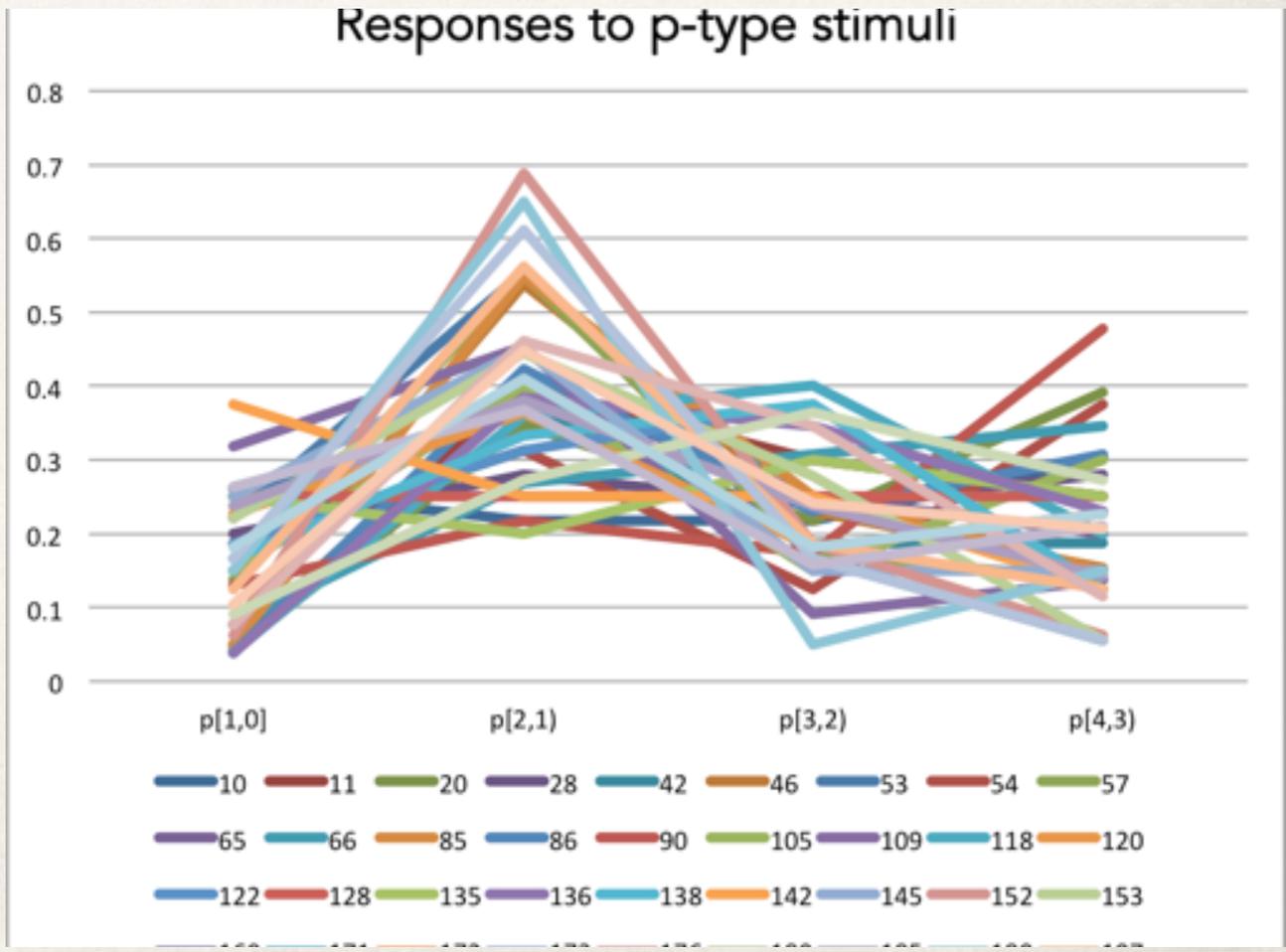




Comparison by edit type 3/5

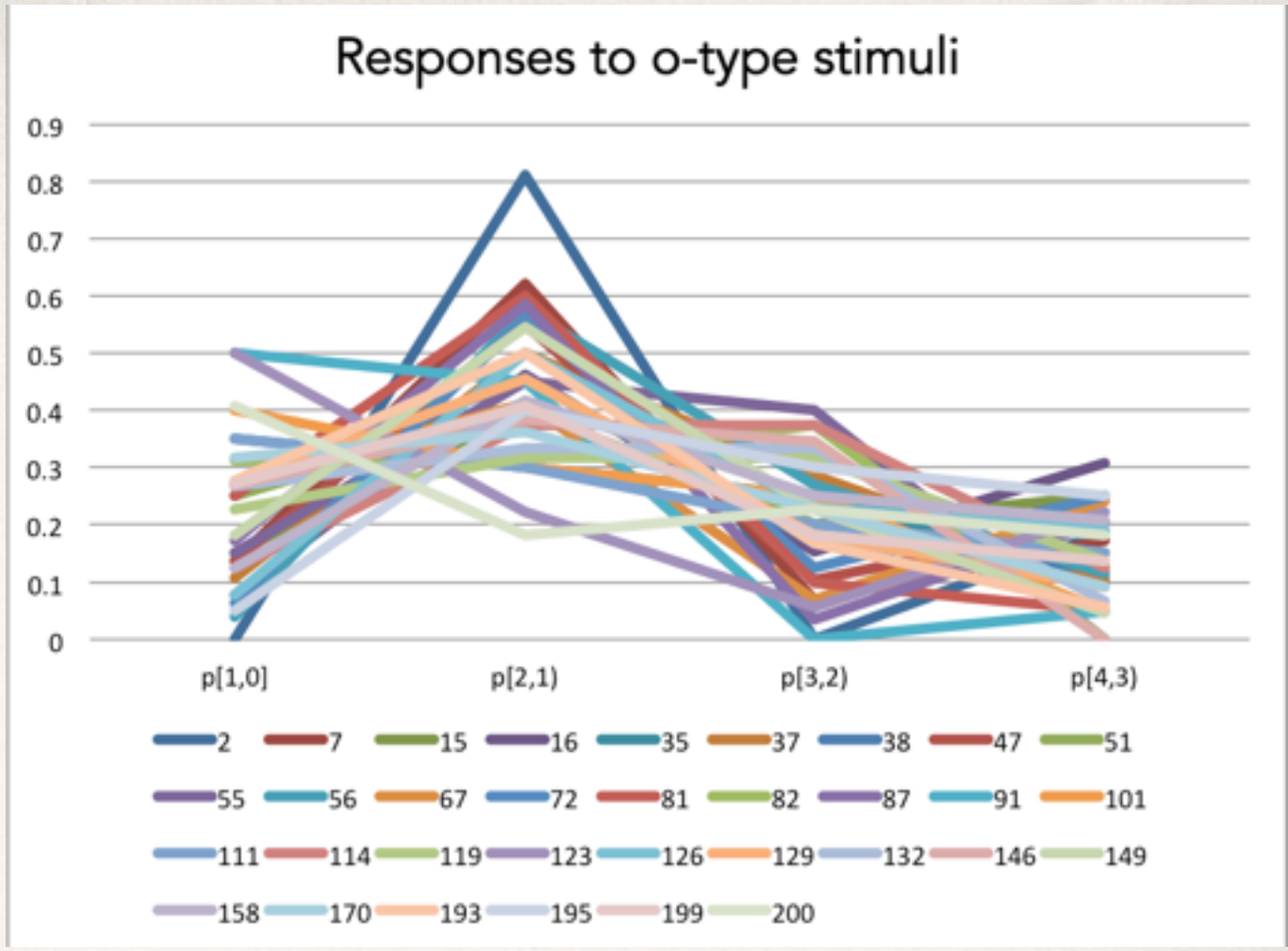
- Comparison of originals (above) and case-marker changes
- Again, they look the same.
  - This was what we didn't expect.

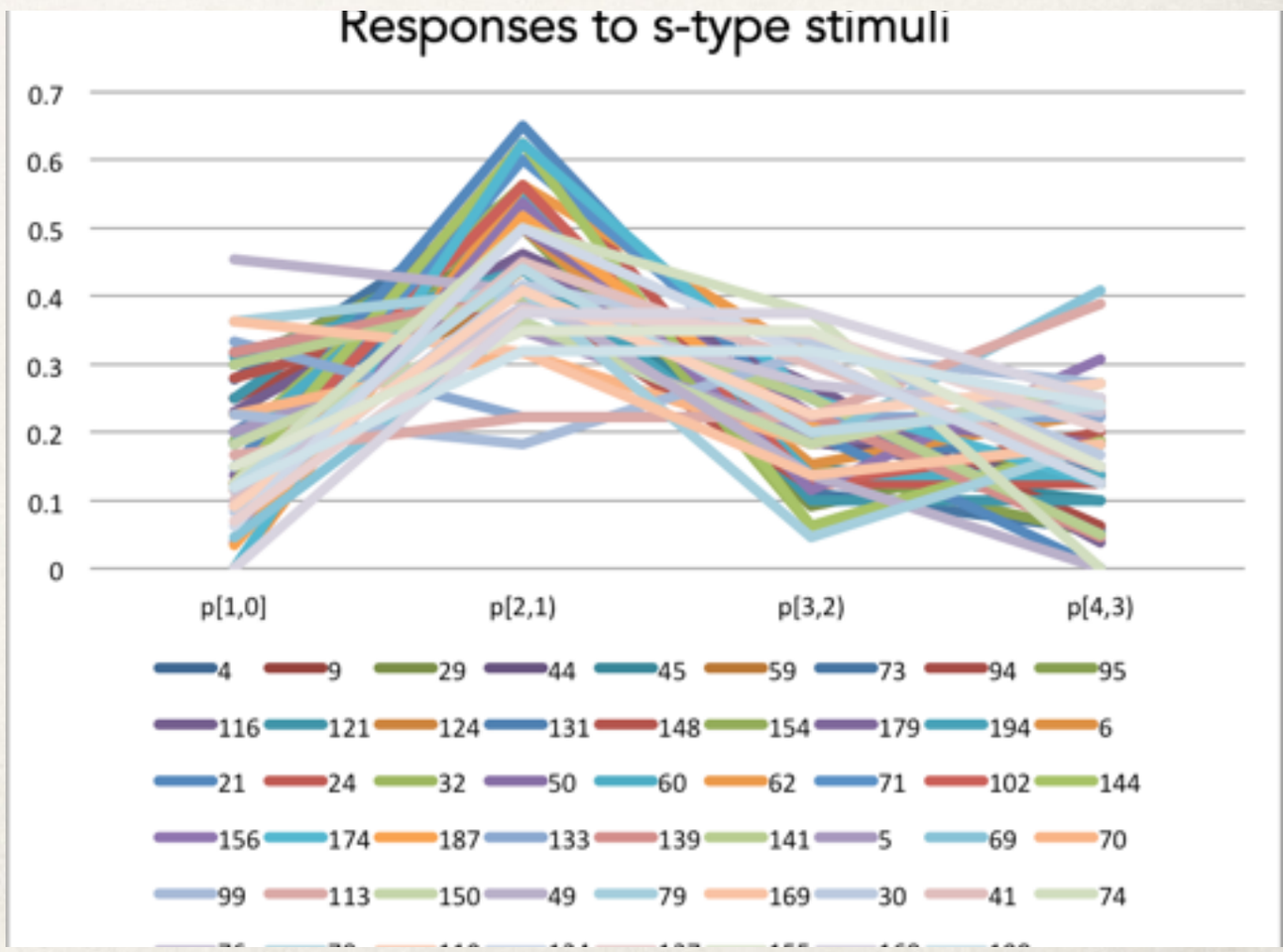




Comparison by edit type 4/6

- original (above) and swapping/ phrasal dislocation
- All over again, they look the same.
  - This was what we didn't expect.





Comparison by edit type

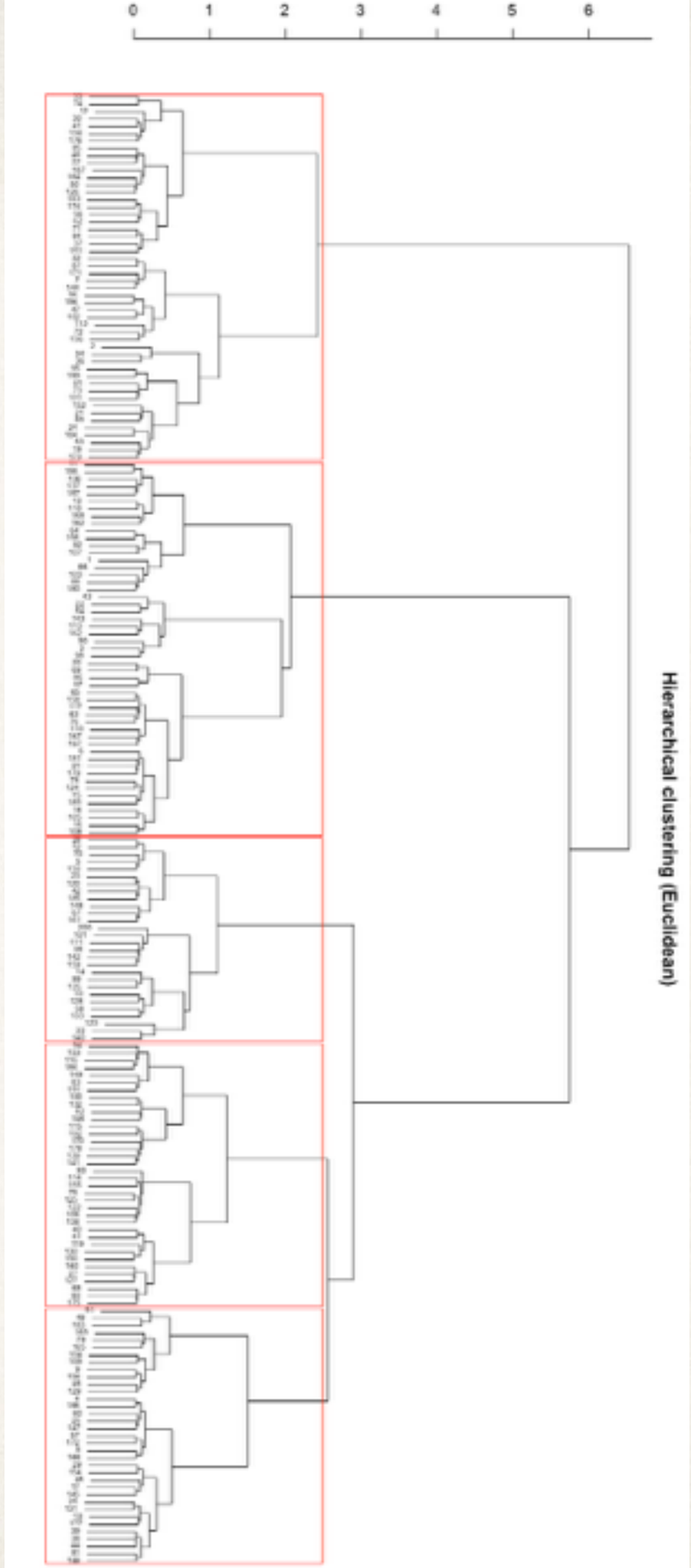
- Overall, no significant differences were recongized among different types of edit/mutation.
- This result was totally unexected, honestly.
  - I expected otherwise.

New Questions and Directions

- Questions
  - Q1. Are there qualitatively different subsets of responses?
  - Q2. How to identify them, if any?
- Actions
  - Go exploratory!

Hierachical clustering

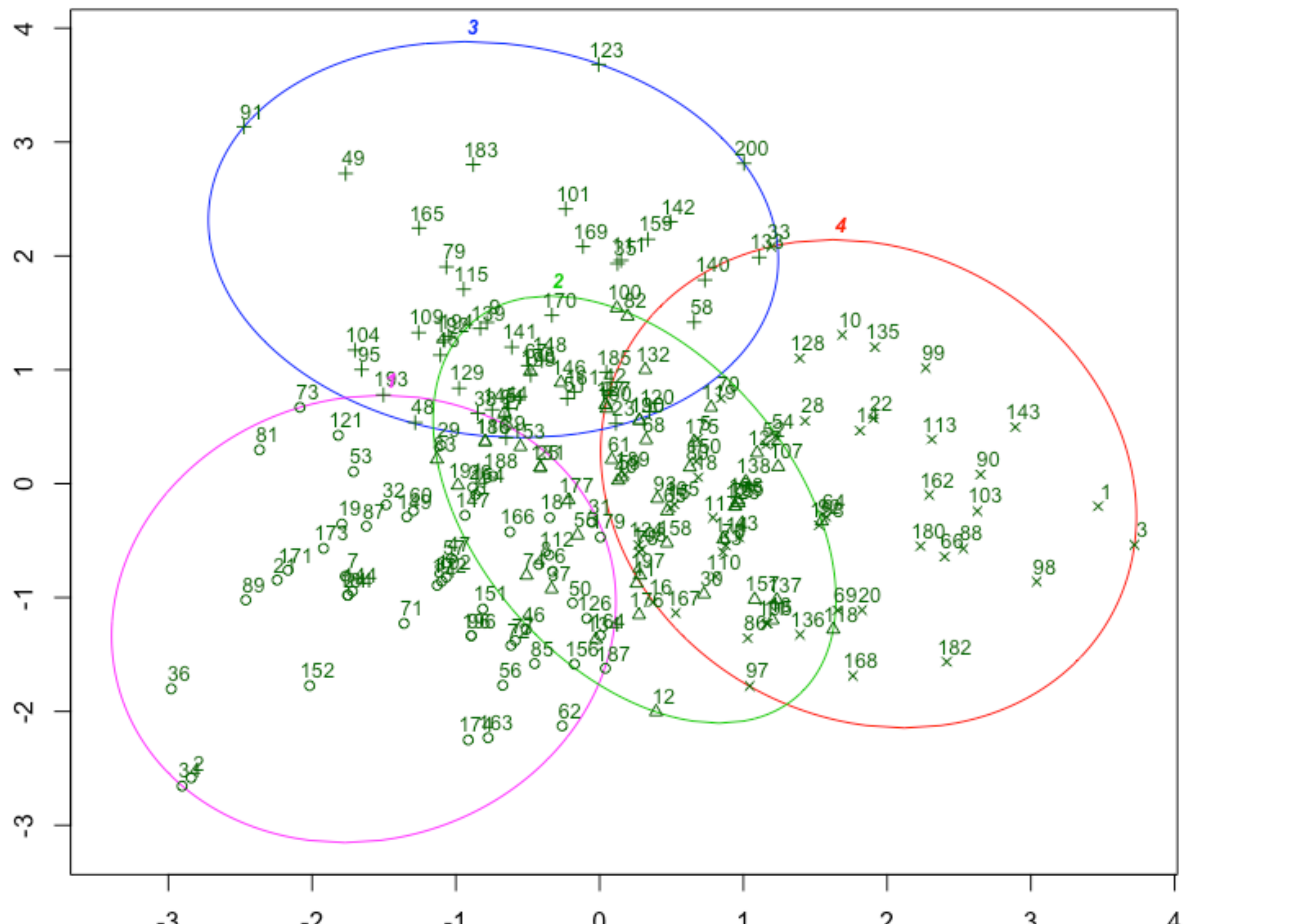
- All responses (213) are clustered by Ward method (with Euclidean distance).
- There seem to be different response patterns.
- But hierarchical clustering hardly describes how they relate to each other.
- We need another approach.



k-means Clustering

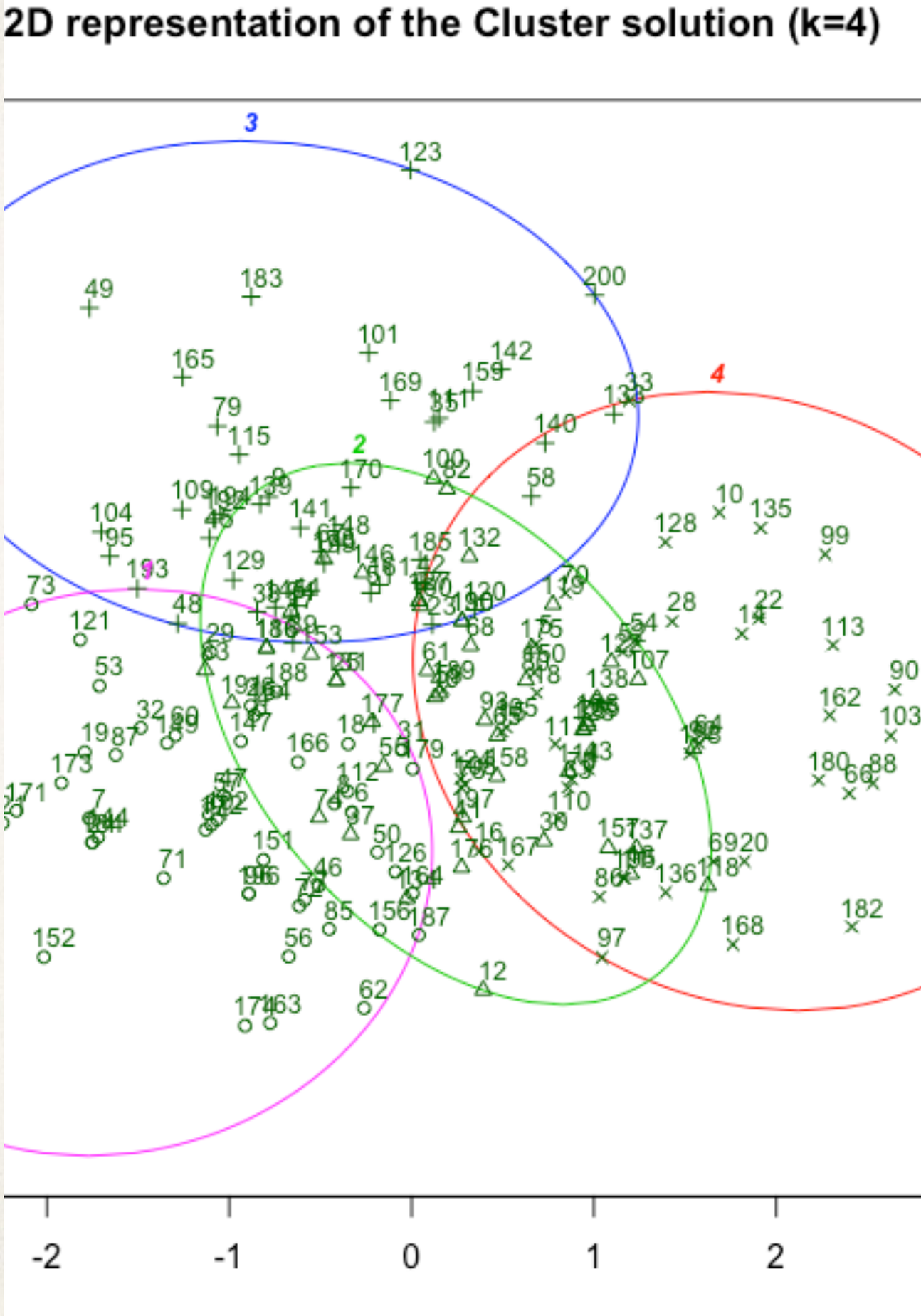
- We tried casese where k=4, 5, 6 and 7 and concluded the case of k=6 gives the best result for analysis.

2D representation of the Cluster solution (k=4)

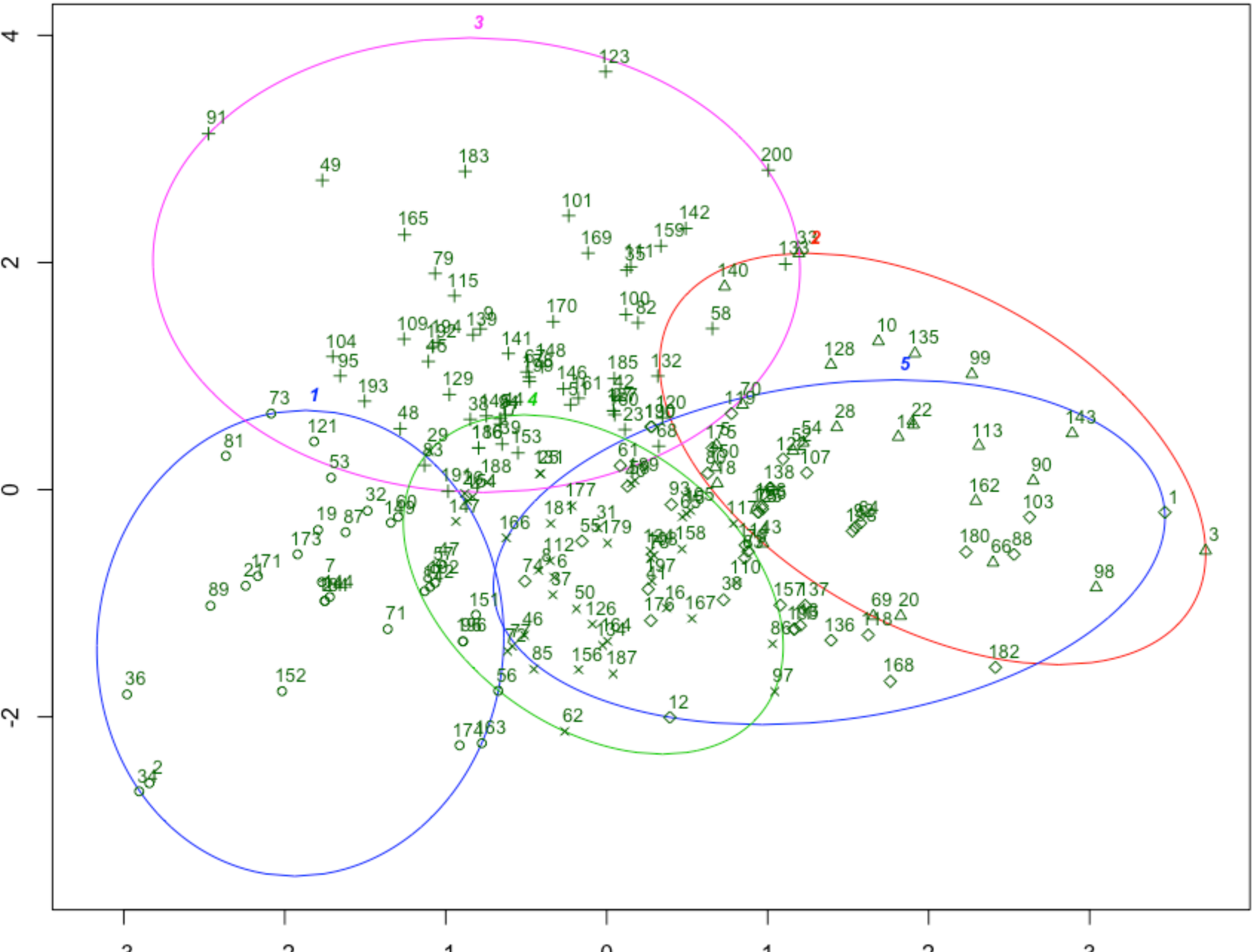


Clustering of k=4

- Looks underclassified
  - failing to identify transitions
    - between C1 and C4
    - between C1 and C3
    - between C3 and C4

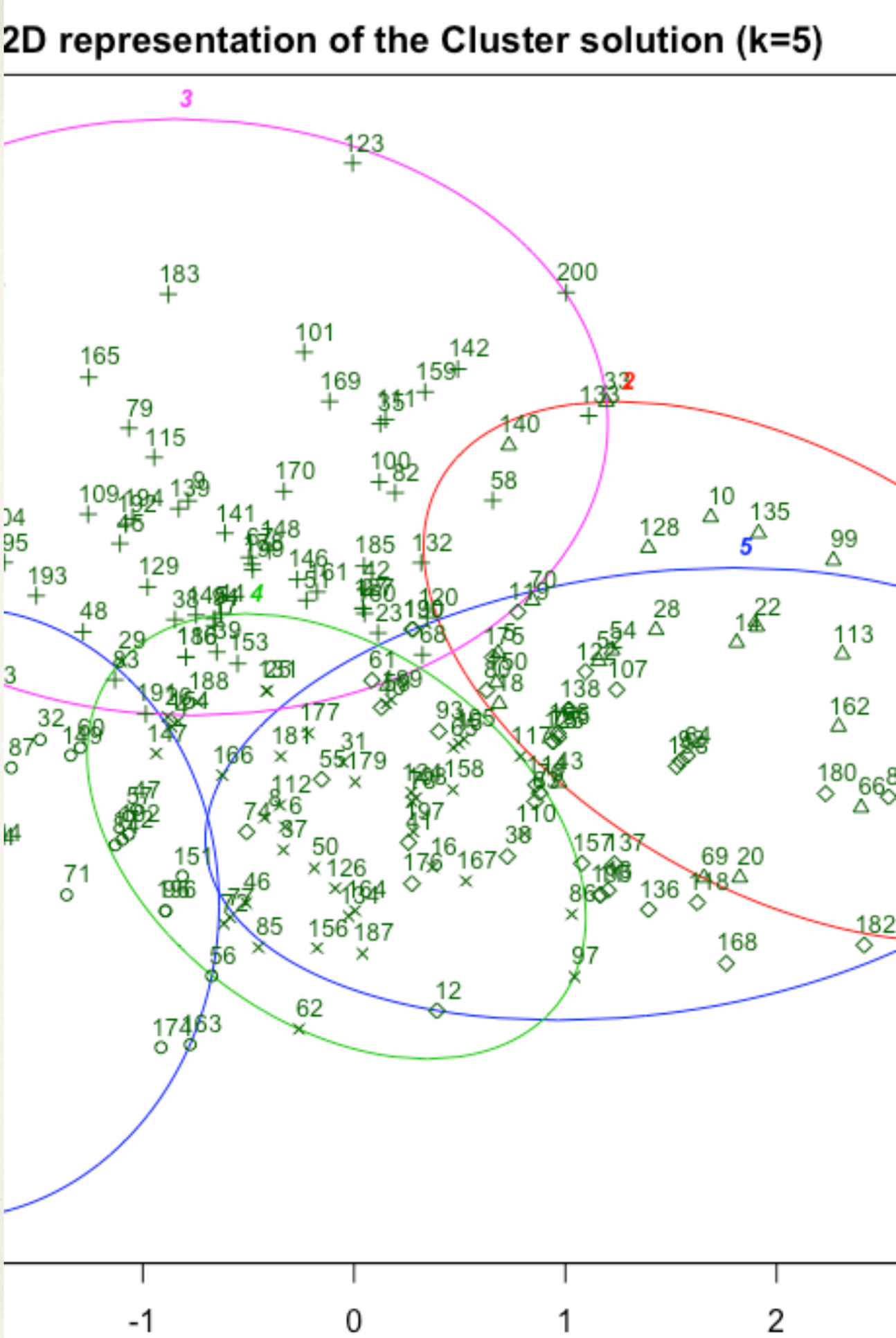


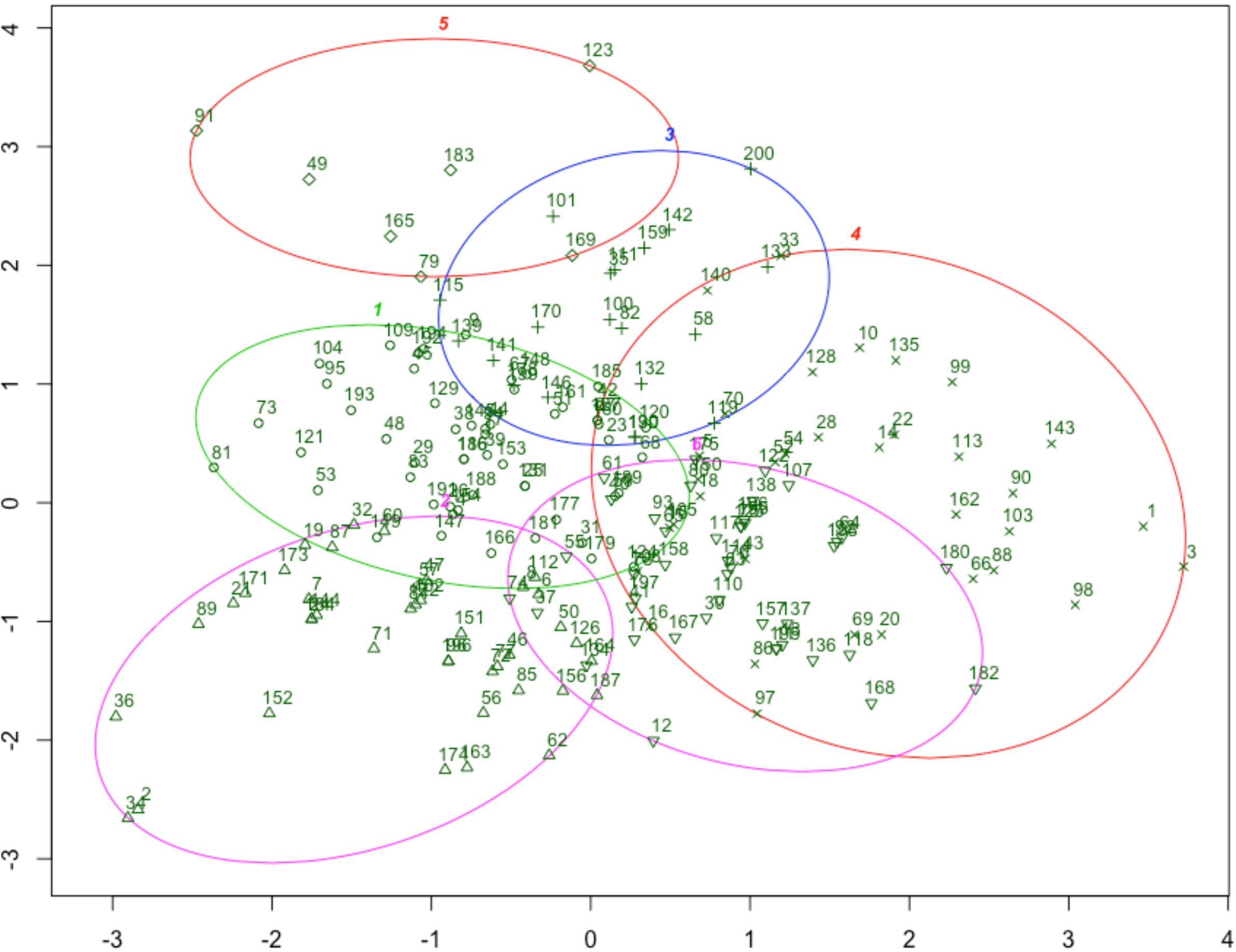
2D representation of the Cluster solution (k=5)



Clustering of k=5

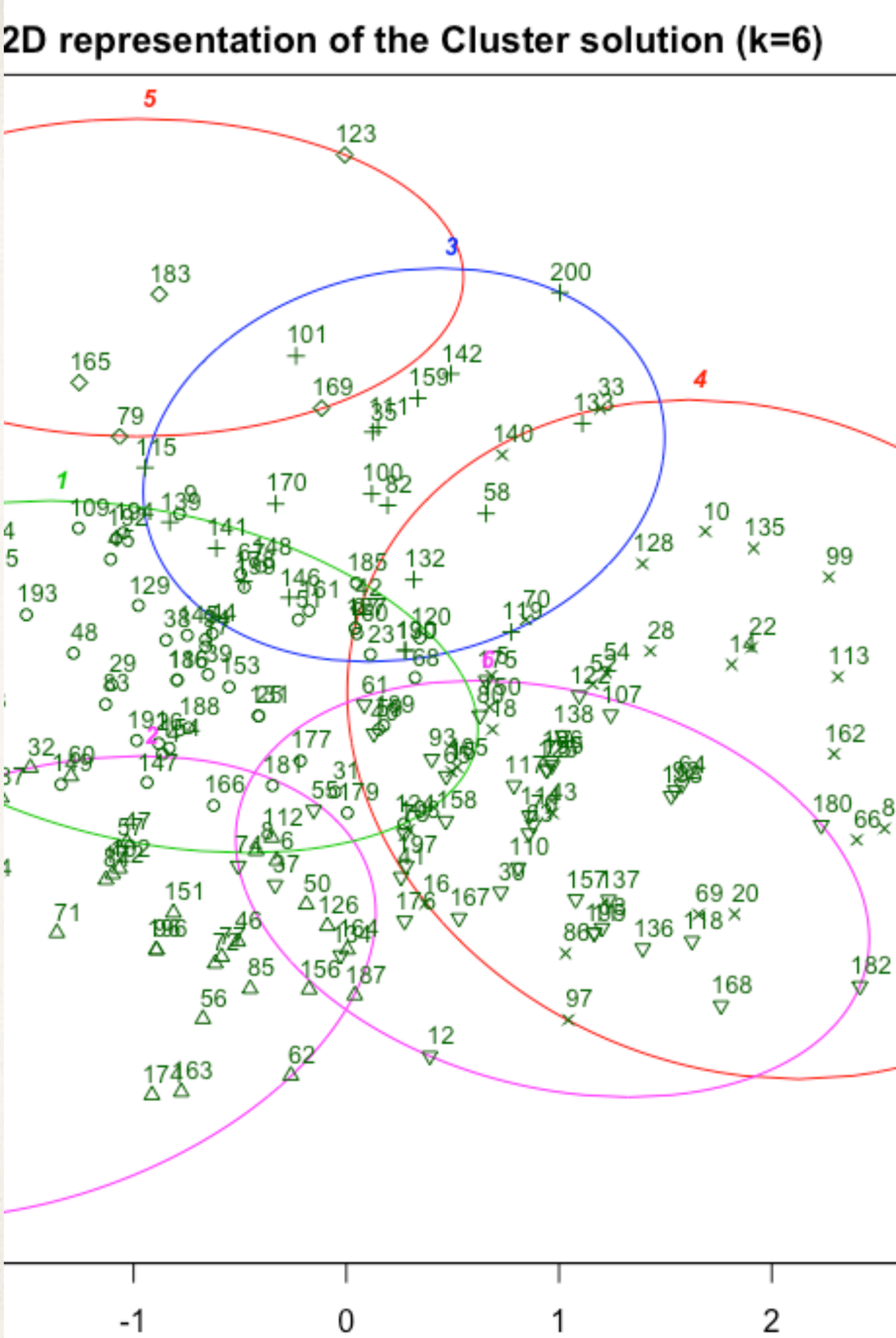
- Looks underclassified
  - Too much overlap
    - between C2 and C5
    - between C4 and C5



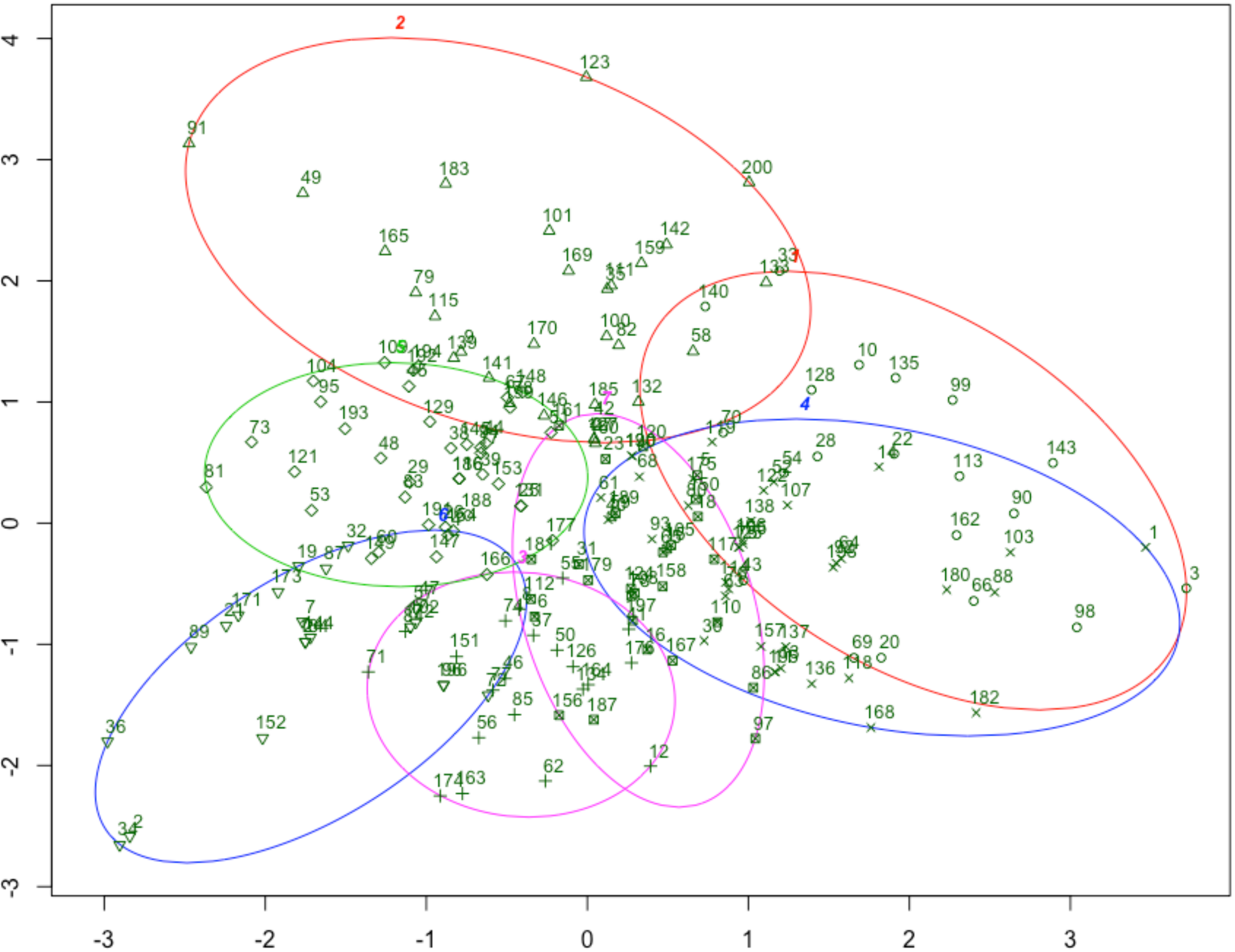

2D representation of the Cluster solution (k=6)
1
2
3
4
5
6
-3
-2
-1
0
1
2
3
4

Clustering of k=6

- Looks good
  - balancing between separation and overlapping

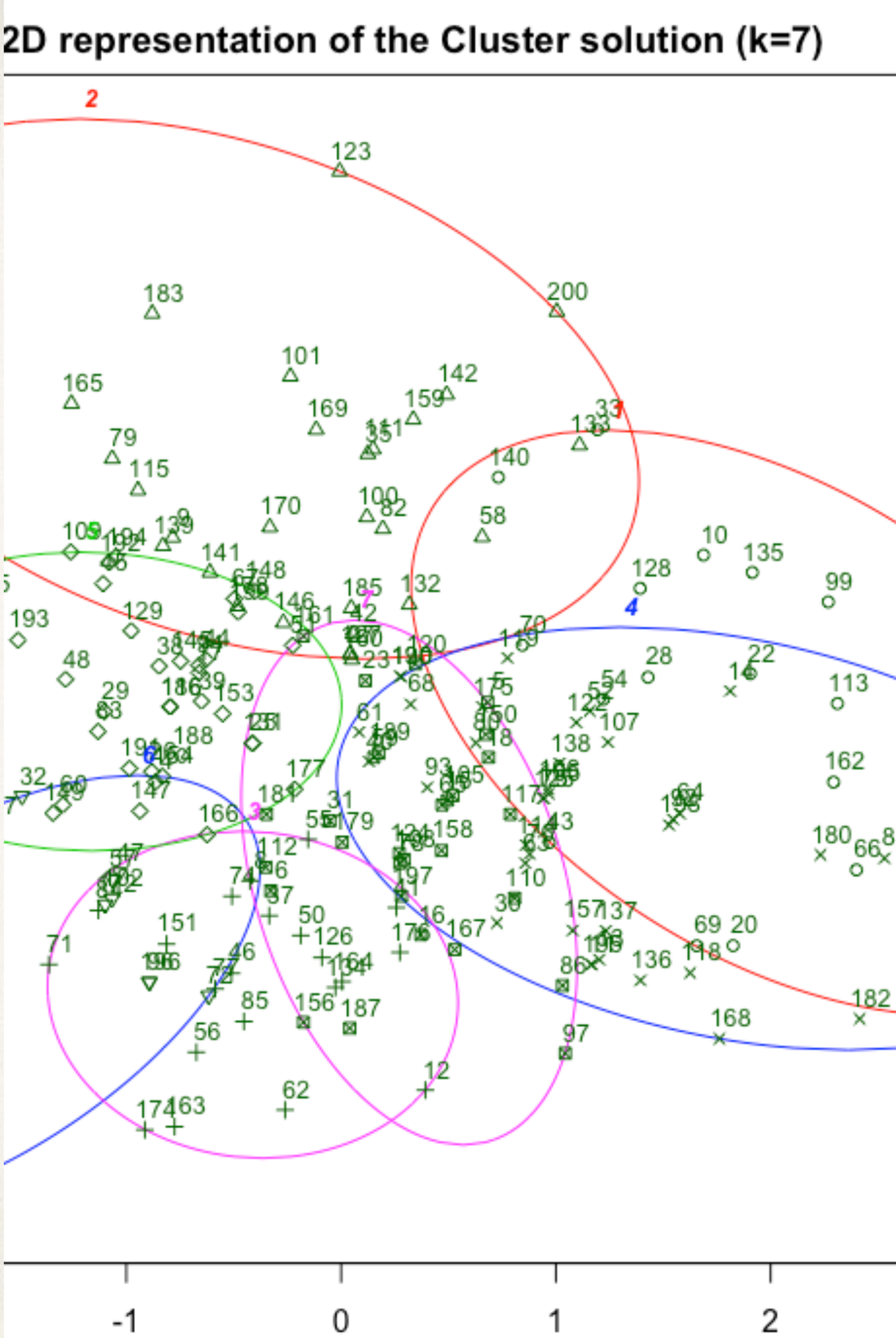


2D representation of the Cluster solution (k=7)

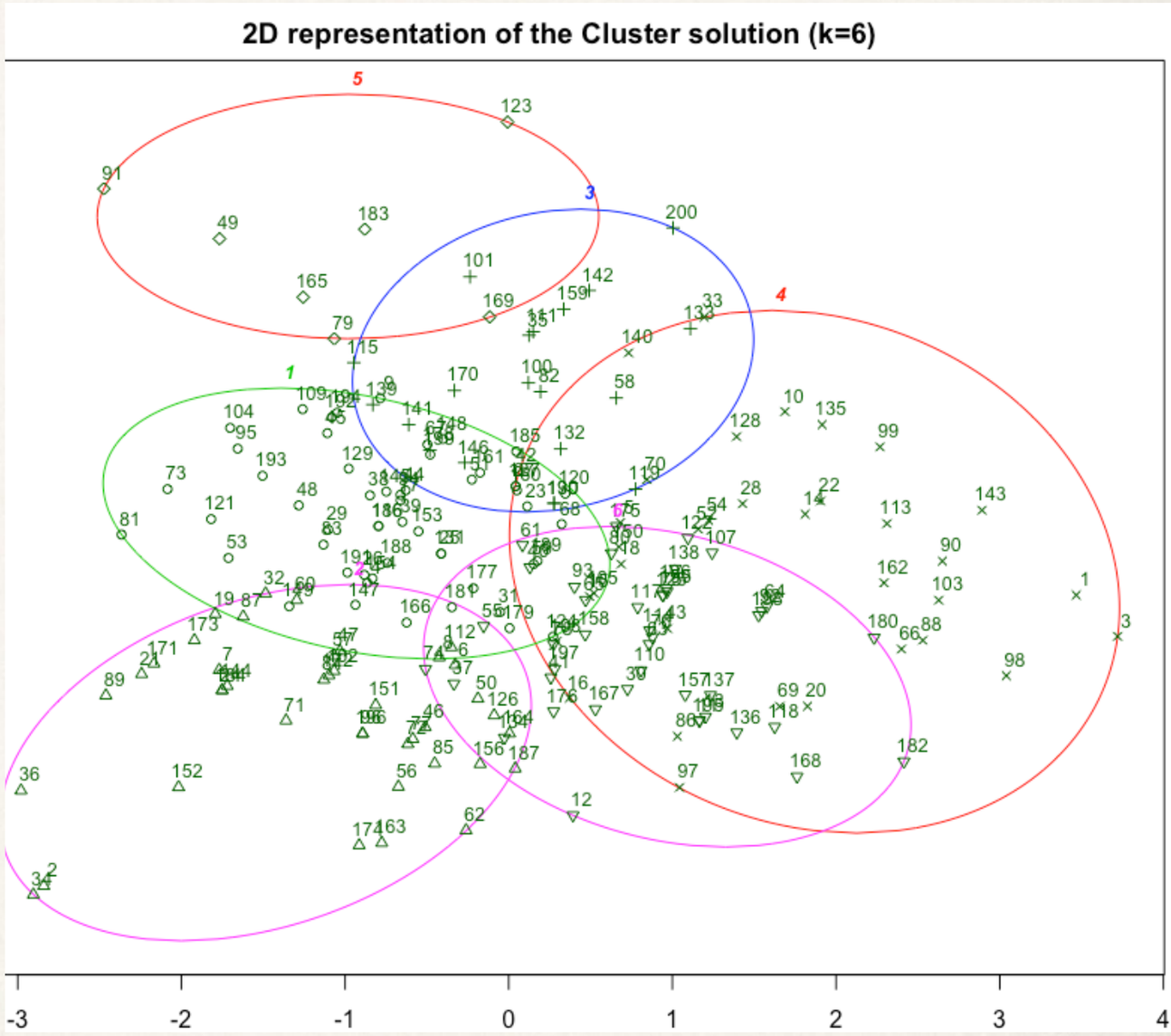


Clustering of k=7

- Looks overclassified
  - Too few proper clusters

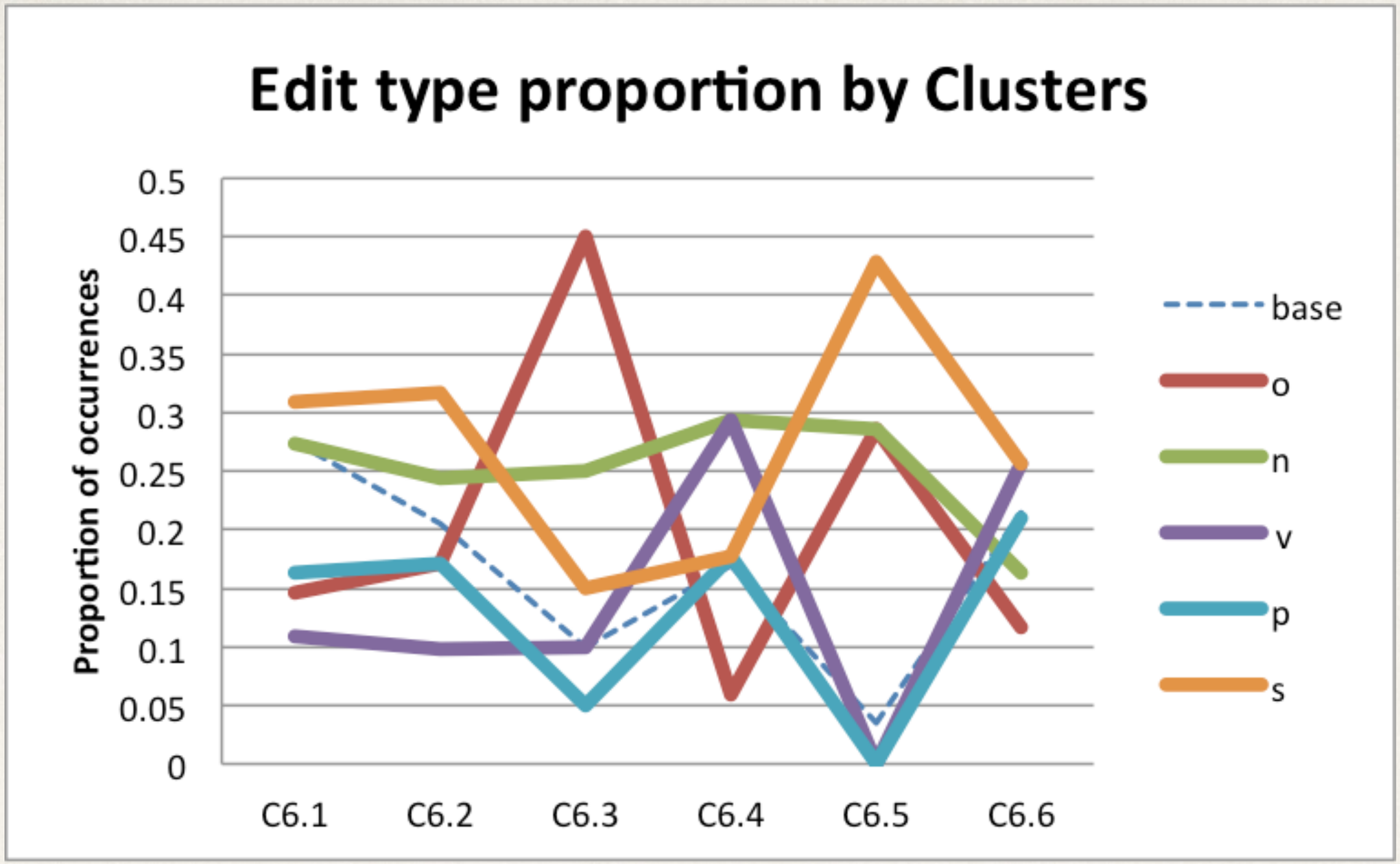


Closer look into k=6 clustering



Partial orderings of dominant response ranges

- Left is more dominant than right.
  - C6.1: [2,1] > [1,0] > [3,2], [4,3] (mild deviance 1)
  - C6.2: [2,1] > [1,0], [3,2], [4,3] (mild deviance 2)
  - C6.3: [1,0], [2,1] > [3,2] > [4,3] (slight deviance)
  - C6.4: [4,3], [2,1] > [3,2] > [1,0] (strong deviance)
  - C6.5: [1,0] > [2,1] > [4,3] > [3,2] (no deviance)
  - C6.6: [2,1], [3,2] > [1,0], [4,3] (mild deviance 3)



Proportions of edit types in Clusters 1/2

Proportions of edit types in Clusters 2/2

- C6.1 of deviant stimuli
  - contains fewer examples of o-, p- and v-types
- C6.2 of deviant stimuli
  - contains significantly more examples of s-type, relatively more examples of n-type, and relatively fewer examples of v-type
- C6.3 of deviant stimuli
  - contains significantly more examples of o-type and relatively more examples of n-type
- C6.4 of deviant stimuli
  - contains significantly more examples of v-type, significantly fewer examples of o-type, relatively more examples of n-type
- C6.5 of normal stimuli
  - contains no examples of v- and p-types, and significantly more examples of s- and o-types
- C6.6 of deviant stimuli
  - contains examples of all types equally

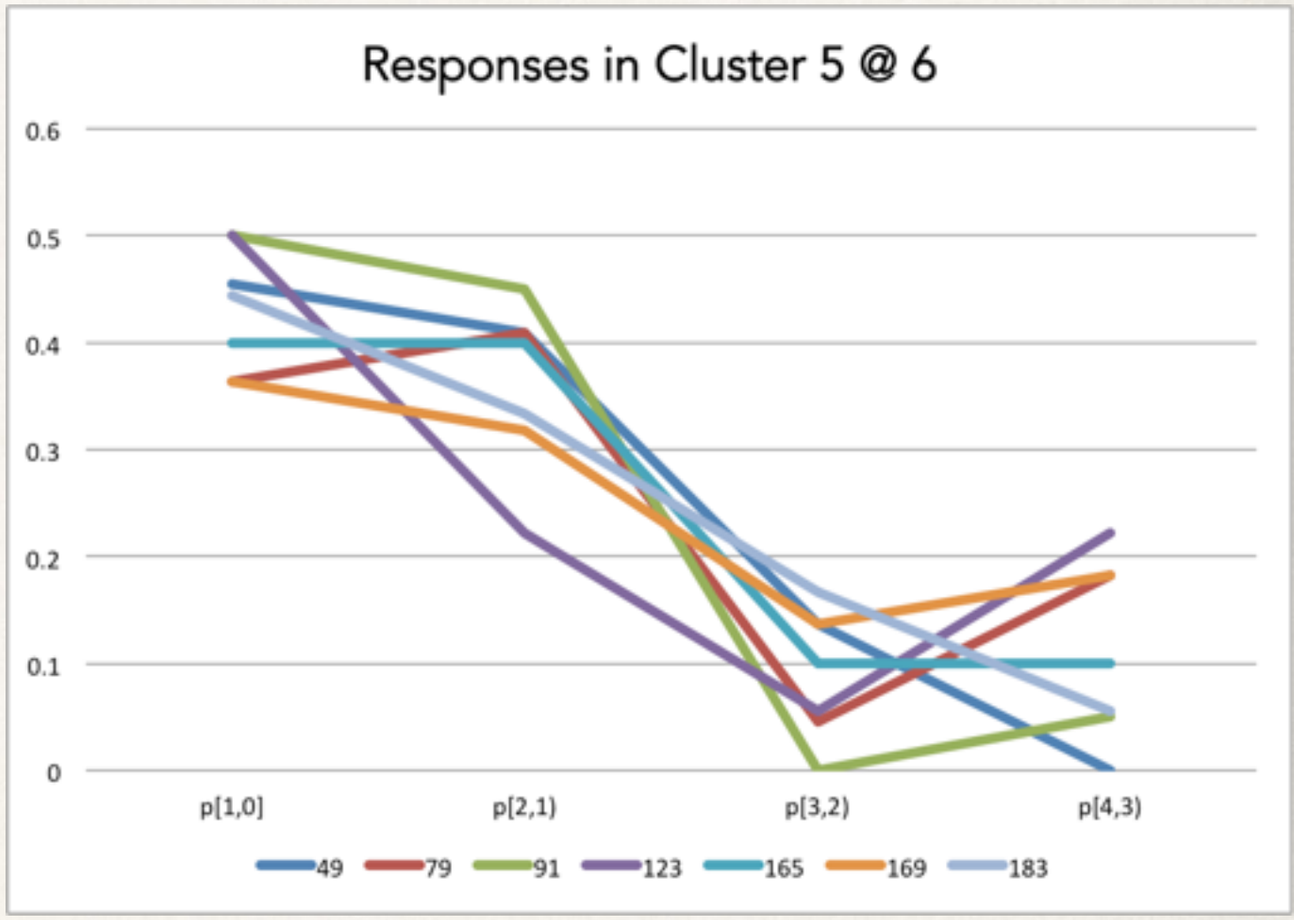
Pairwise (dis)similarity by KL-divergence

- Most similar pairs
  - C6.1 and C6.3
  - C6.2 and C6.1
- Rather dissmilar pairs
  - C6.2 and C6.5
  - C6.4 and C6.5
- Most dissimilar pair
  - C6.6 and C6.5

	C6.1	C6.2	C6.3	C6.4	C6.5	C6.6
C6.1	0.000	0.070	0.062	0.156	0.106	0.101
C6.2	0.086	0.000	0.264	0.200	0.389	0.103
C6.3	0.062	0.228	0.000	0.245	0.102	0.129
C6.4	0.137	0.200	0.196	0.000	0.327	0.091
C6.5	0.100	0.281	0.133	0.336	0.000	0.368
C6.6	0.102	0.096	0.160	0.099	0.402	0.000

Response curves in Cluster 6.5

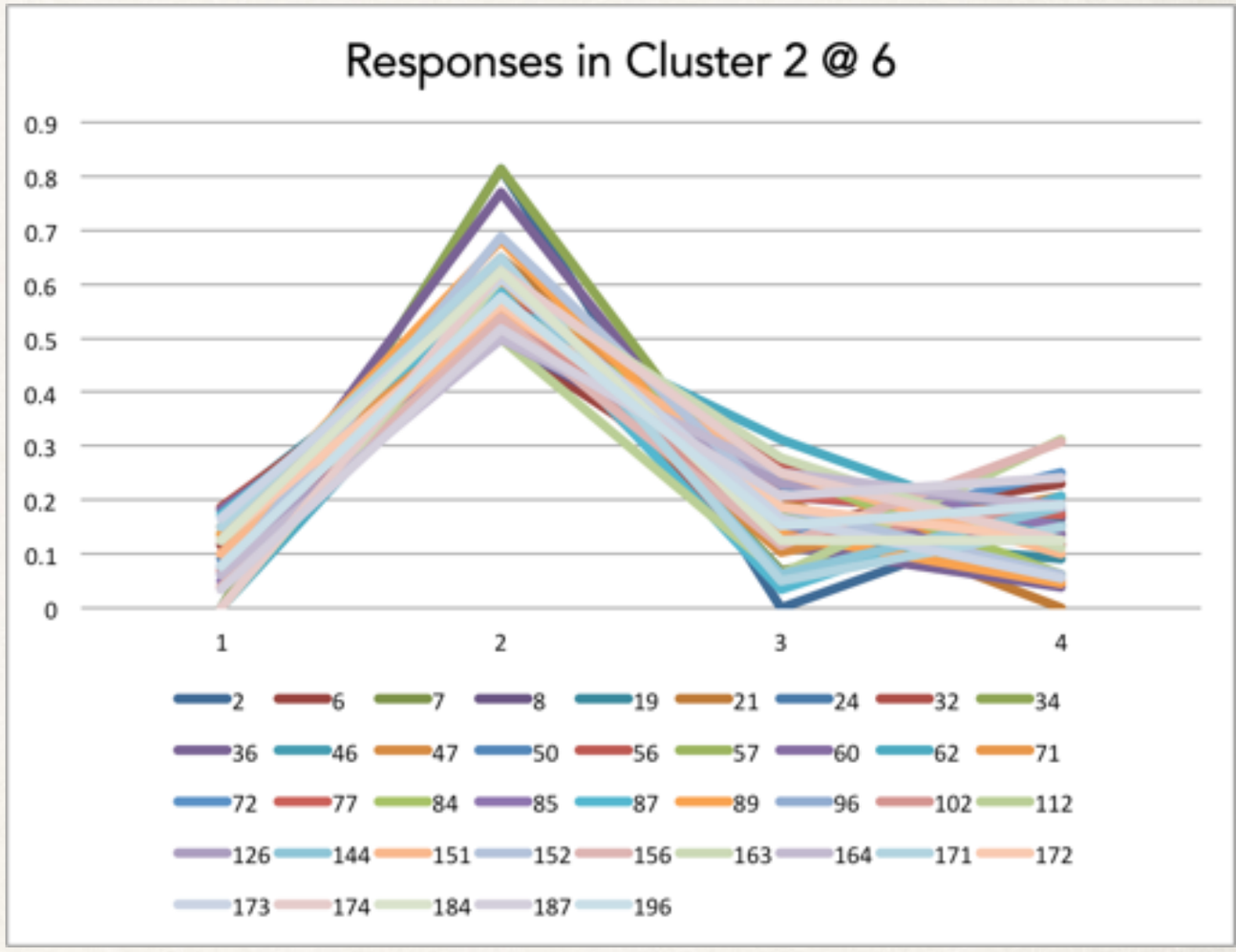
- Interpretation
  - This cluster collects responses to stimuli with no deviance.



Examples of C6.5 stimuli

- 知人が職場で同僚から奇病を感染した。[o-type]
- 夫が墓場で真夜中に妻と知り合った。[n-type]
- 講師が学生に良い解法を受験対策で教えた。[s-type]
- 特徴
  - v-type, p-type がない

Response curves in Cluster 6.2

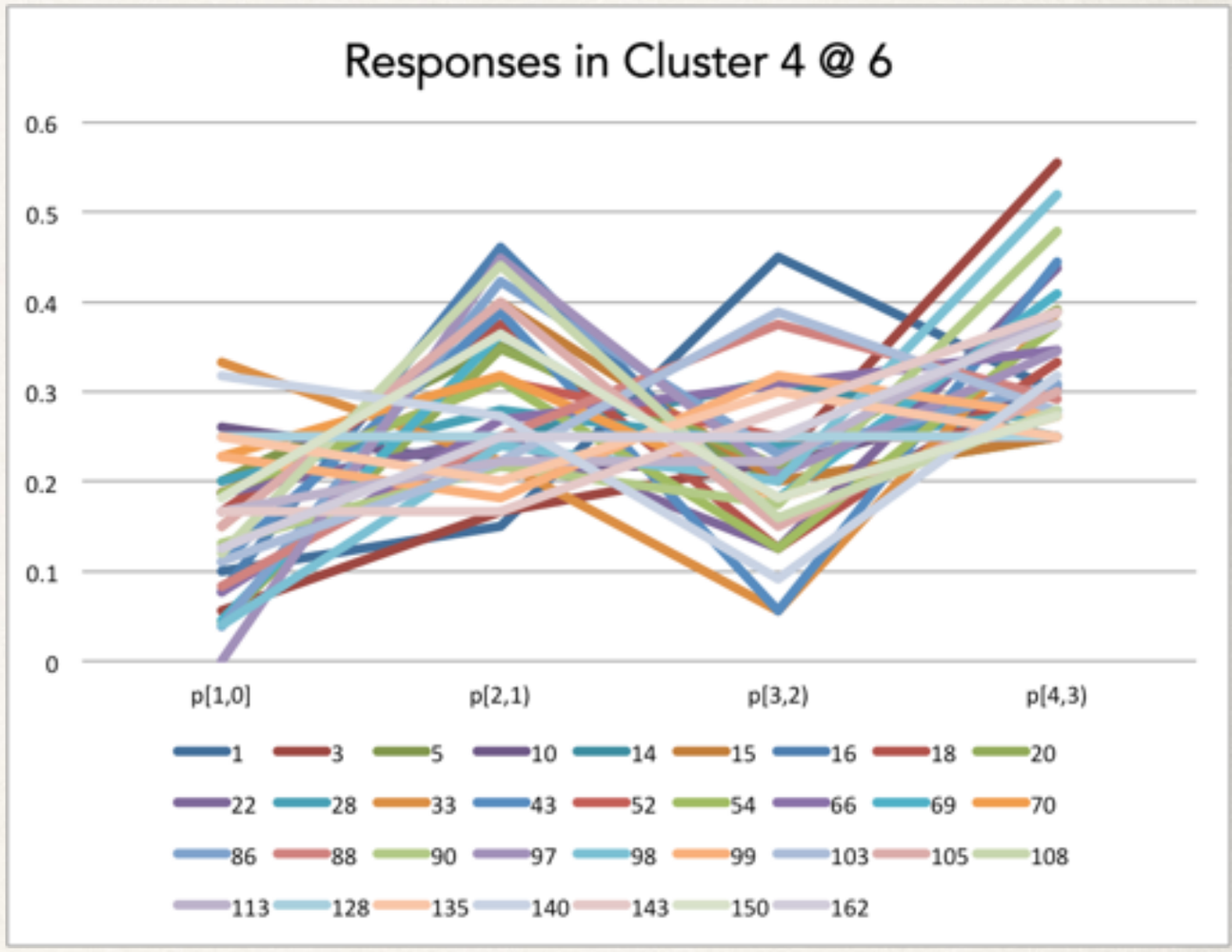


- Interpretation
  - This cluster collects responses to stimuli with mild deviance (of category 1).

Examples of C6.2 stimuli

- 青年が合格発表の場で足下から幸福を感じた。[o-type]
- 彼女が手紙で根源を不意に知った。[n-type]
- 高校生がデートの場でしらじらしさを恋人に感じ取った。[v-type]
- 同僚が質問で相手から否と答えた。[p-type]
- 格下相手に大会でその候補が楽な試合を負けた。[s-type]
- 知人が学会で否と質問に答えた。[s-type]

Response curves in Cluster 6.4



- Interpretation
  - This cluster collects responses to strong deviant stimuli (of category 2).

Examples of C6.4 stimuli

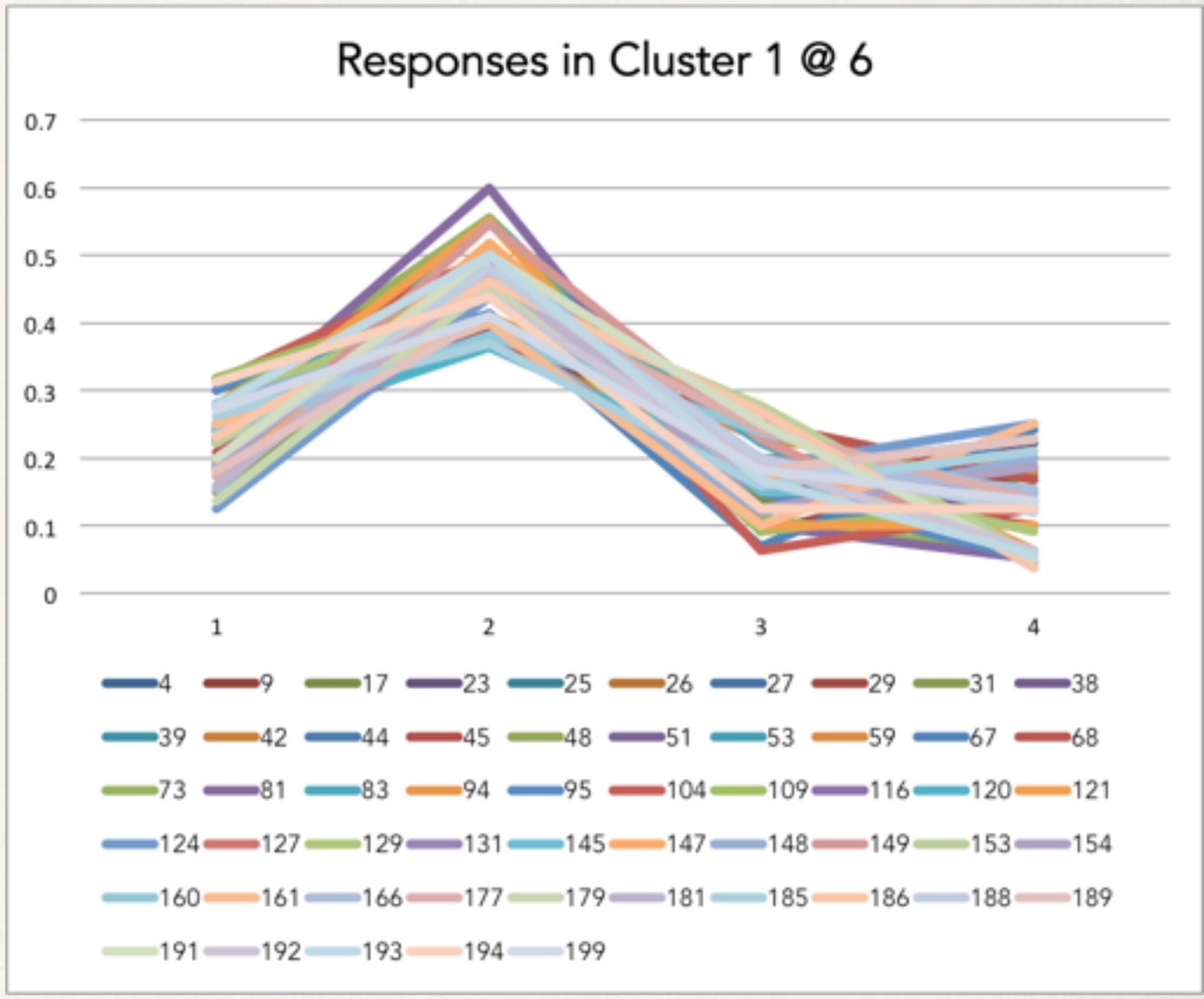
- 知人が学会で質問に否と答えた。[o-type]
- 研究者がアンケートで無作為に漠然と探した。[n-type]
- 彼女が手紙で真実を不意に話しかけた。[v-type]
- 捕虜に尋問で忠誠心から秘密を黙った。[p-type]
- 段々とネットワークに不注意でコンピュータウイルスが感染した。[s-type]
- 特徴
  - o-type が少ない

Response curves in Cluster 6.1

- Interpretation

- This cluster collects responses to mildly deviant stimuli.

- This cluster is a mixture of C6.5 and C6.2.

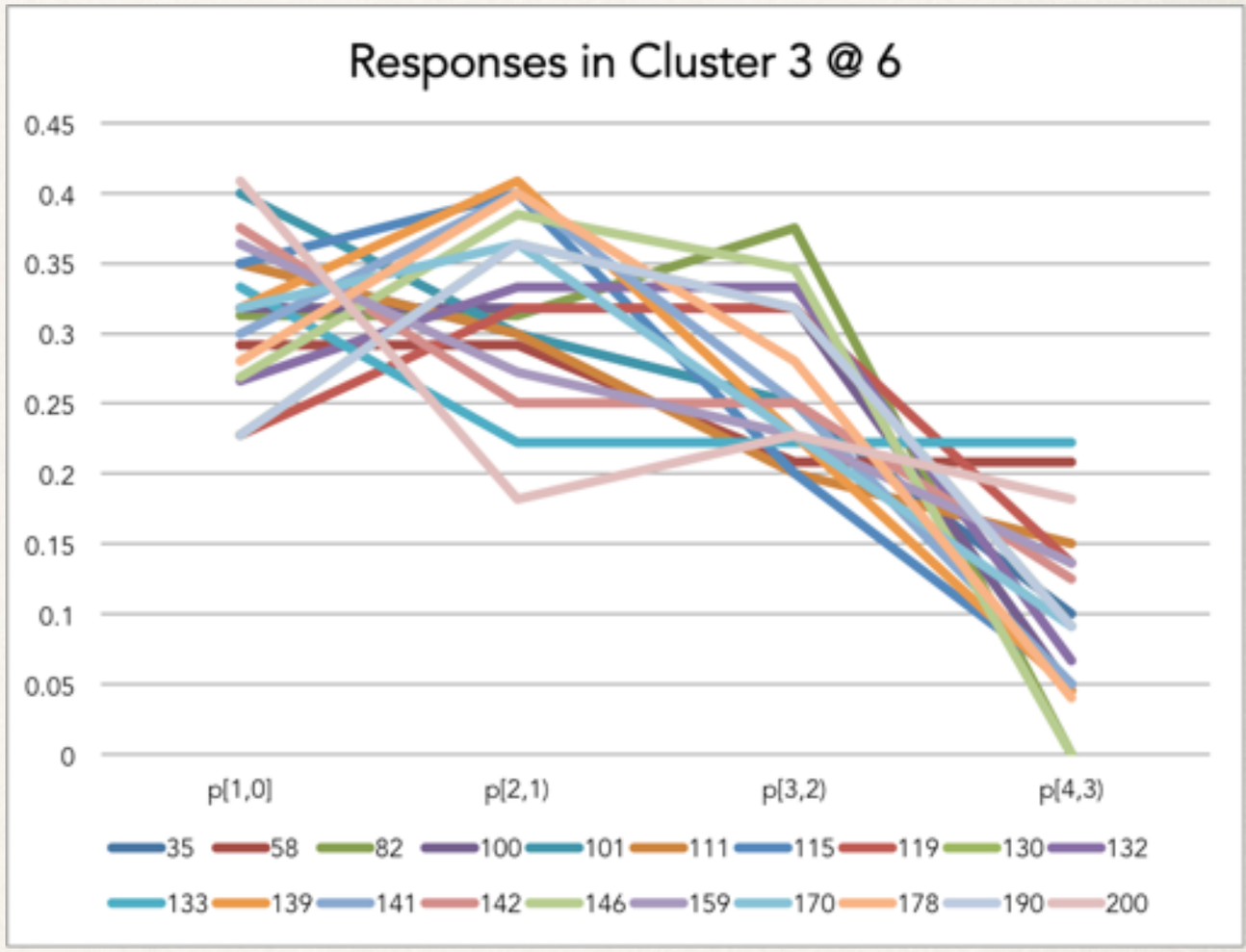


Examples of C6.1 stimuli

- 講師が受験対策で学生に良い解法を教えた。[o-type]
- 同僚が学会で質問に否と答えた。[n-type]
- 青年が合格発表の場で足下から幸福をひたった。[v-type]
- 字からうまい青年が外国で生活の必要から書道を教えた。[p-type]
- 彼女が真実を不意に手紙で知った。[s-type]

Response curves in Cluster 6.3

- Interpretation
  - This cluster collects responses to stimuli with mild deviance.
  - This cluster is a mixture of C6.5 and C6.4.

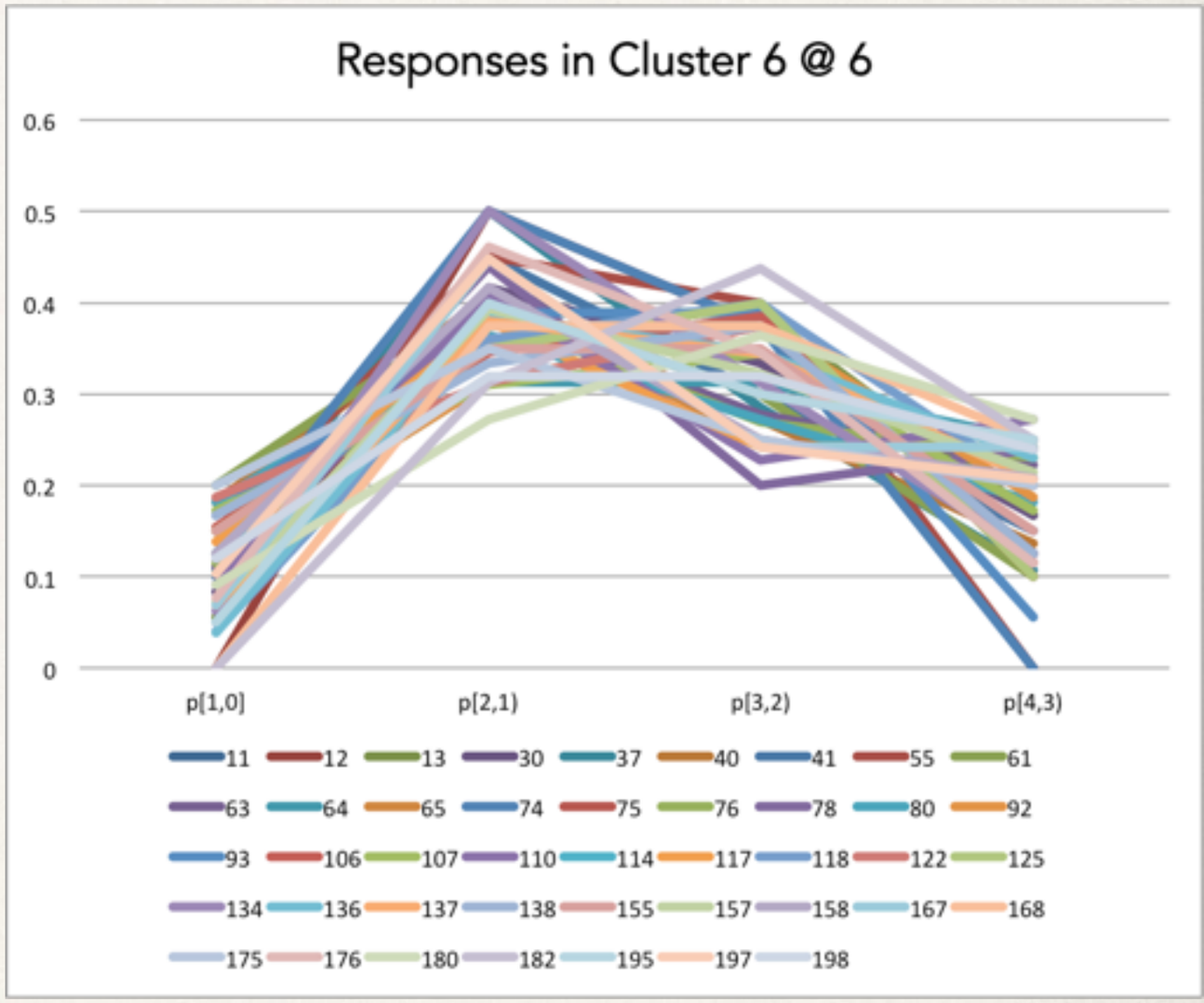


Examples of C6.3 stimuli

- 船が遠回りで海路を安全に行った。[o-type]
- 友人が同僚で同僚から難病を感染した。[n-type]
- 同僚が質問で相手に否と近寄った。[v-type]
- 質問で同僚が否と相手に答えた。[s-type]
- 特徴
  - o-type の割合が高い, v-type の割合が低い

Response curves in Cluster 6.6

- Interpretation
  - This cluster collects responses to stimuli with mild deviance (of mixed categories).
  - This cluster is a mixture of C6.2 and C6.4

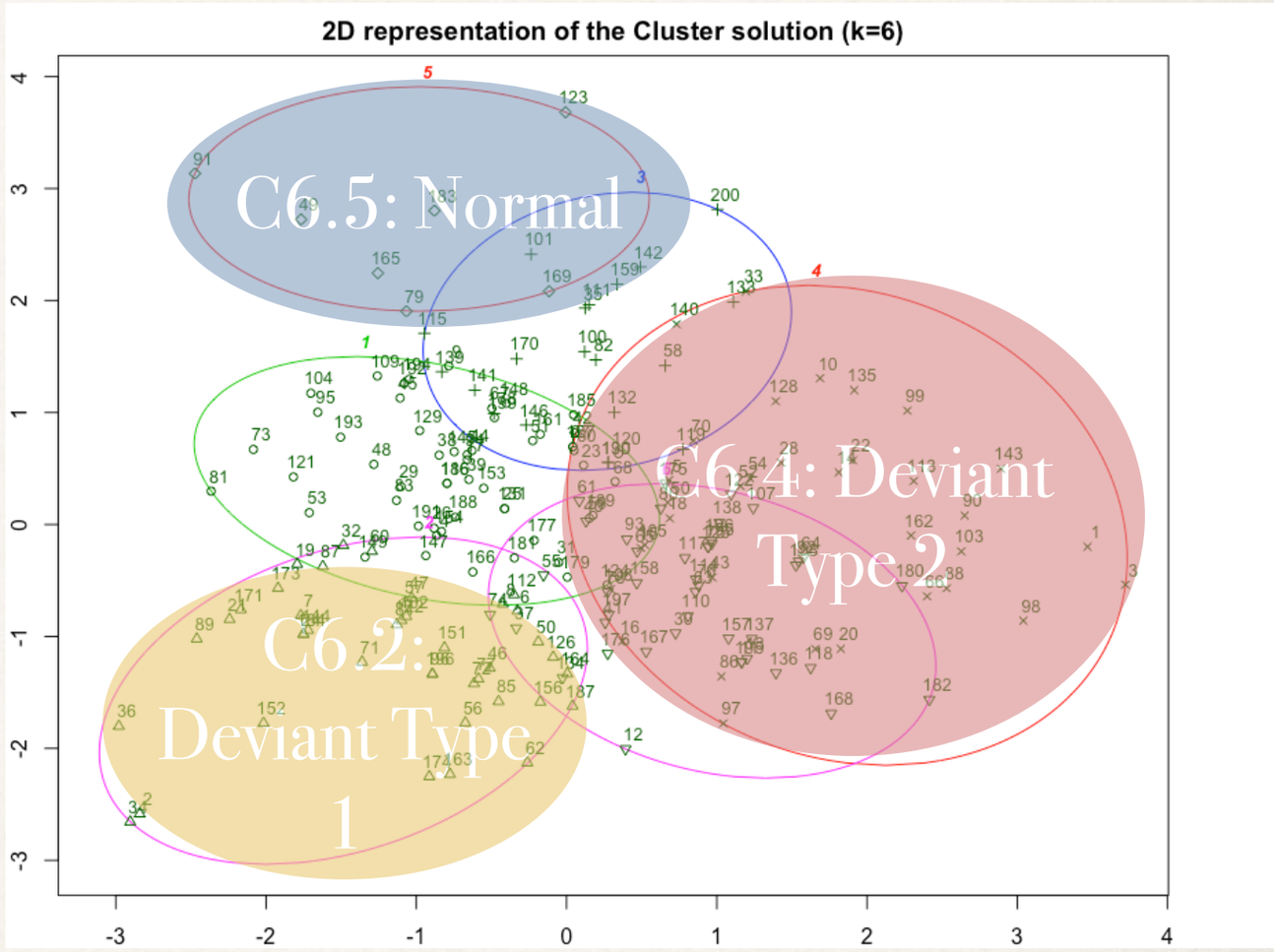


Examples of C6.6 stimuli

- 刑事が捜査で手がかりを手当たり次第に探した。[o-type]
- 船が遠回りで海路を安全に出かけた。[v-type]
- 有名選手が大事な試合で得意種目で無名選手に負けた。[p-type]
- 捕虜が忠誠心から秘密を尋問で黙った。[s-type]
- 周りの影響で感動が外向きに観客を伝わった。[s-type]
- 特徴
  - v-type の割合が高い, n-type の割合が少ない

Discussion

- Finding
  - It was suggested that C6.2 and C6.4 represent distinct classes/ dimensions of deviance.
- Questions to address in future
  - How do they differ?
  - What do they represent?



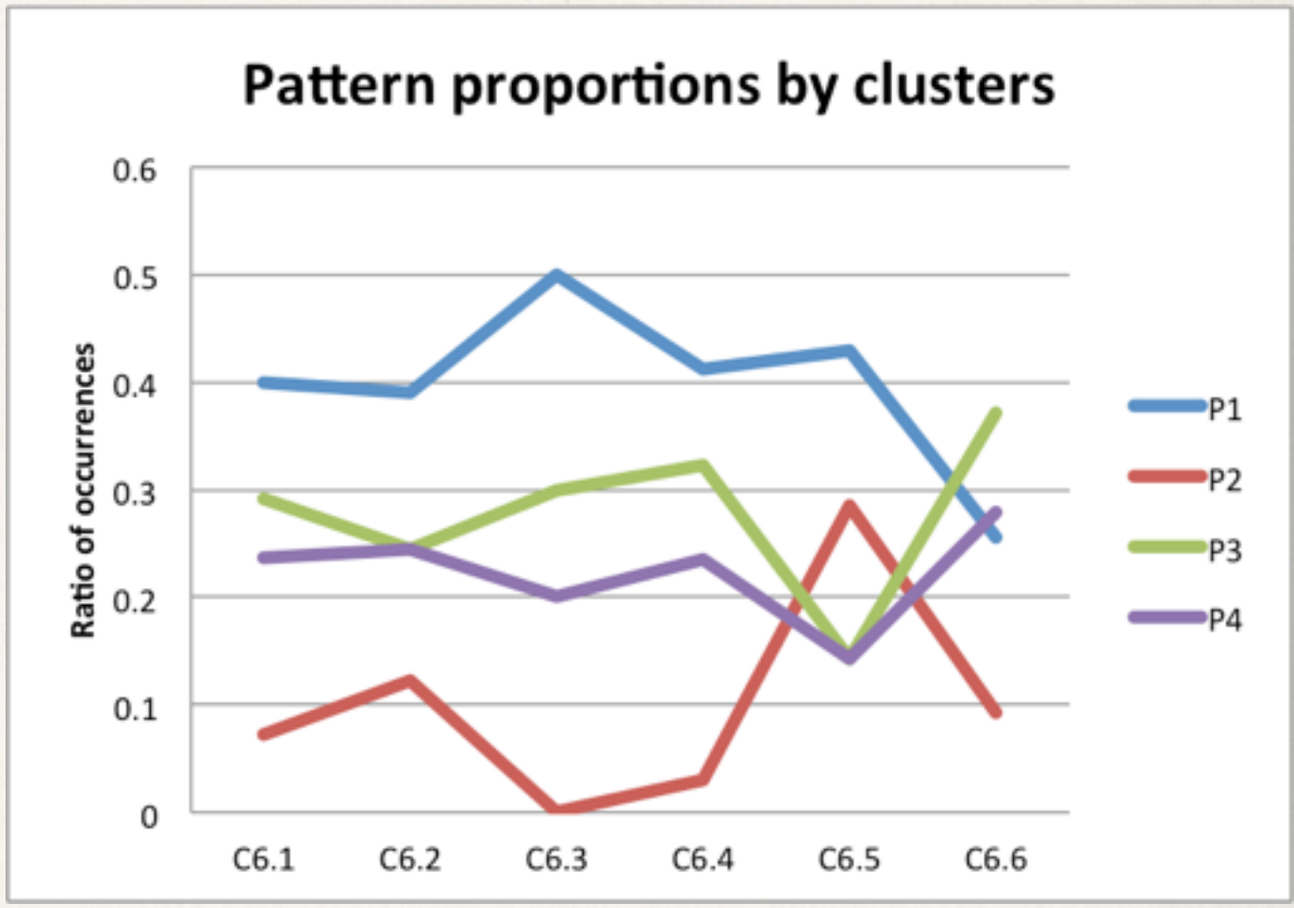
Patternwise comparison

4 Constructions/Patterns used (repeated)

- P1: _-*ga* _-*de* _-*ni* _-*to* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Goal/Place + Committive/Manner + V
  - example: s111: Douryoo-ga shitumon-de aite-ni ina-to kotae-ta.
- P2: _-*ga* _-*de* _-*ni* _-*wo* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Goal/Place + Object/Result + V
  - example: s151: Kazokudure-ga shiohigari-de umi-ni kai-wo sagashi-ta.
- P3: _-*ga* _-*de* _-*wo* _-*ni* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Object/Result + Goal/Place + V
  - example: s197: Kanojo-ga tegami-de shinjitu-wo fui-ni shit-ta.
- P4: _-*ga* _-*de* _-*kara* _-*wo* V-(*shi*)*ta*
  - Gloss: Nominative + Instrument/Locative + Source/Material + Object/Result + V
  - example: s71: Horyo-ga jinmon-de chuuseishin-kara himitu-wo damat-ta.
- I have much to say about how I decided on these 4 patterns, but I don't have enough time.

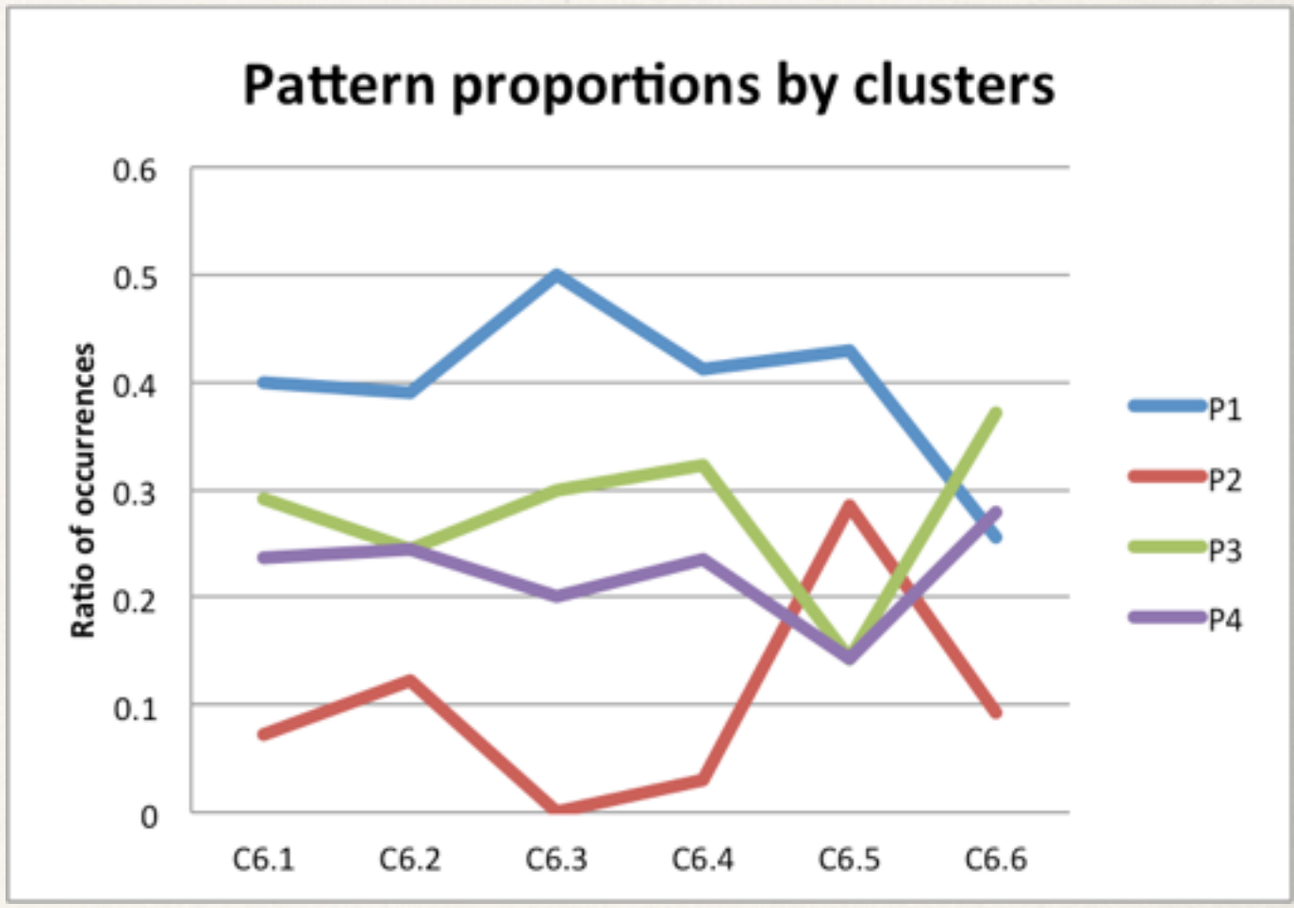
Proportions of patterns in Clusters 1/2

- Proportons of 4 patterns differ for clusters.
  - P1 is relatively more frequent in C6.3 and signifcantly less frequent in C6.6
  - P2 is relatively less frequent in C6.3 and 6.4, and significantly more frequent in C6.5.
  - P3 is significantly less frequent in C6.5.
  - P4 is relatively less frequent in C6.5.
  - P3 and P4 are significantly less frequent in C6.5.

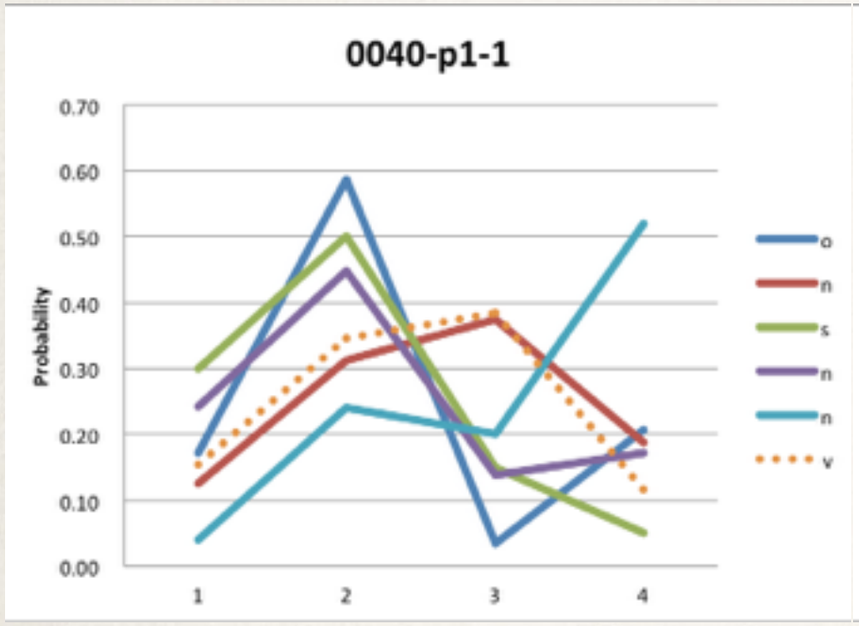


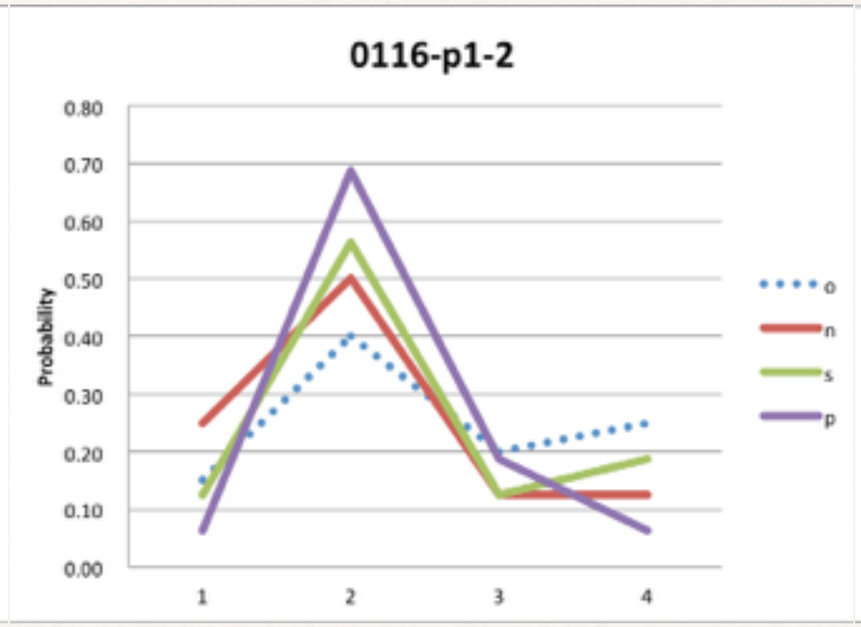
Proportions of patterns in Clusters 2/2

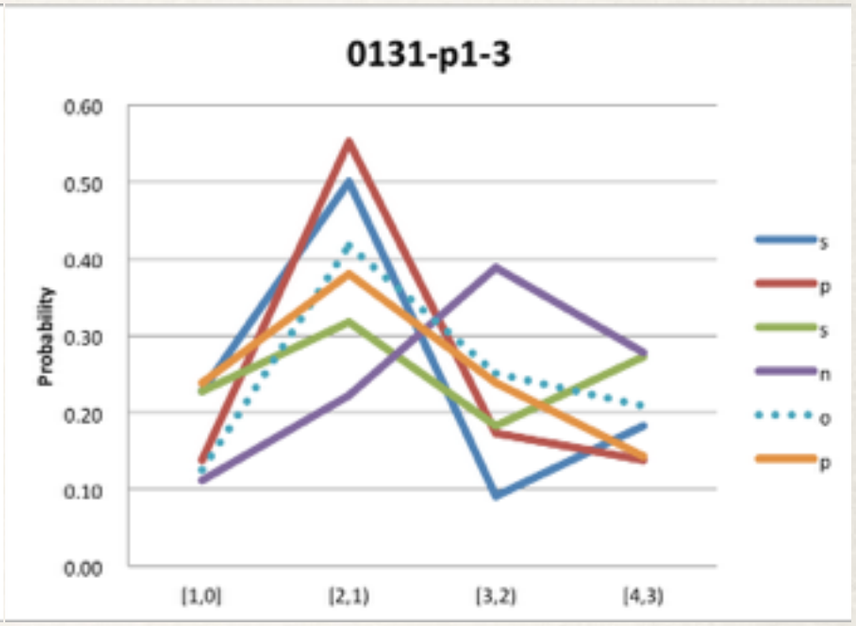
- Proportons of 4 patterns differ for clusters.
  - C6.1 and C6.2 have patterns in their expected ratios. Other four clusters are different.
  - C6.3 contains relatively fewer examples of P2 and relatively more examples of P1.
  - C6.4 contains relatively fewer examples of P2.
  - C6.5 contains significantly (and even drastically) more examples of P2, and relatively fewer examples of P3 and P4.
  - C6.6 contains significantly fewer examples of P1 and relatively more examples of P4.

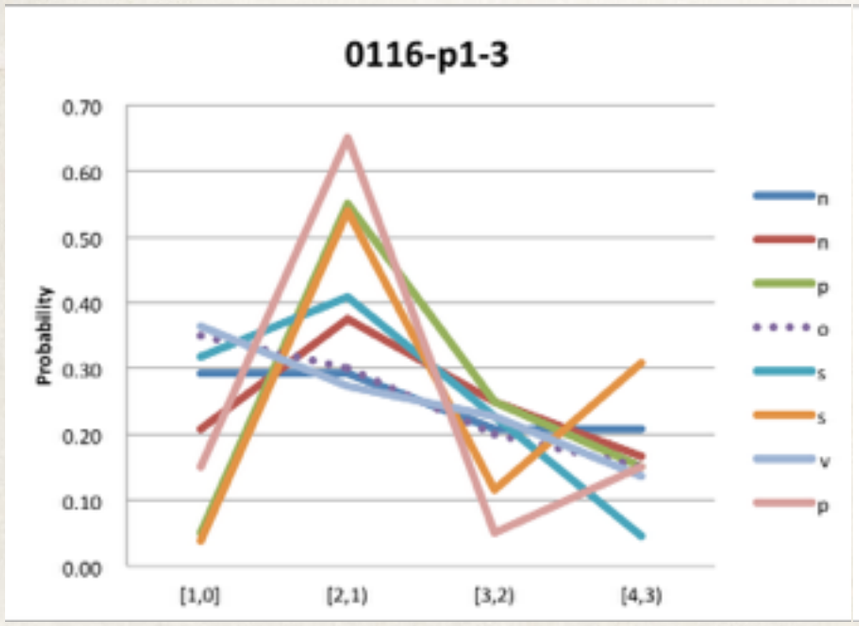


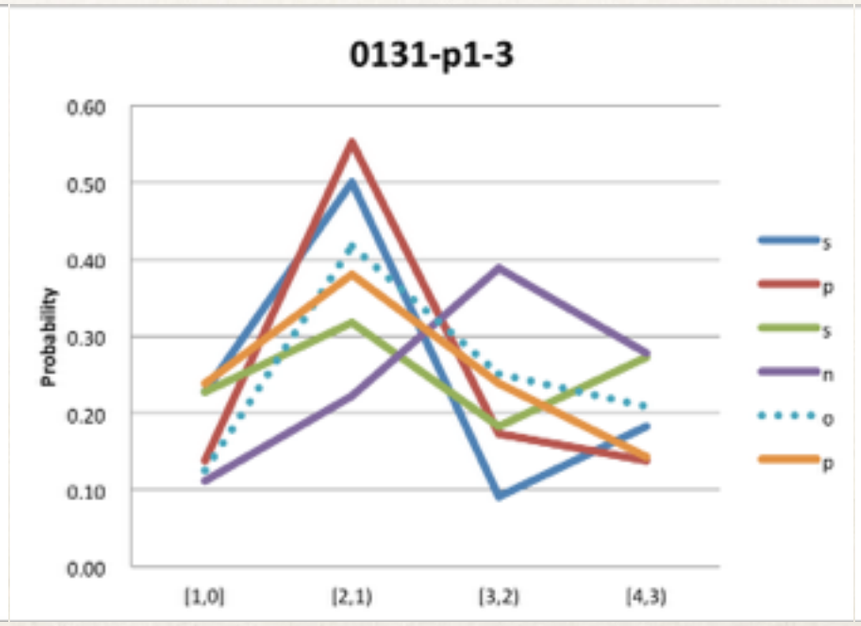
Individual responses to P1 1/2

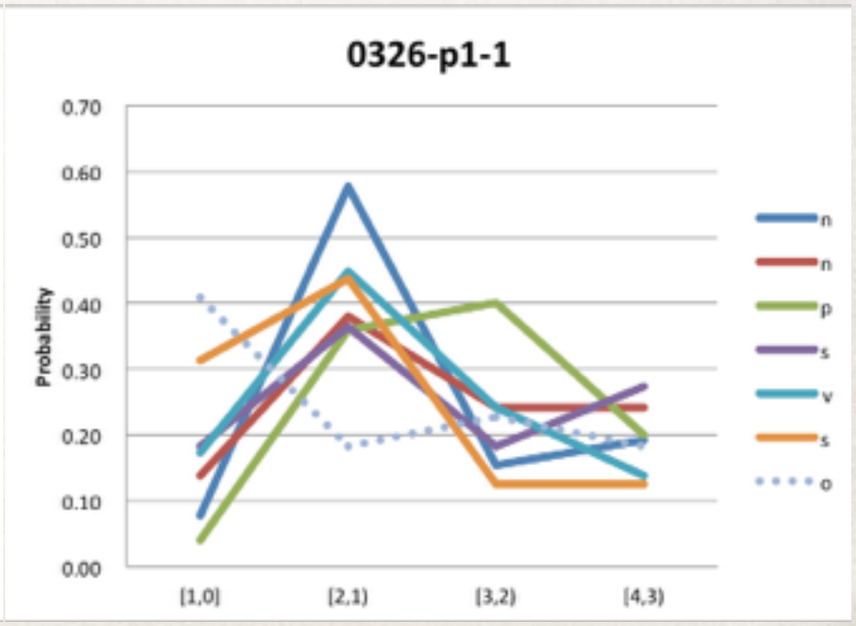




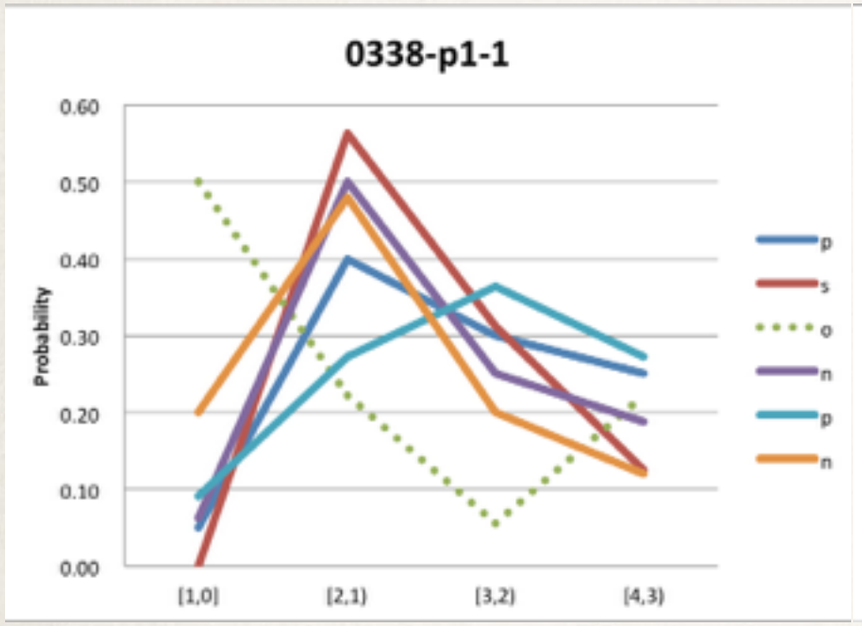


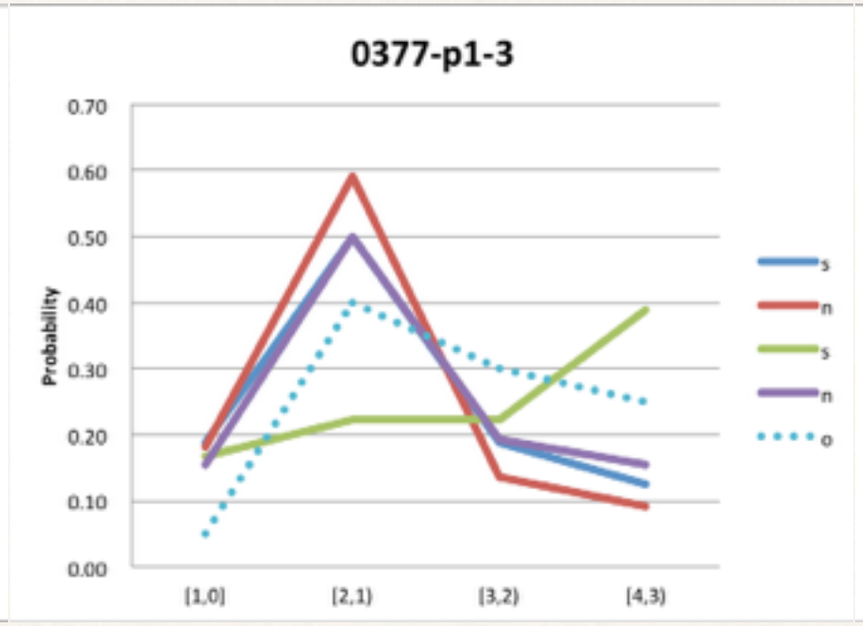


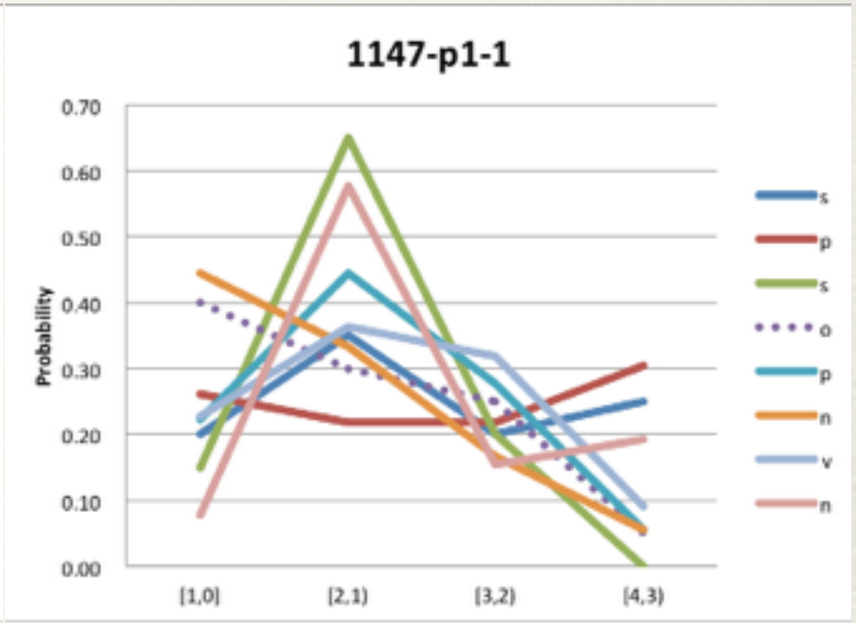


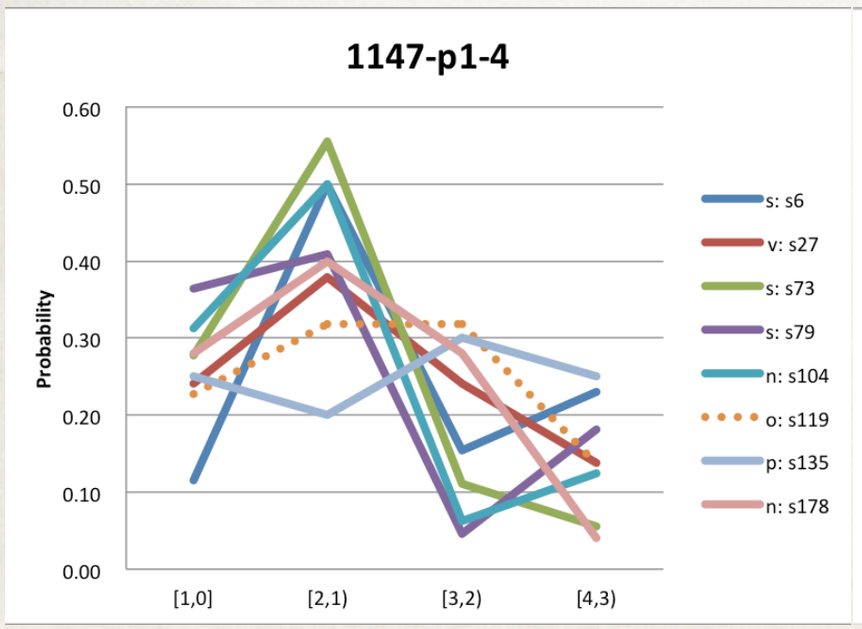


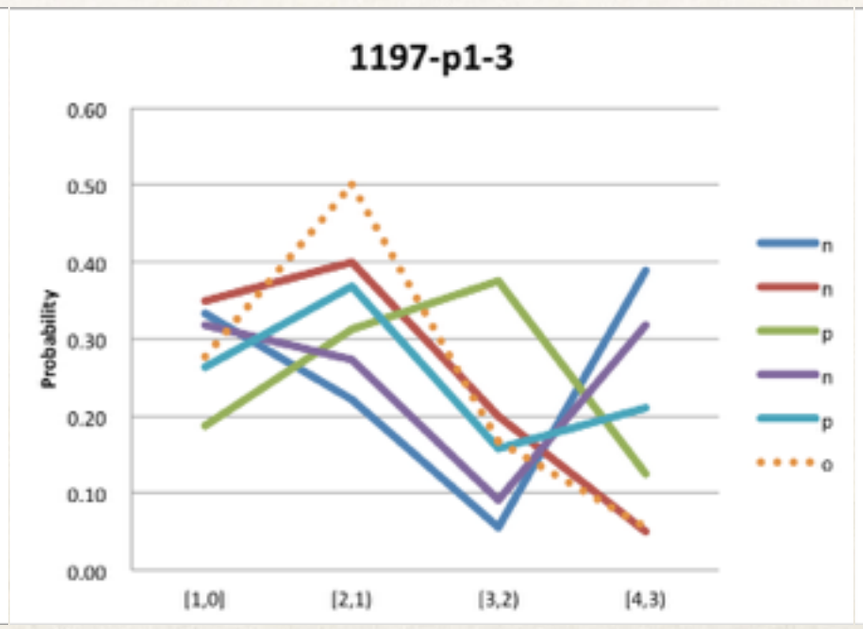
Individual responses to P1 2/2

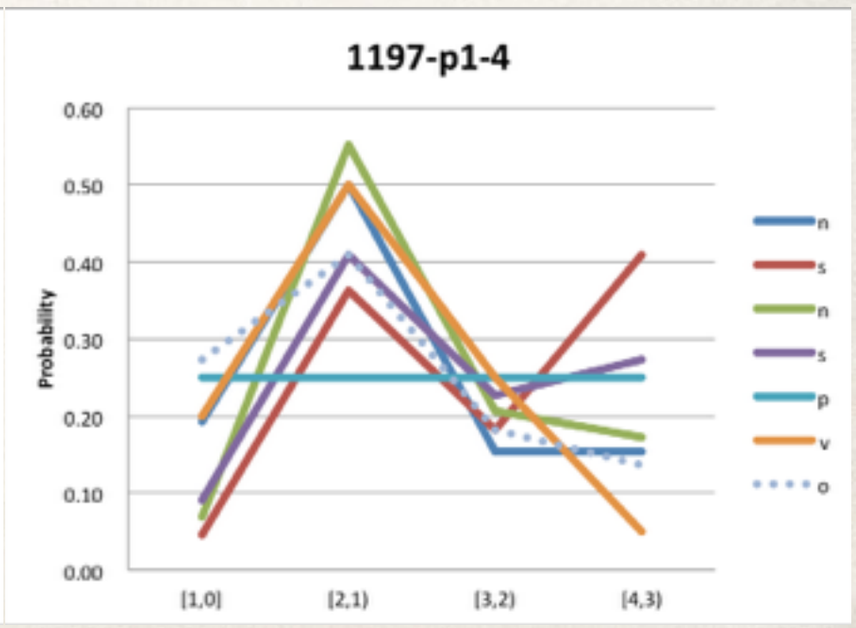




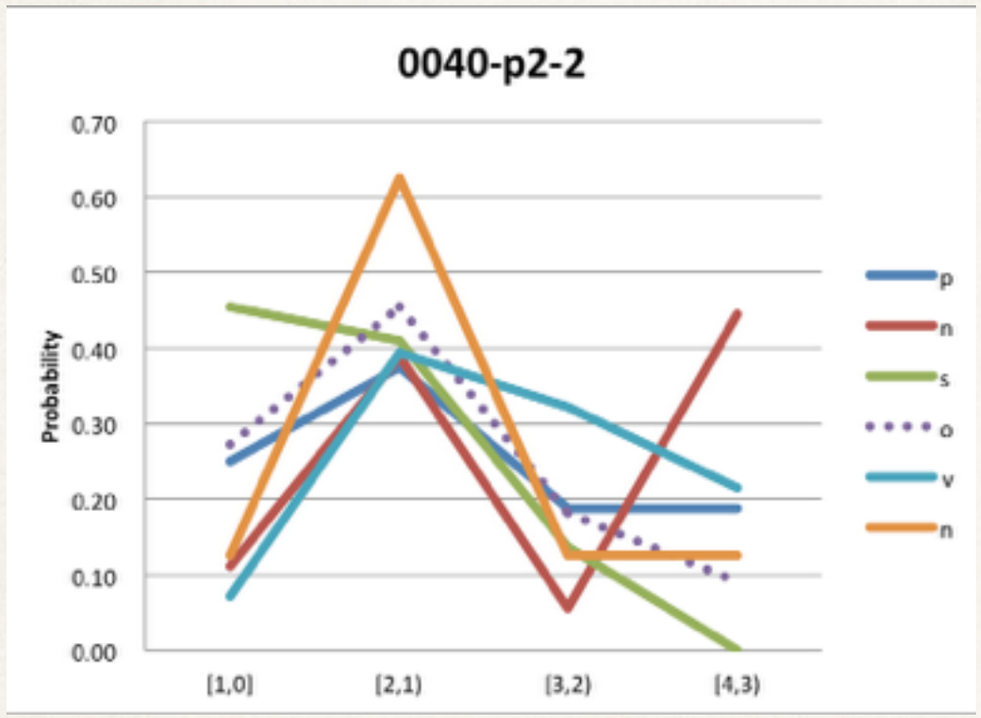


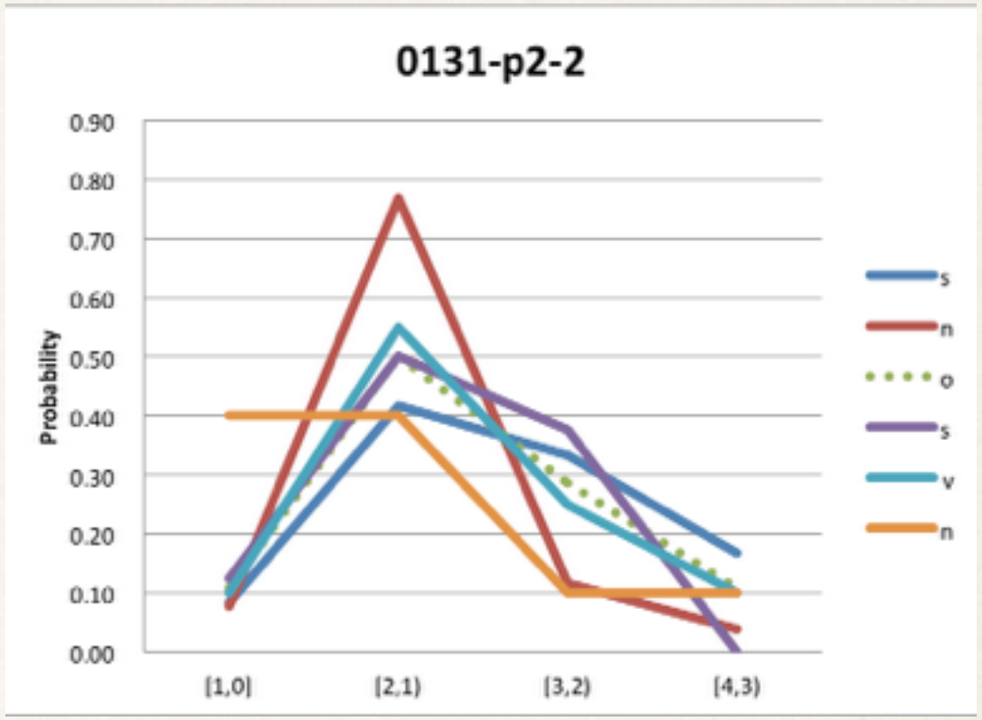


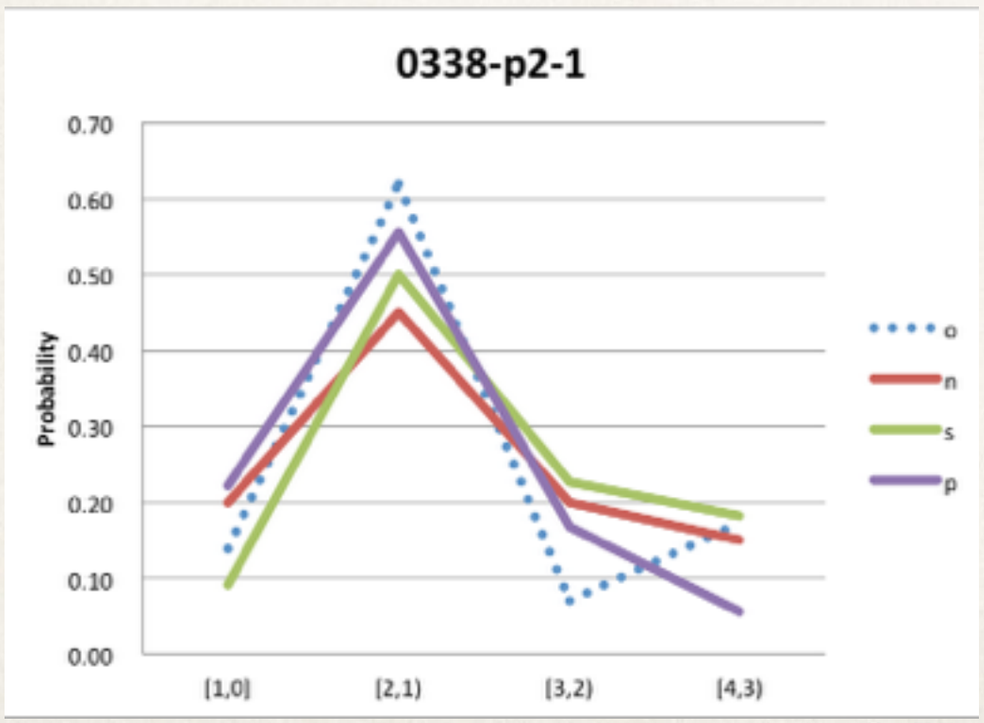




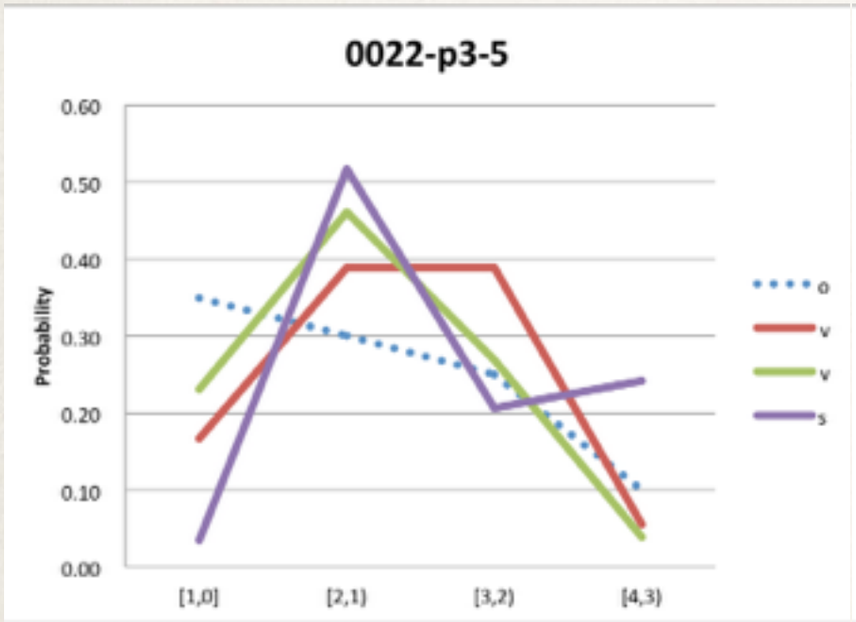
Individual responses to P2

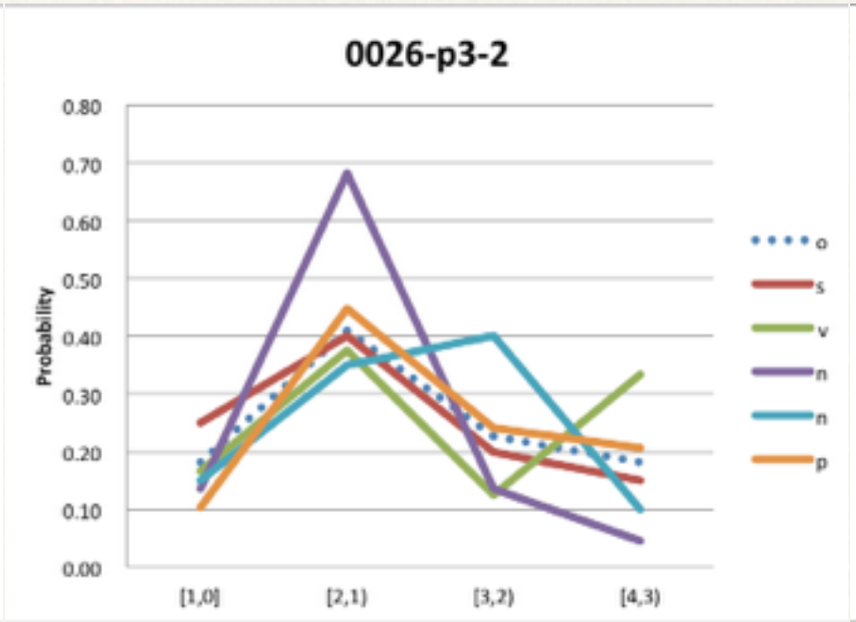


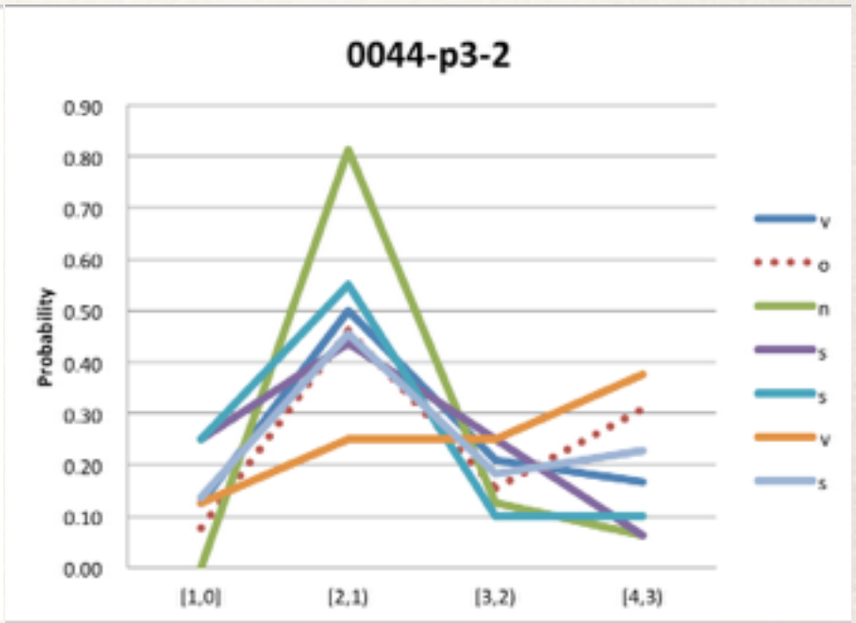


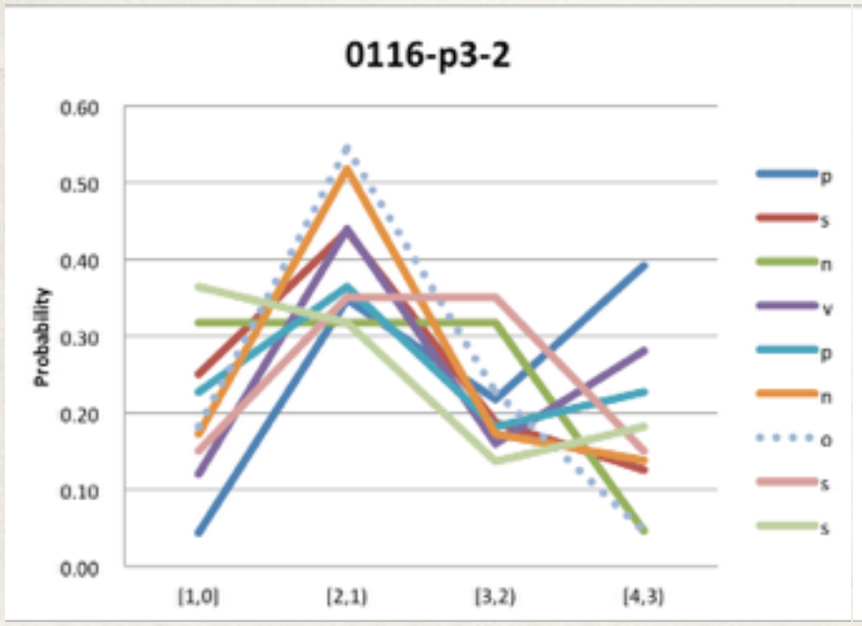


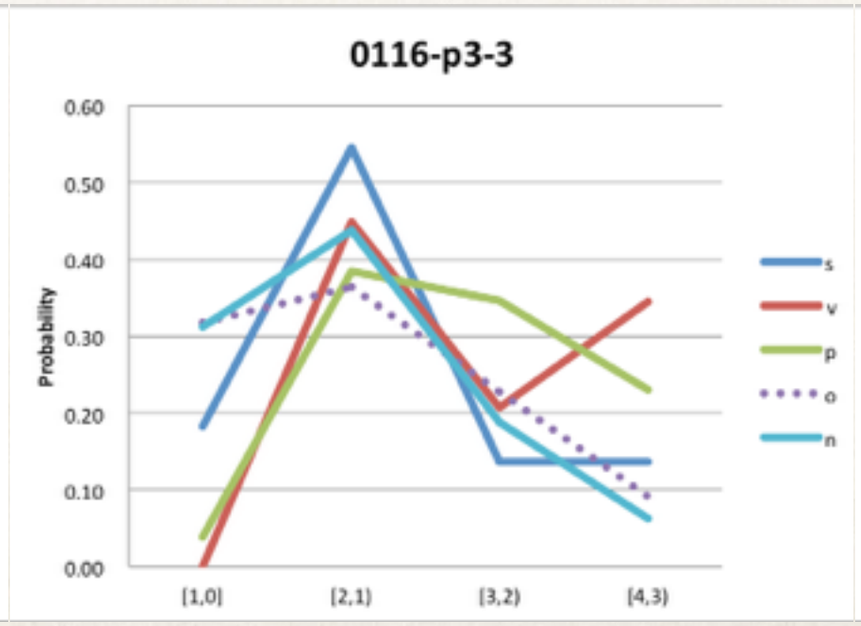
Individual responses to P3 1/2

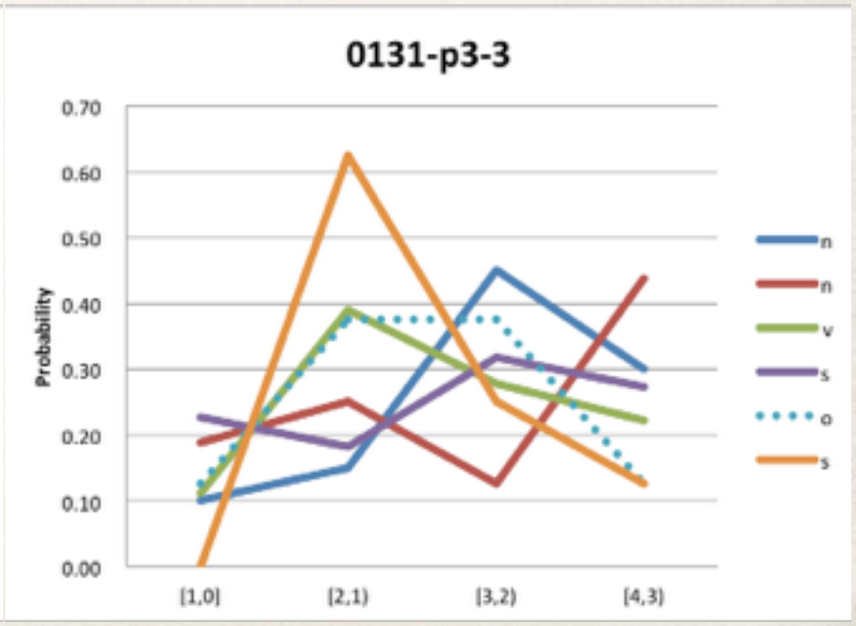




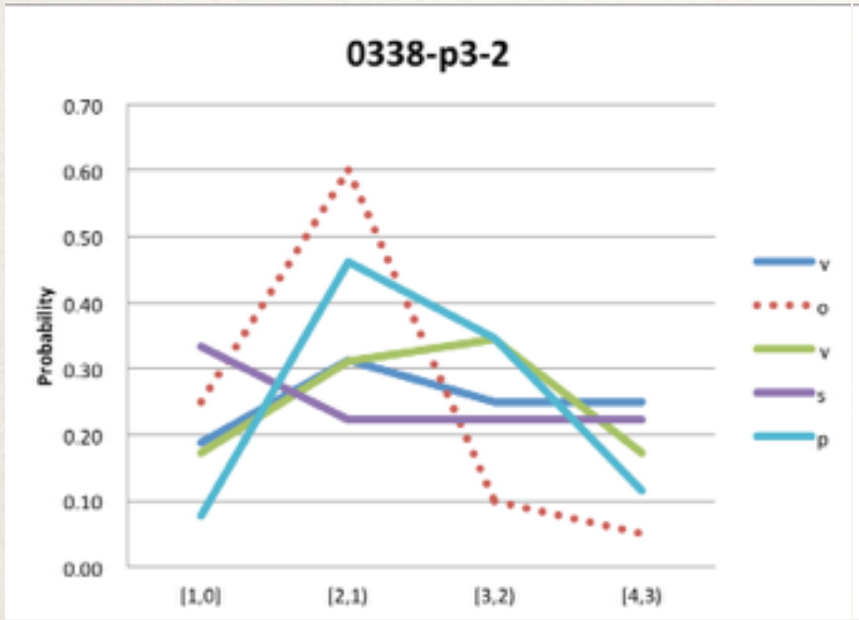


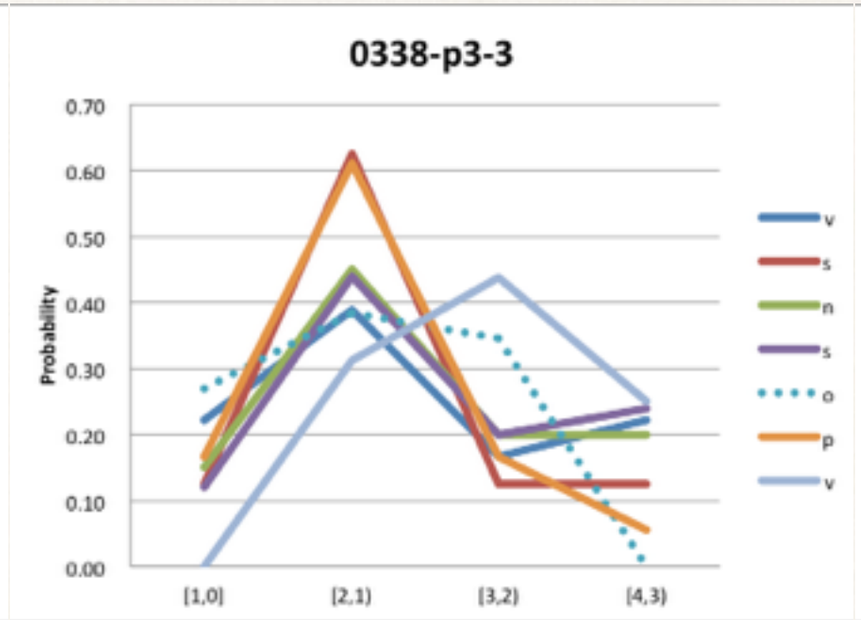


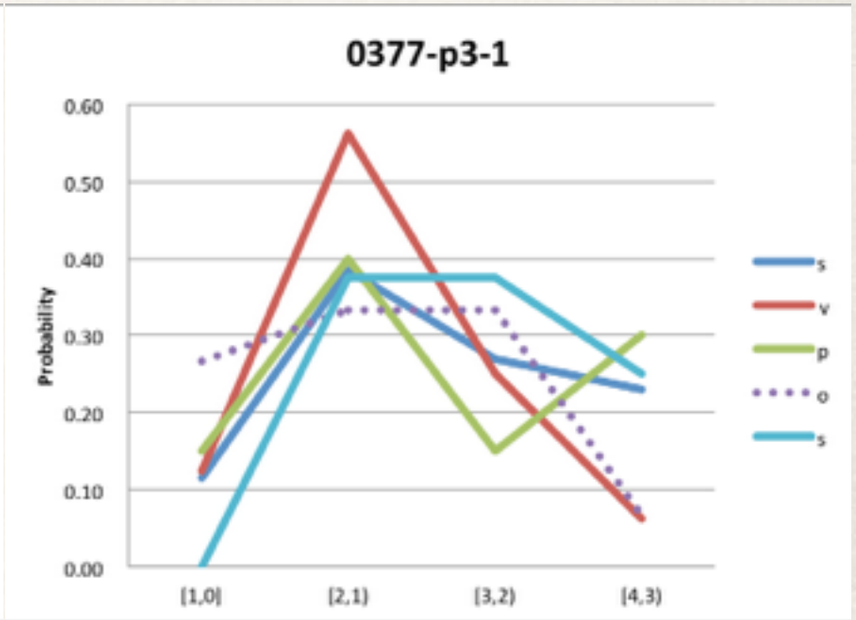


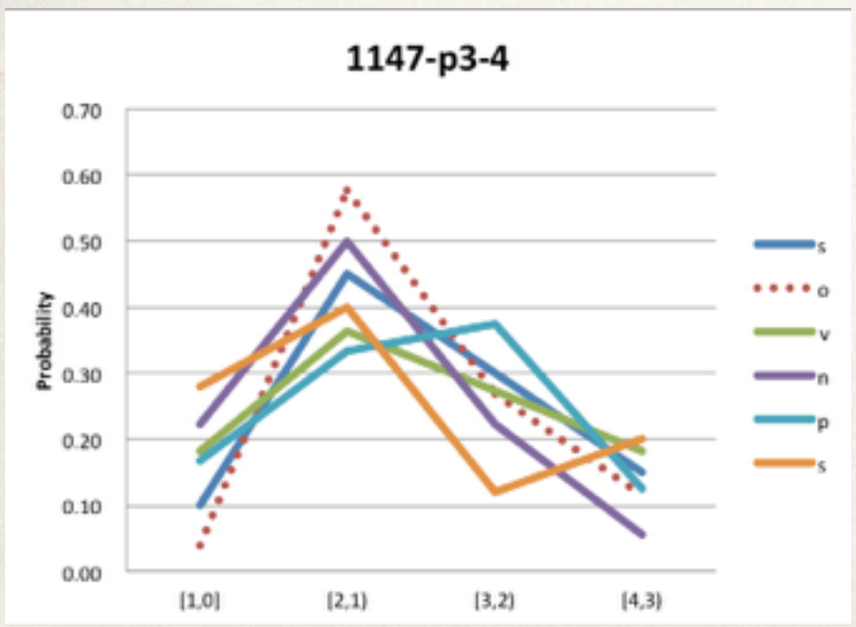


Individual responses to P3 2/2

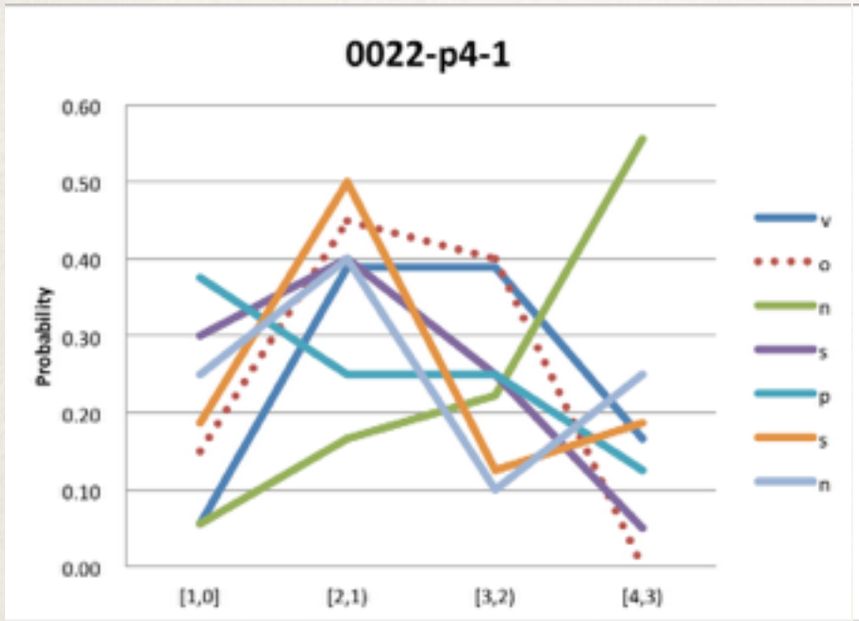


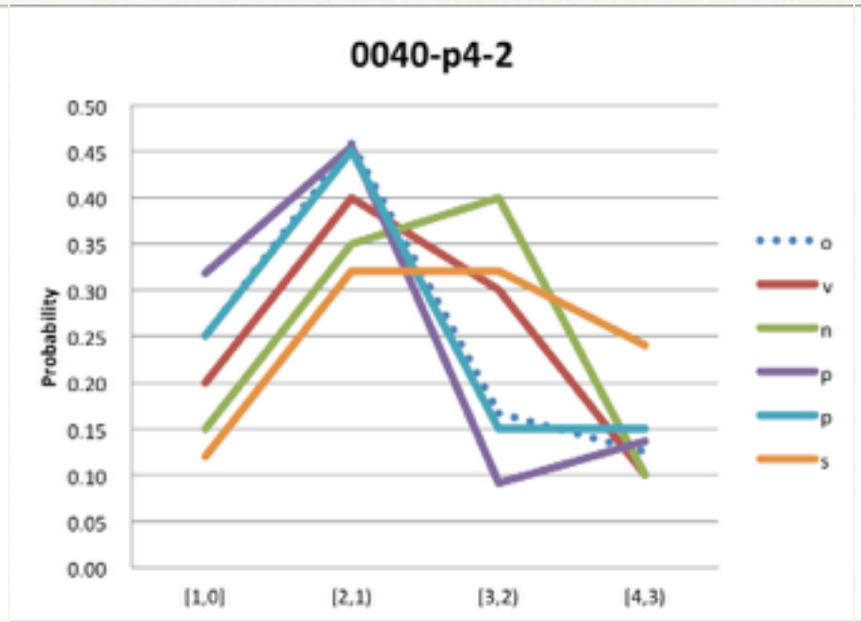


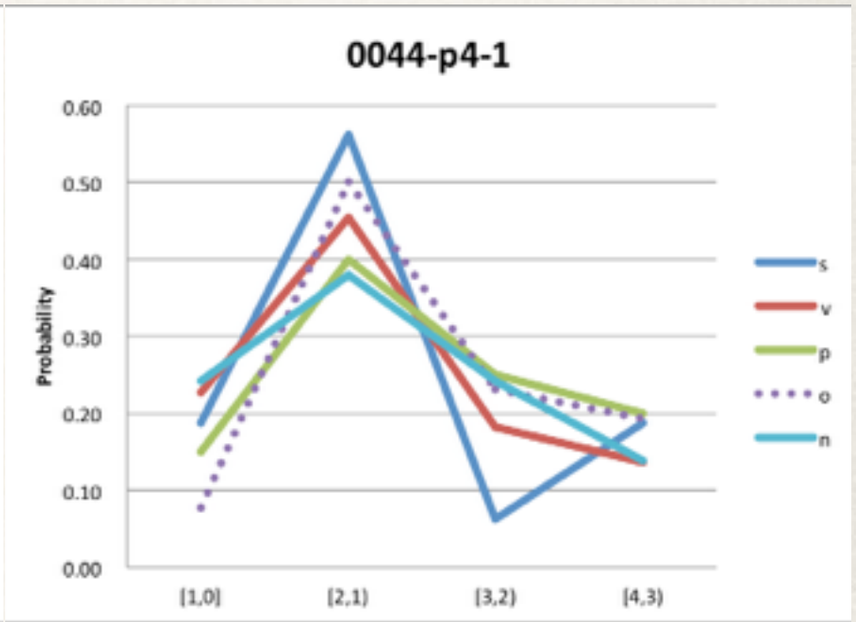


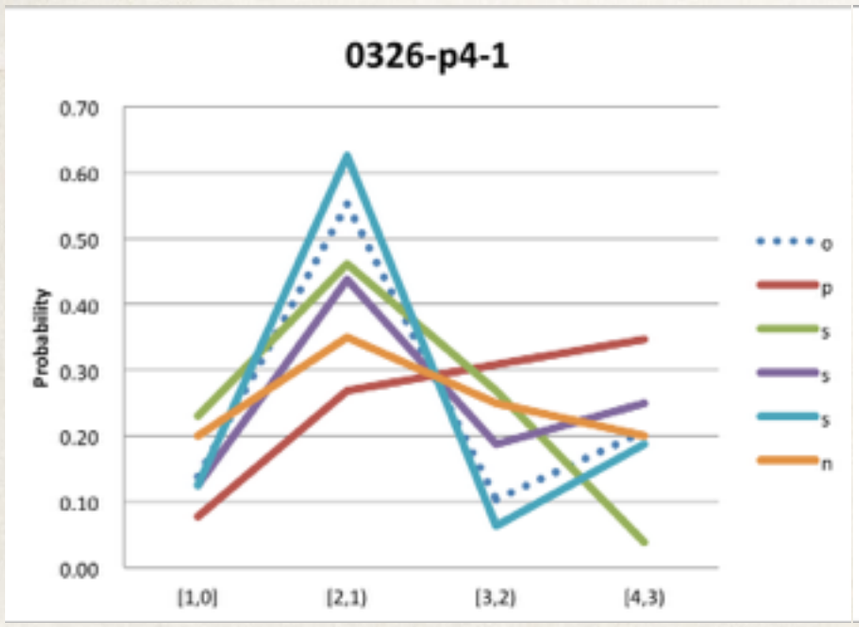


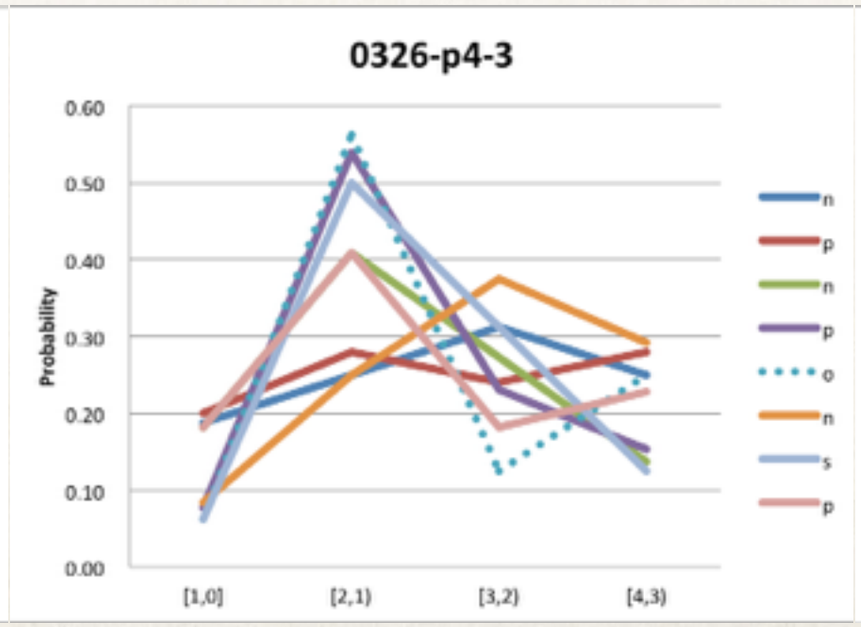
Individual responses to P4 1/2

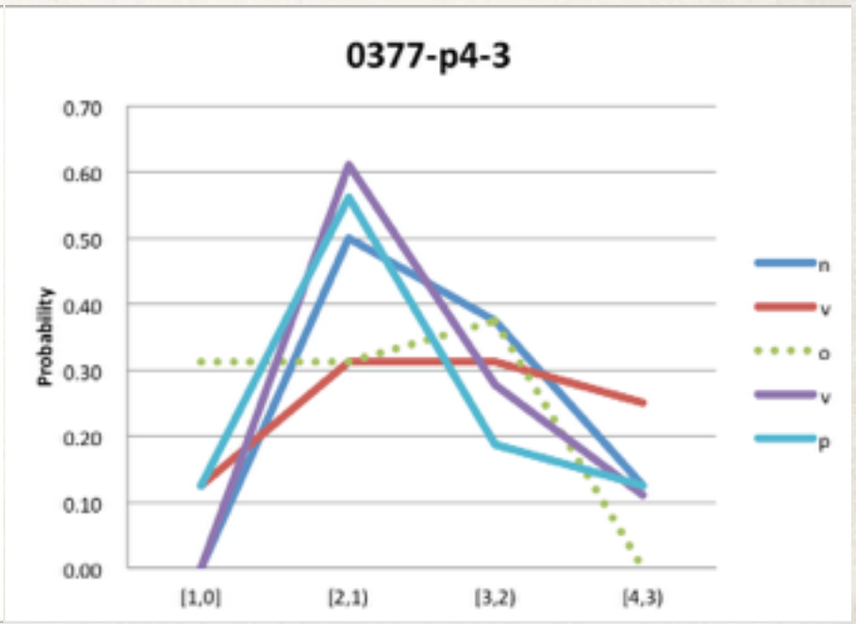




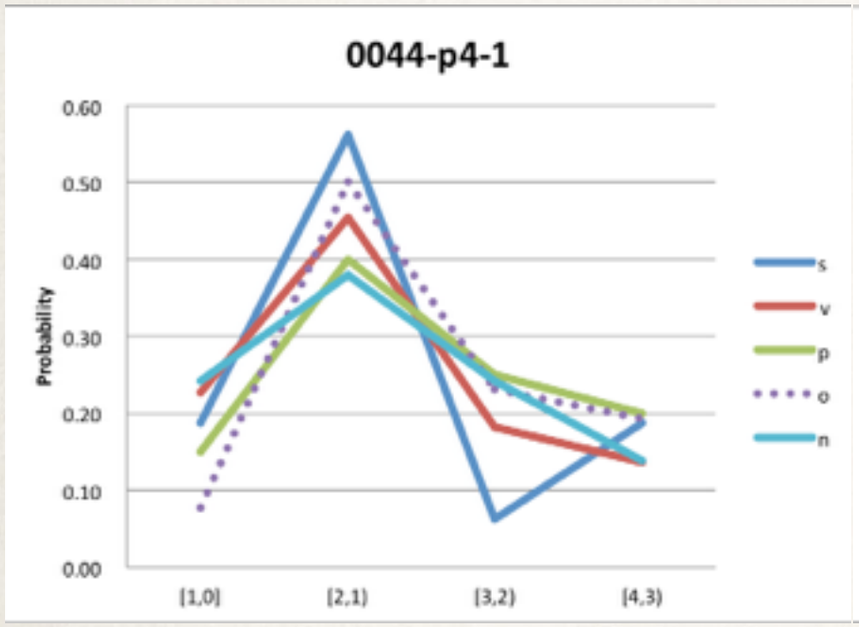


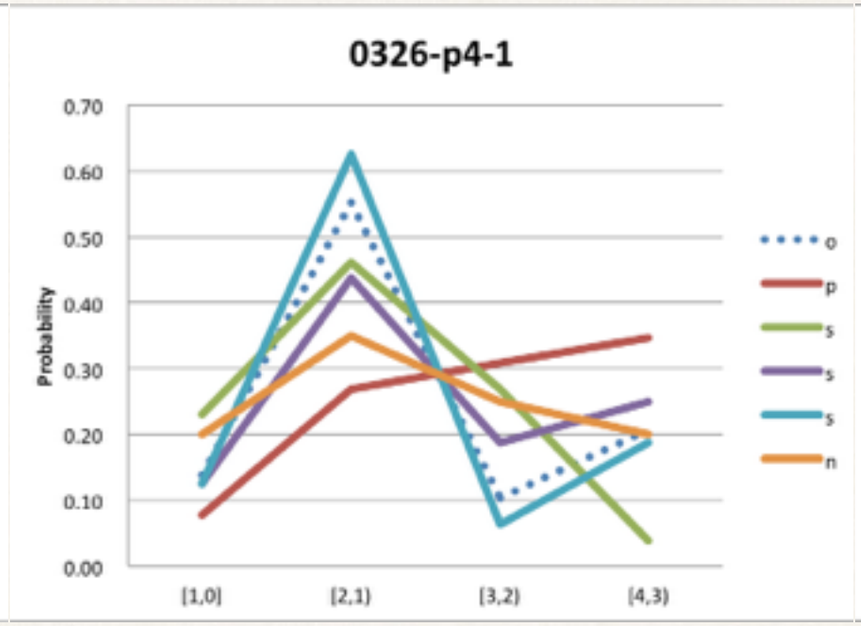


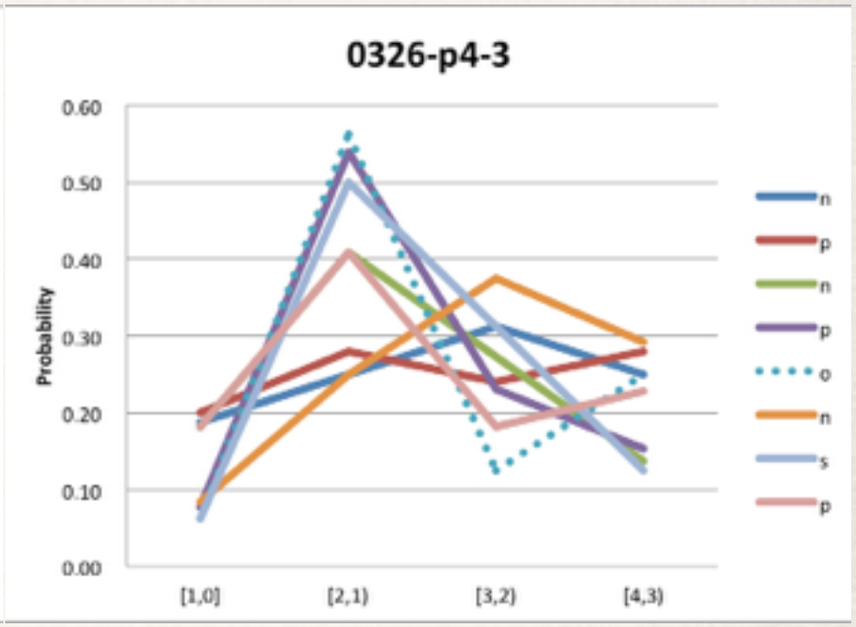


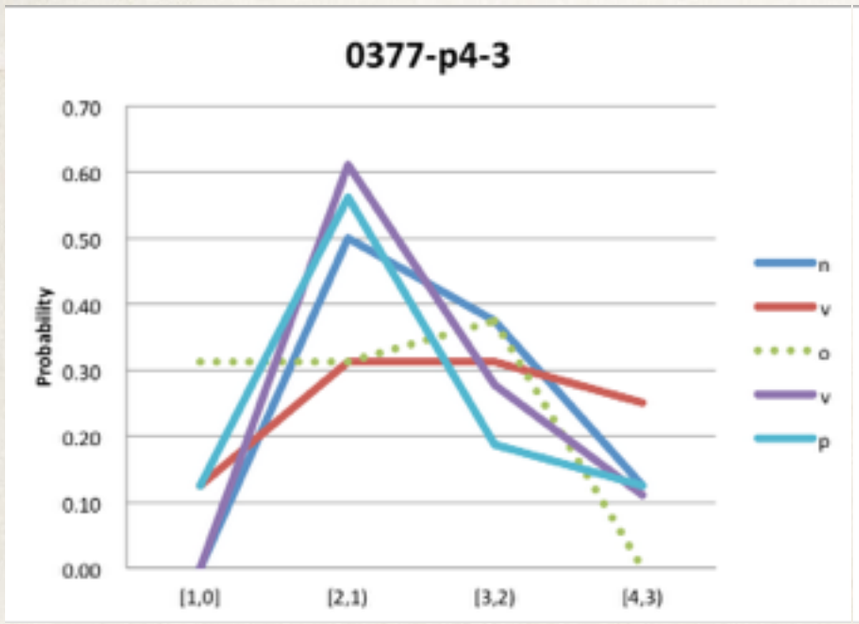


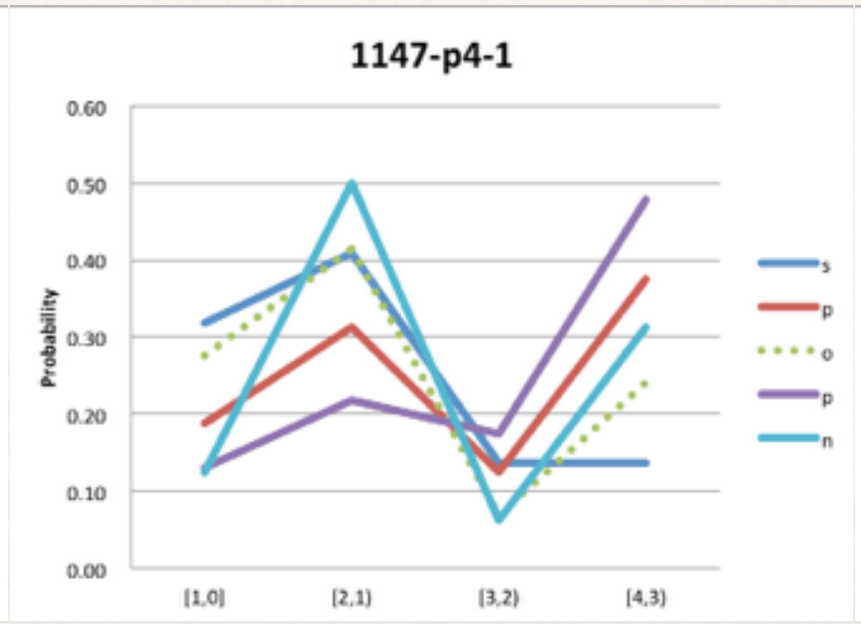
Individual responses to P4 2/2

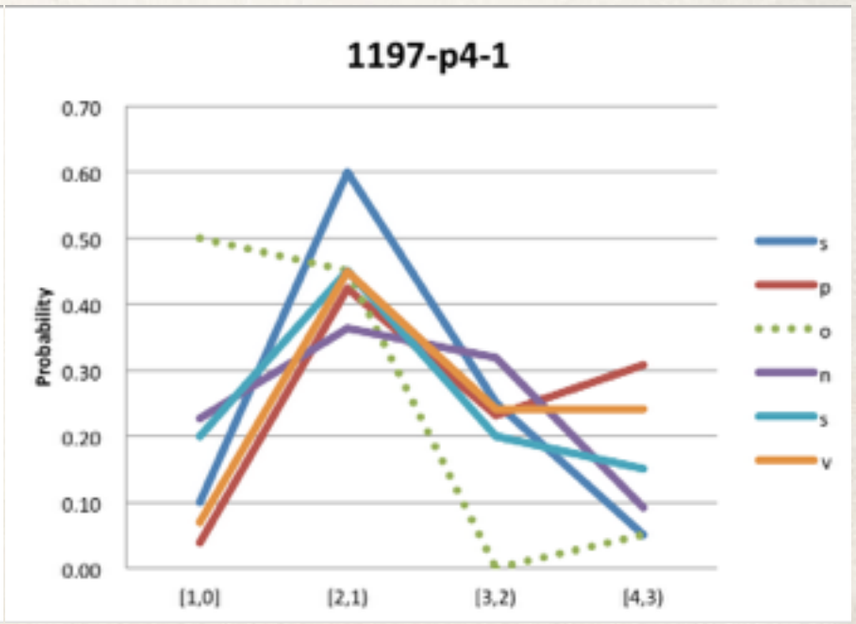






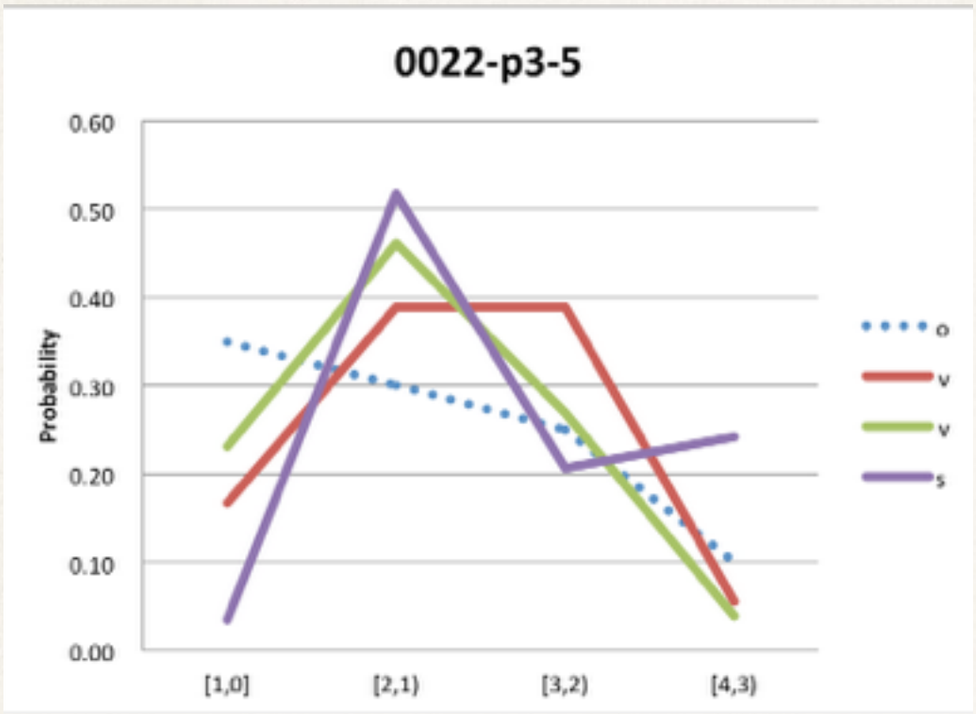


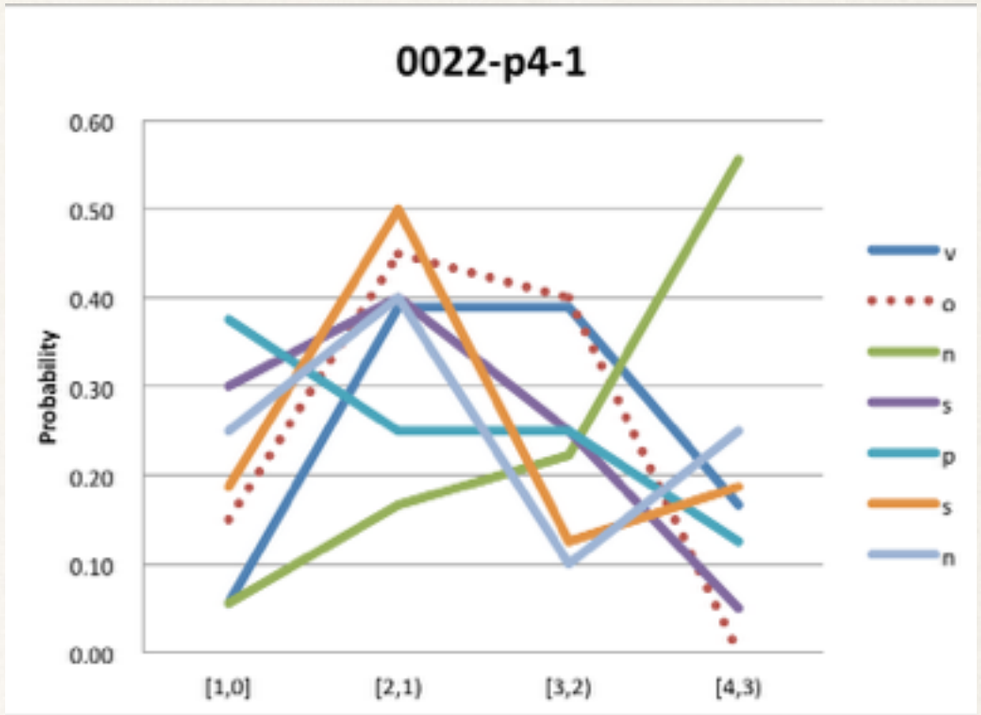




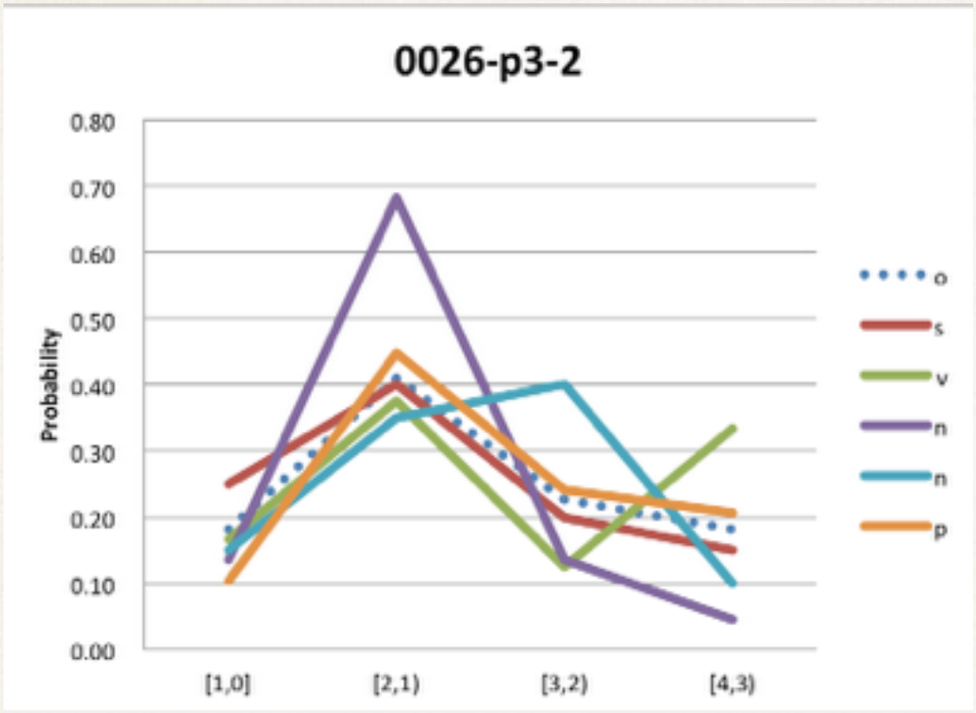
Responses to each stimulus

Individual responses ID=0022 (V= *iku* (go))

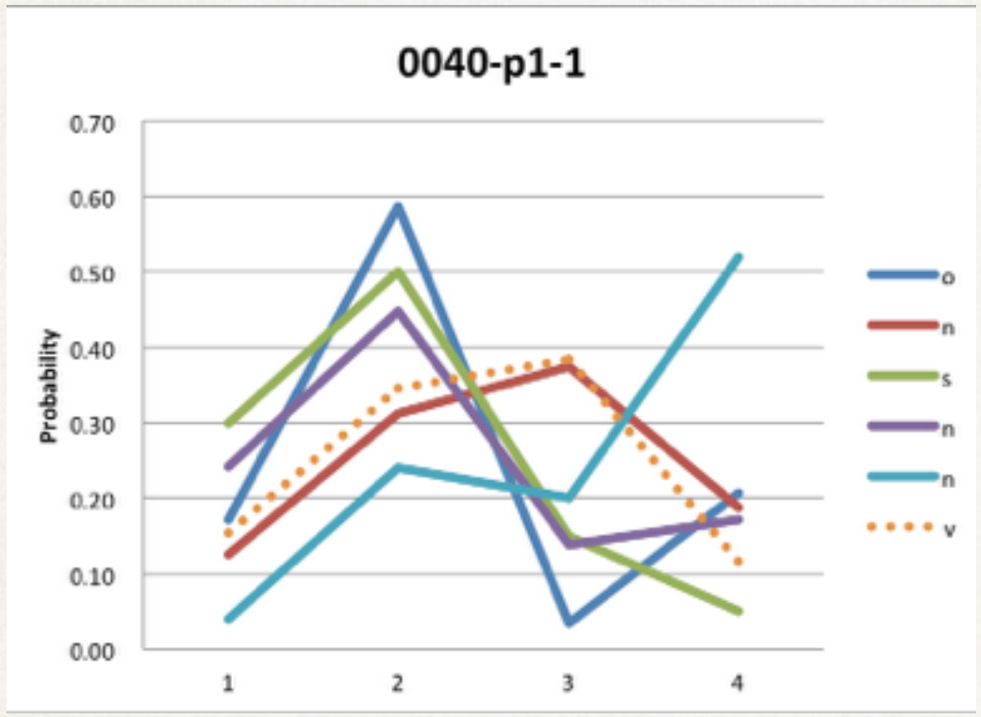


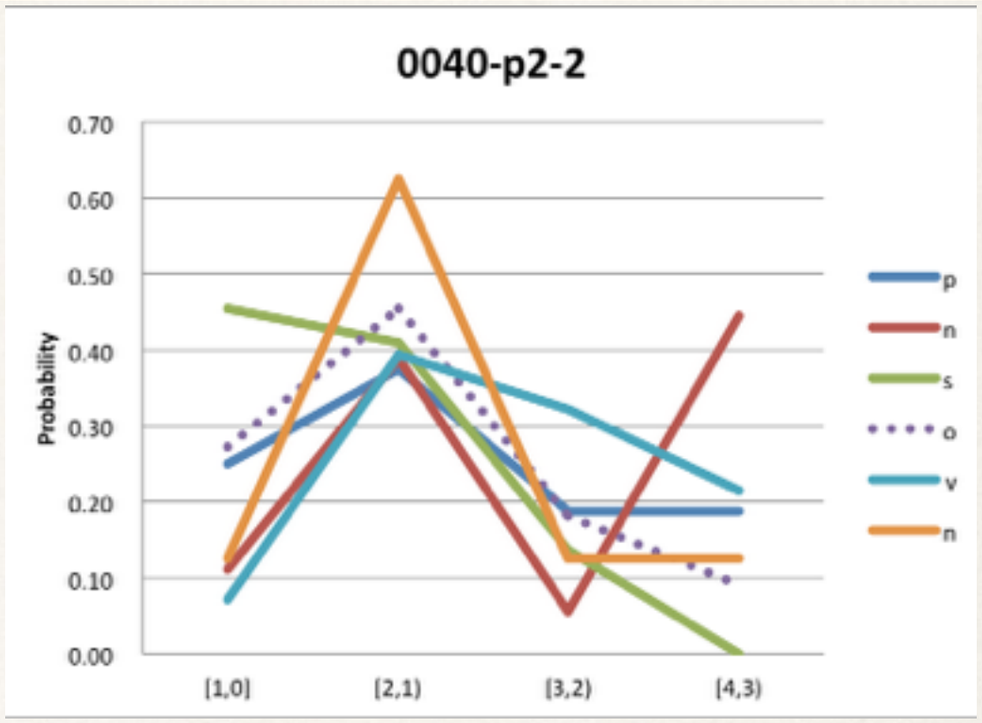


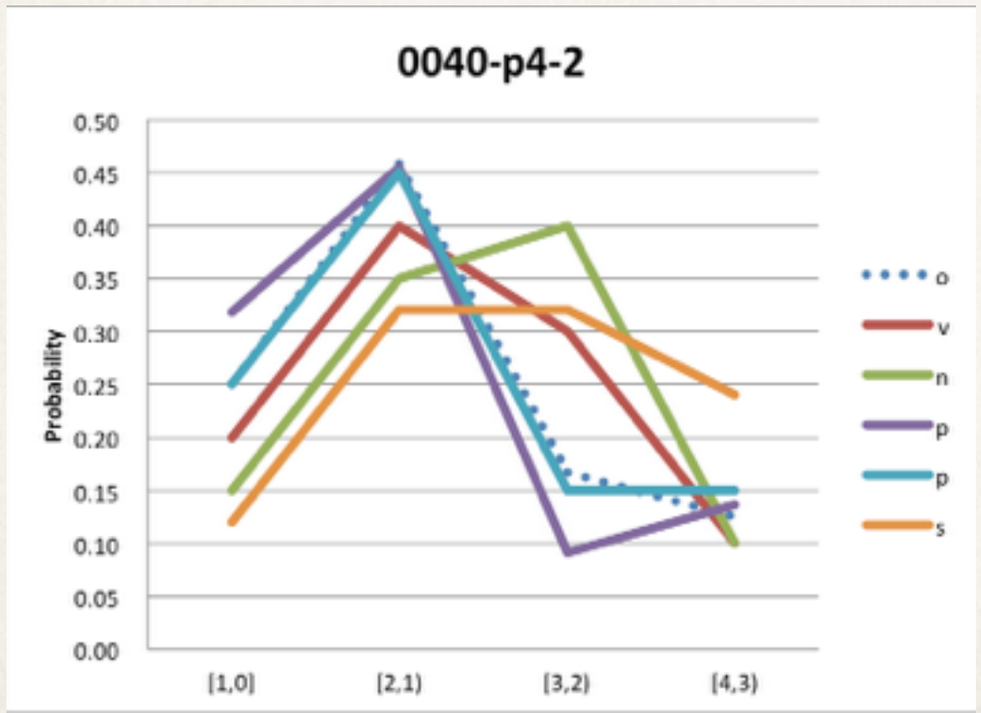
Individual responses ID=0026 (V= *shiru* (know))



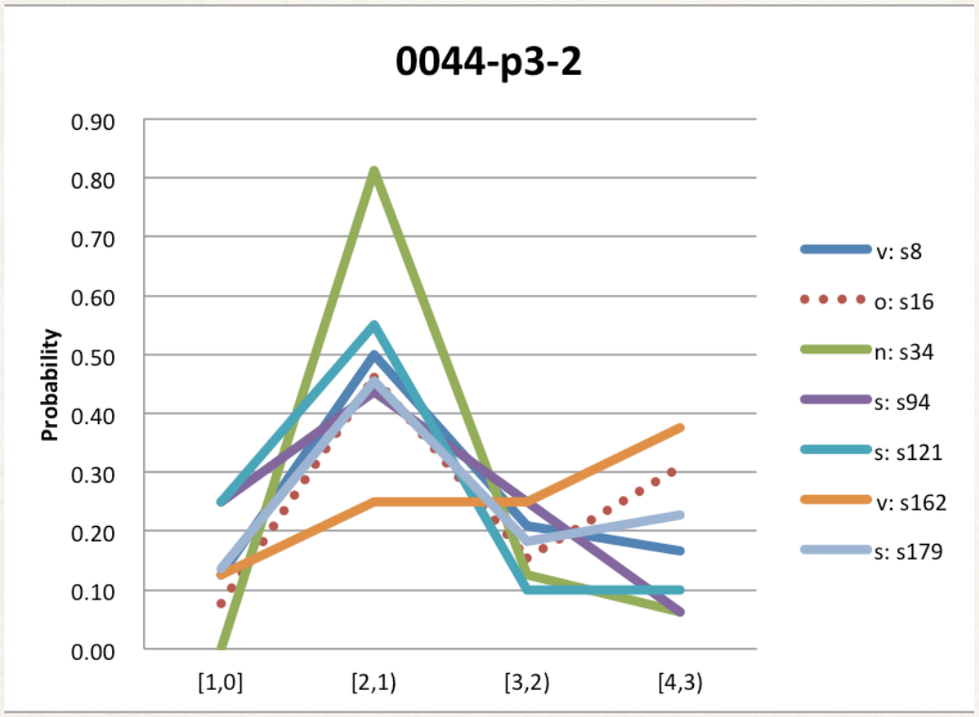
Individual responses ID=0040 (V= *oshieru* (teach))

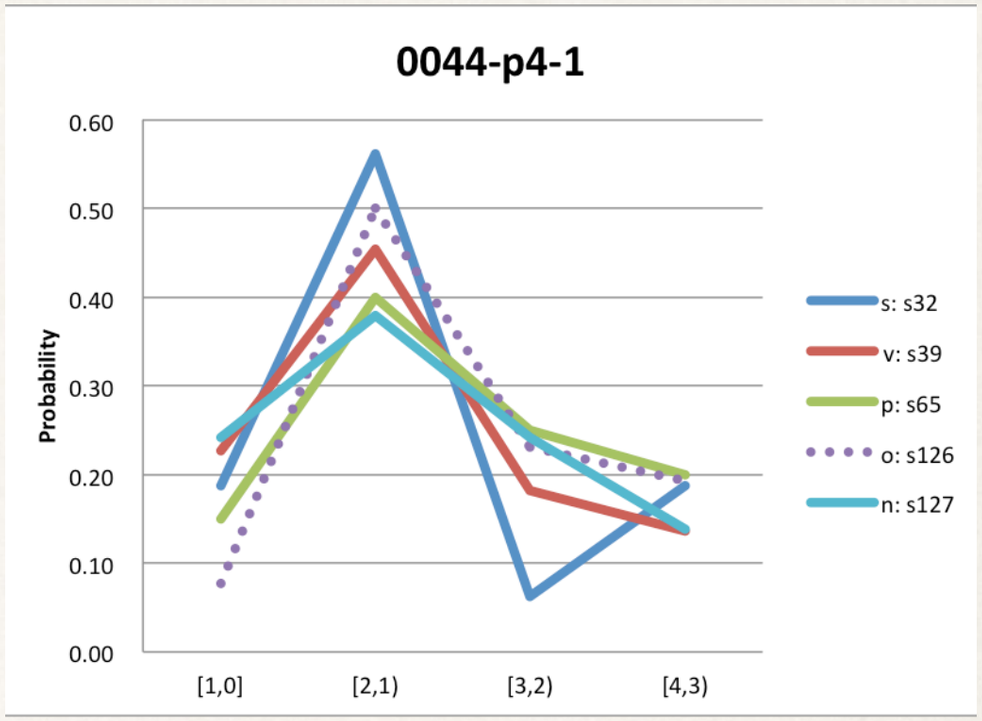




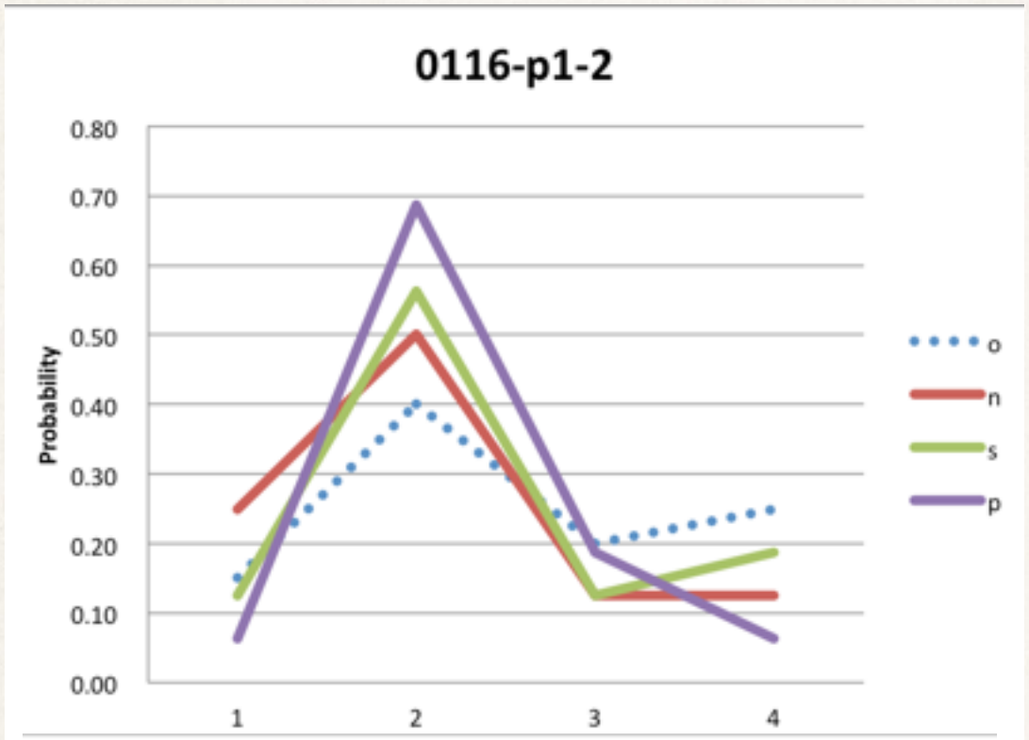


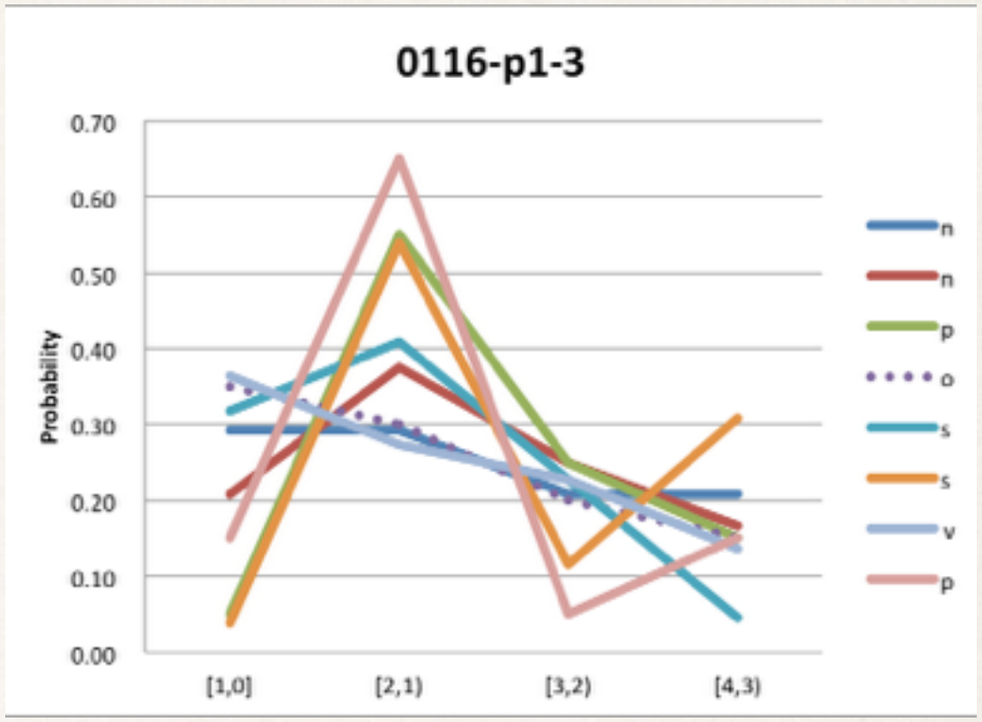
Individual responses ID=0044 (V= *kanjiru* (feel))

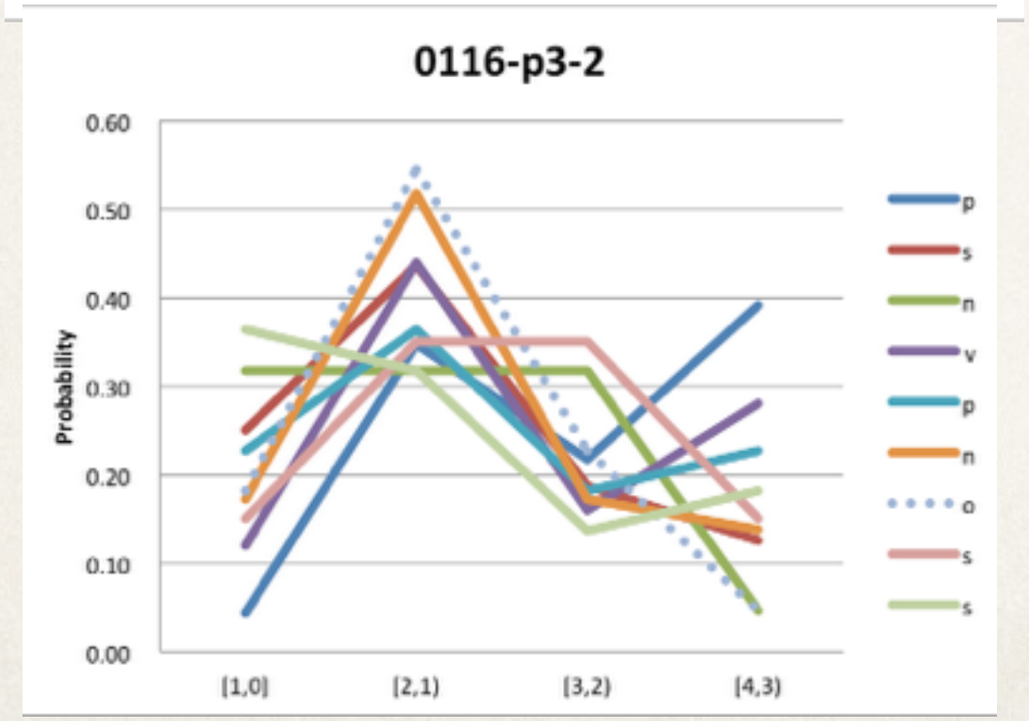


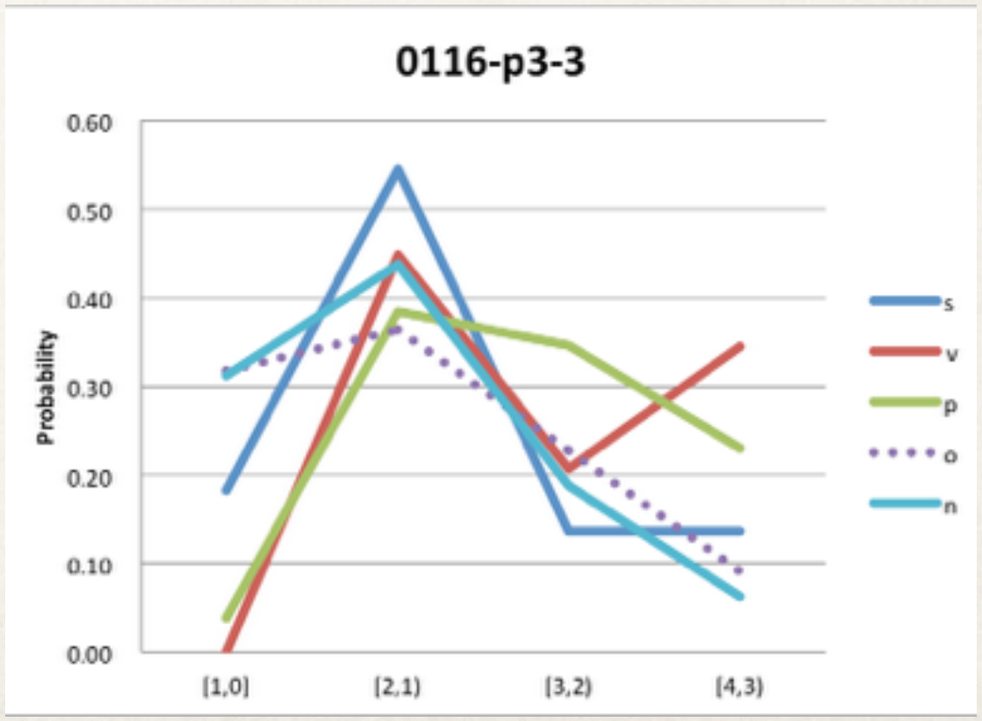


Individual responses ID=0116 (V= *kotaeru* (answer))

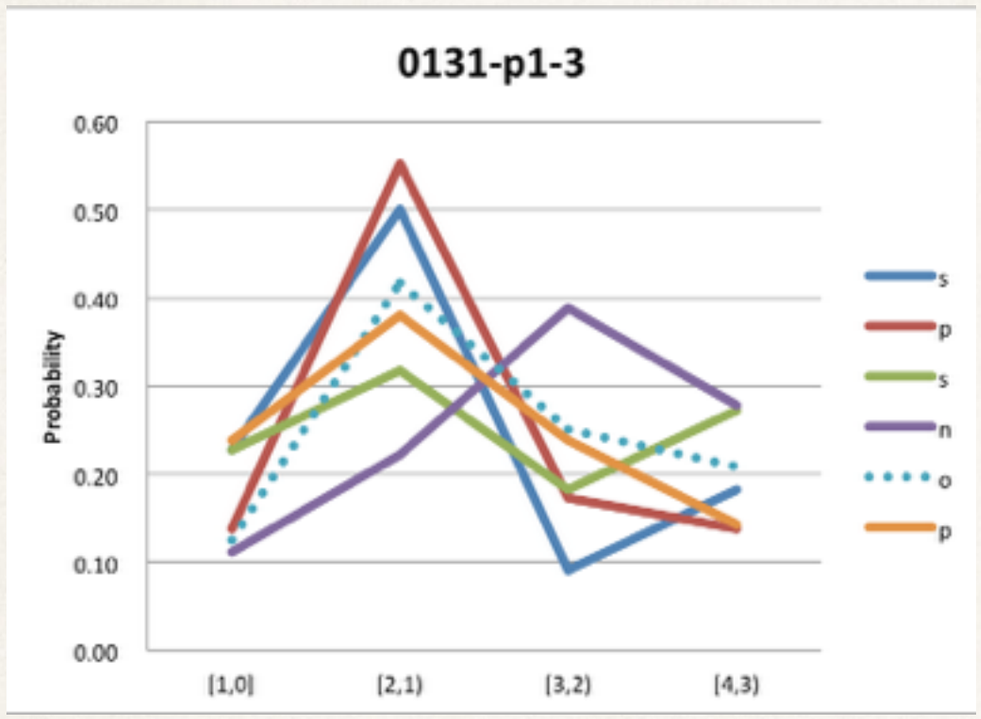


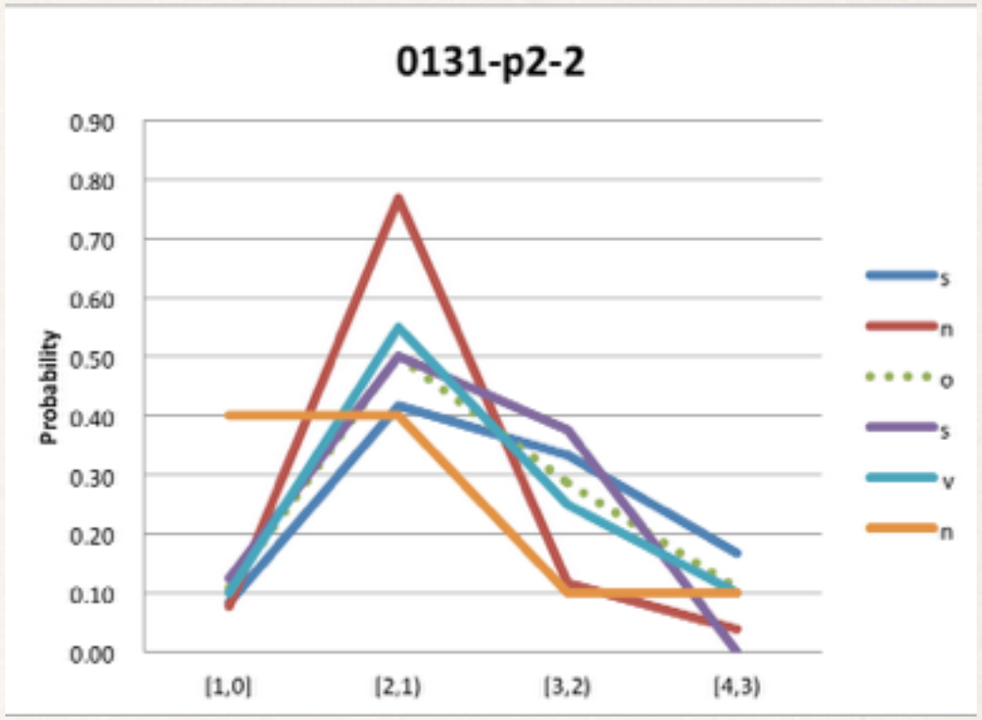


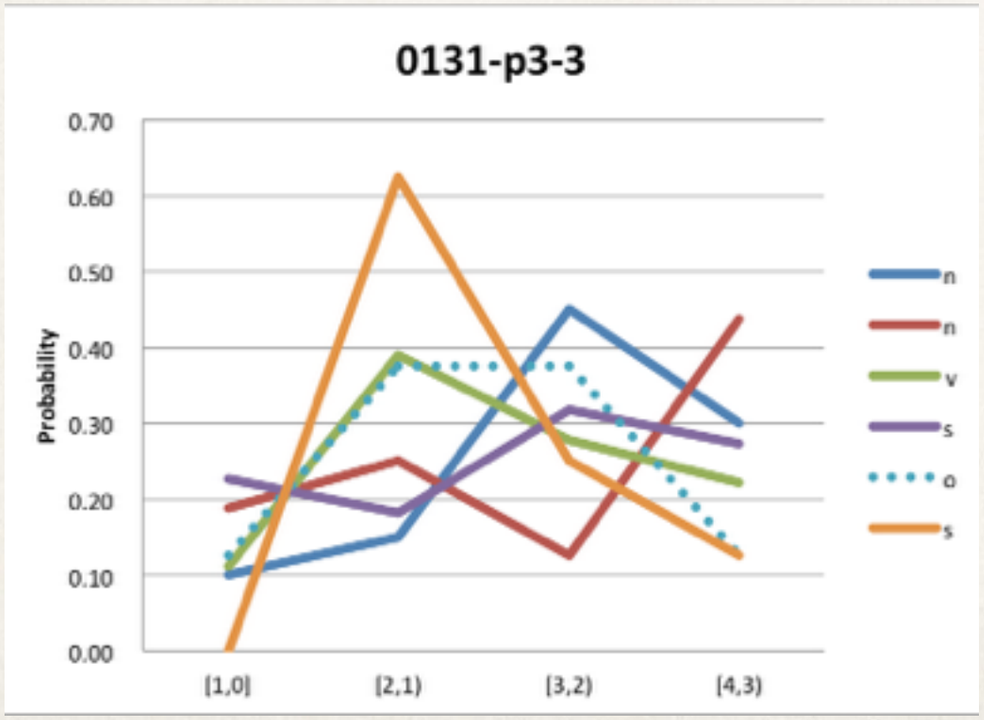




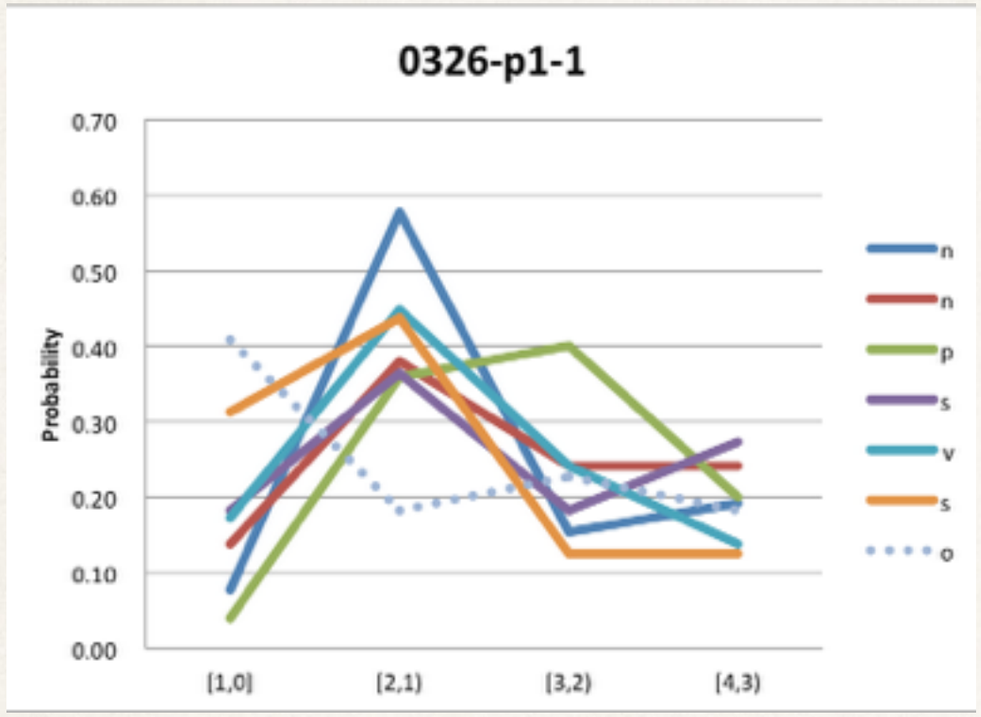
Individual responses ID=0131 (V= *sagasu* (search))

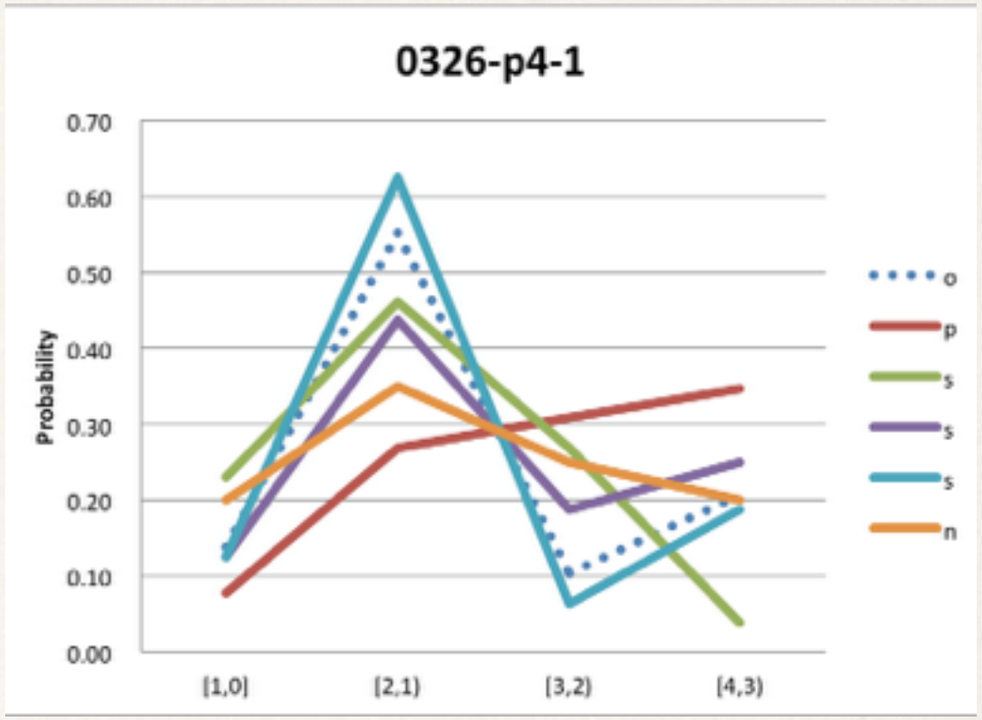


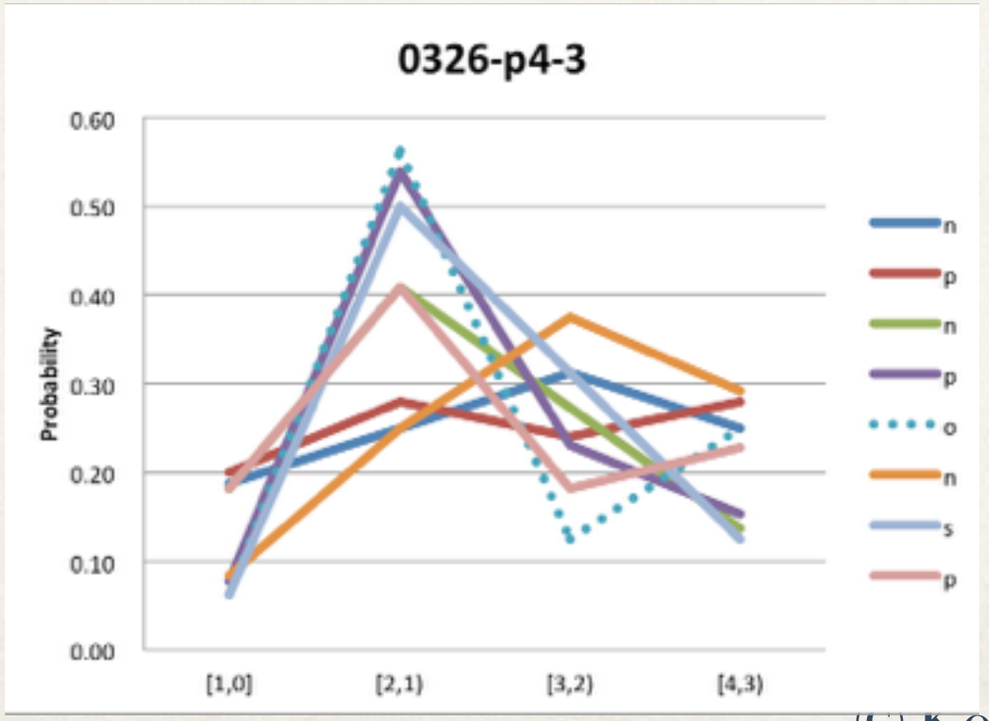




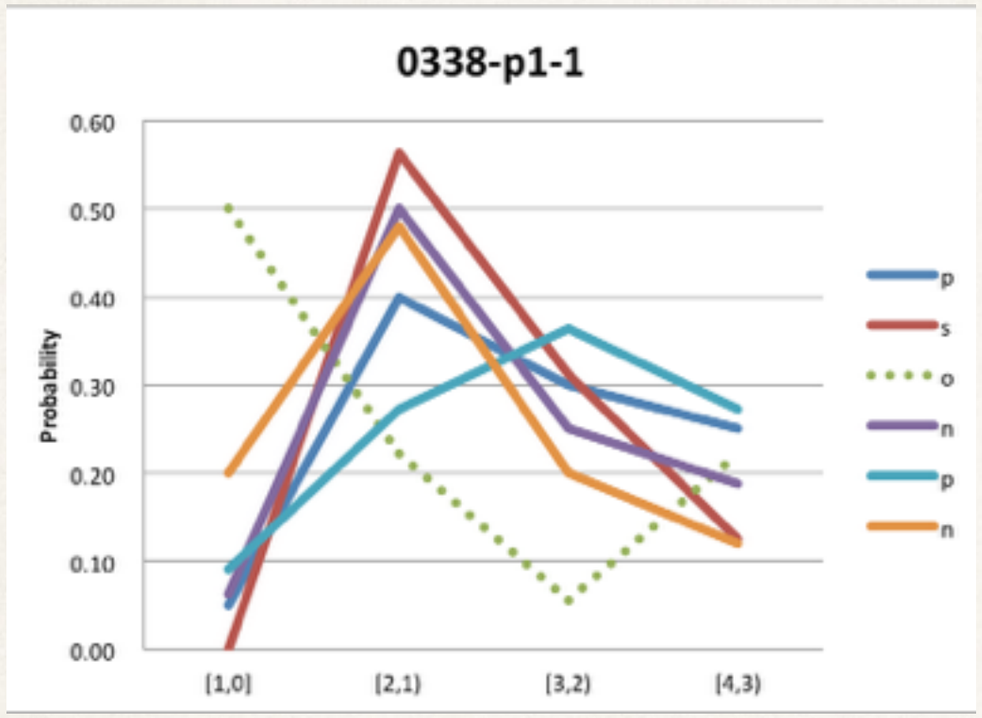
Individual responses ID=0326 (V= *damaru* (become silent))

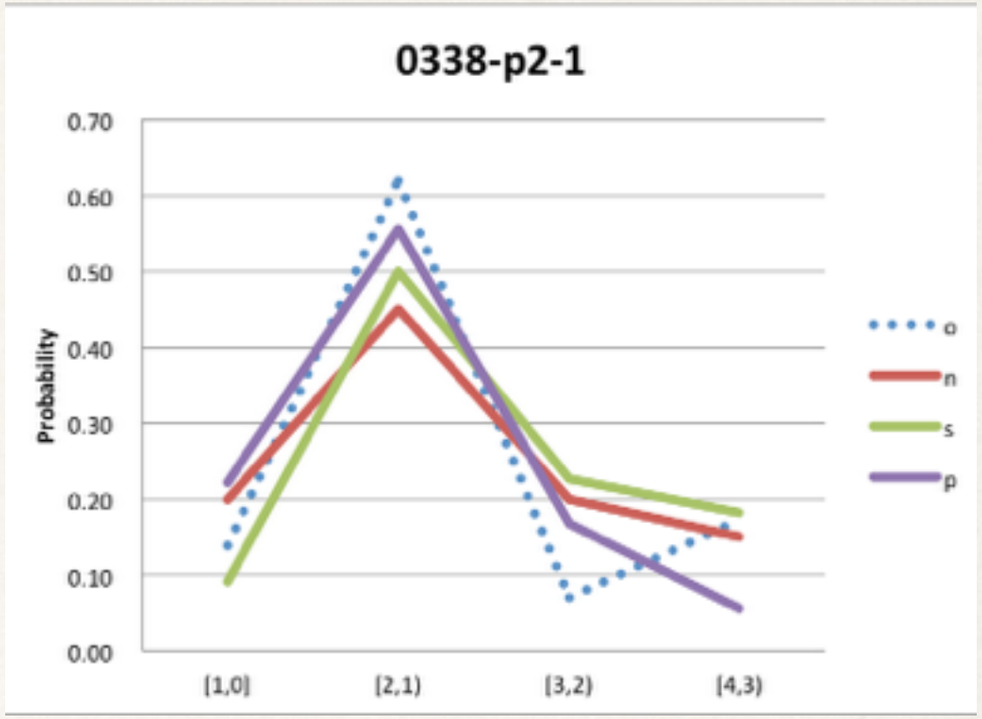


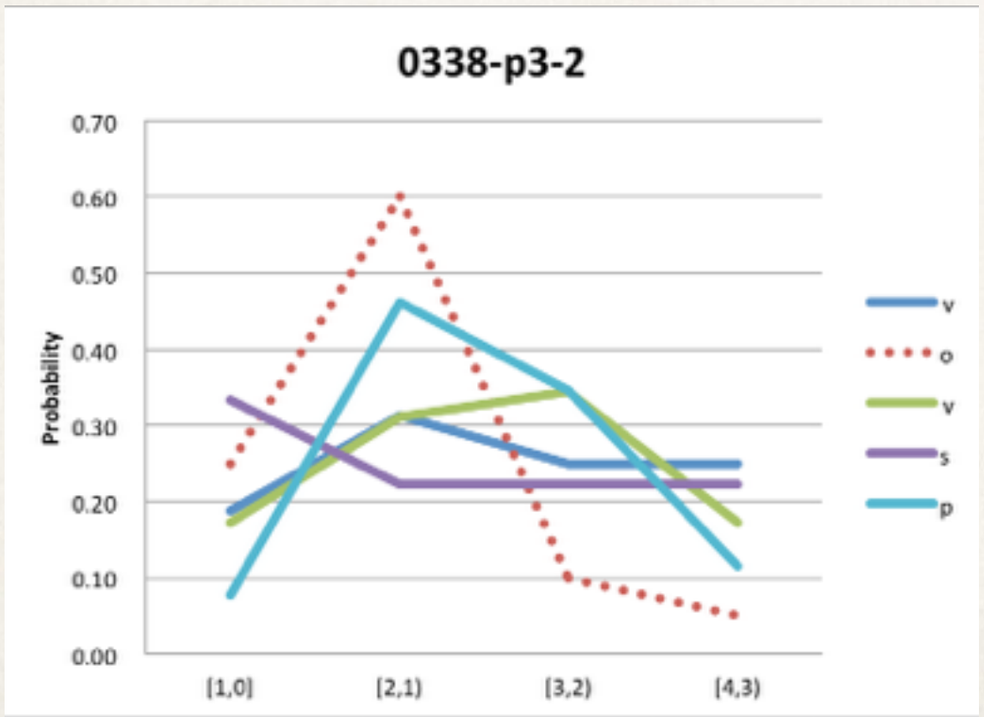


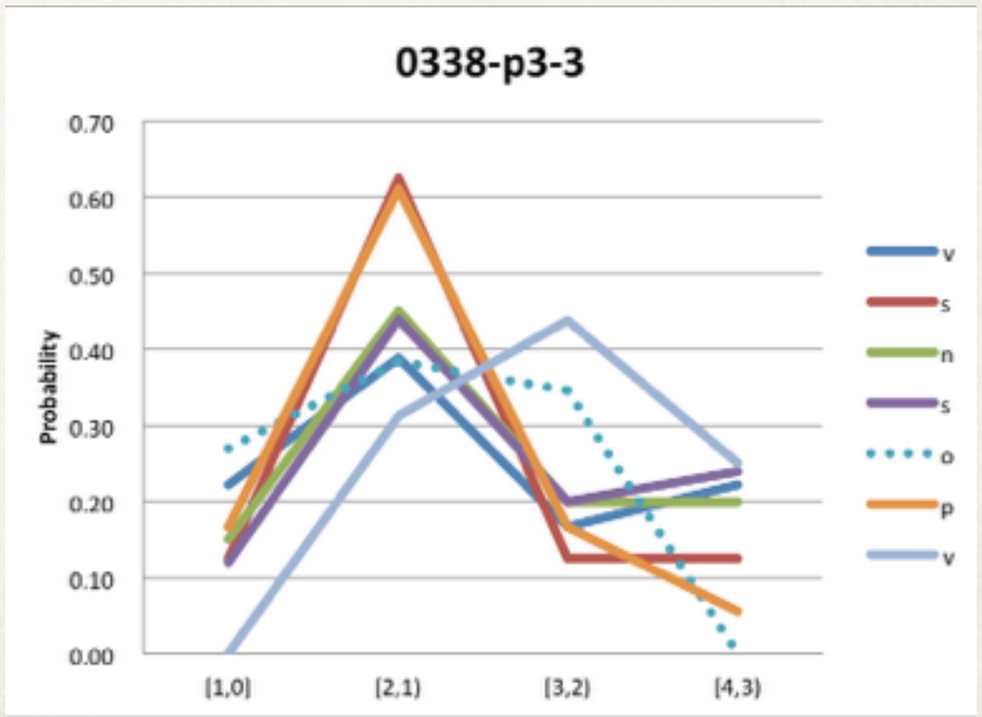


Individual responses ID=0338 (V= *makeru* (lose))

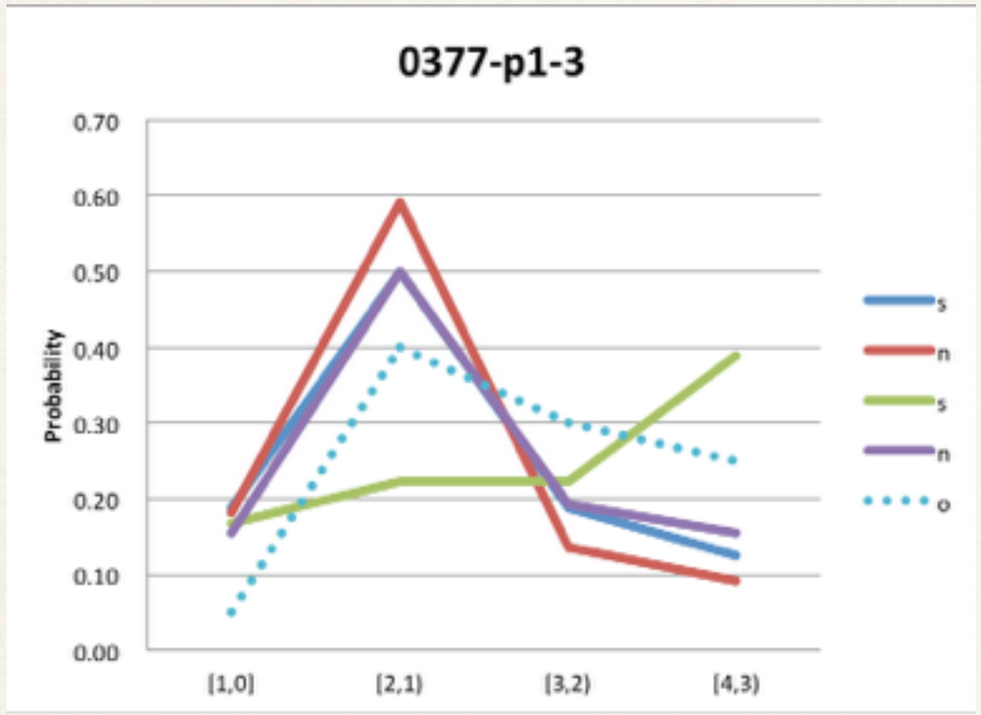


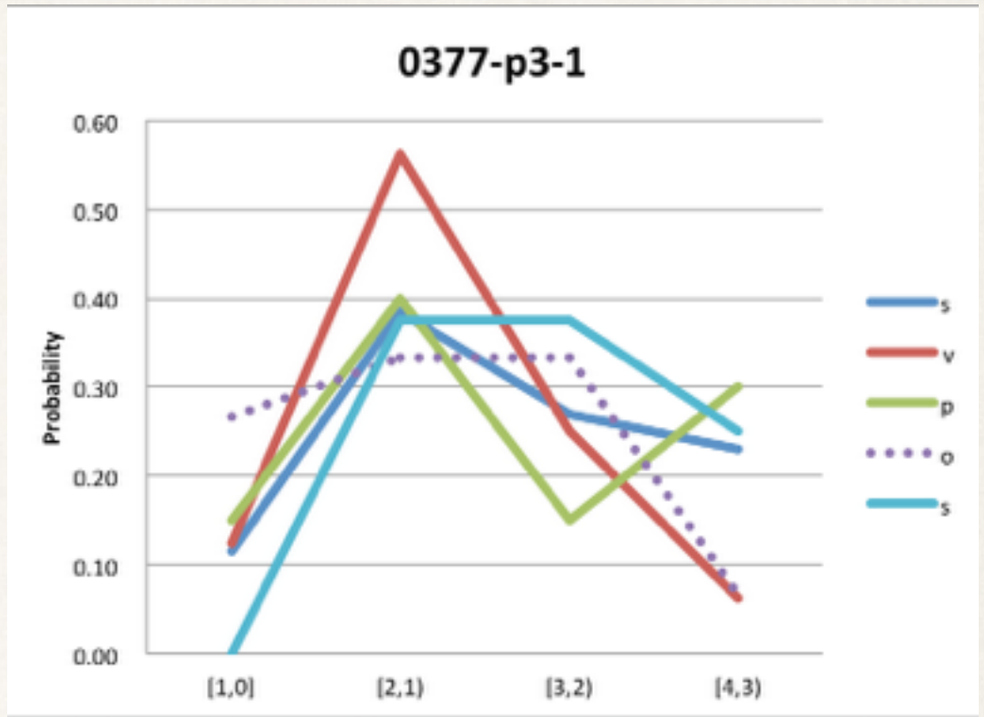


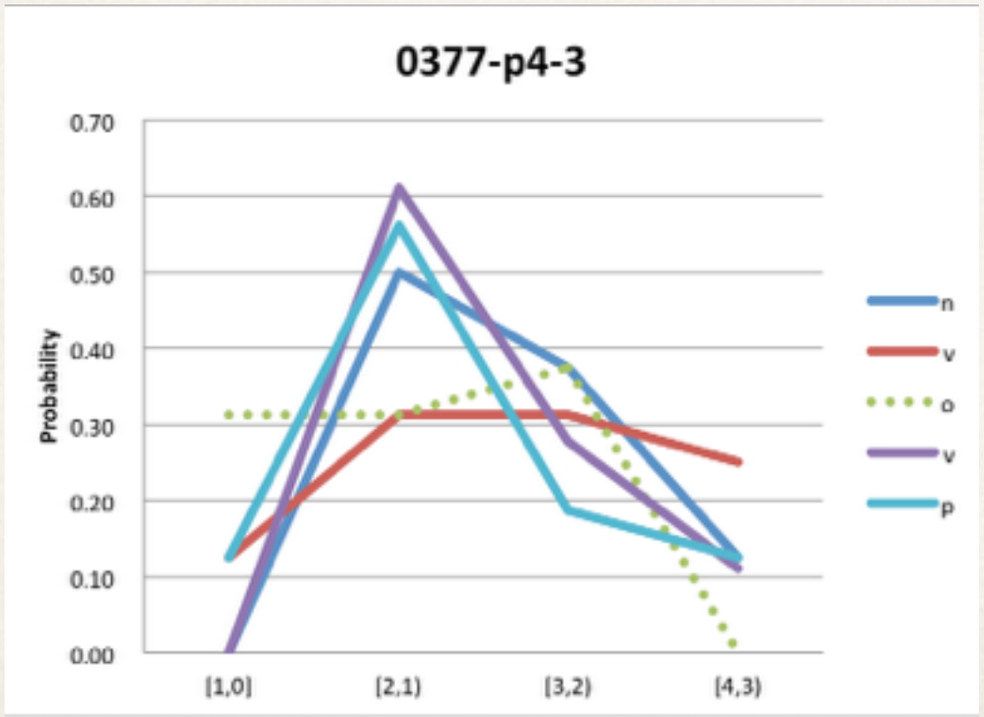




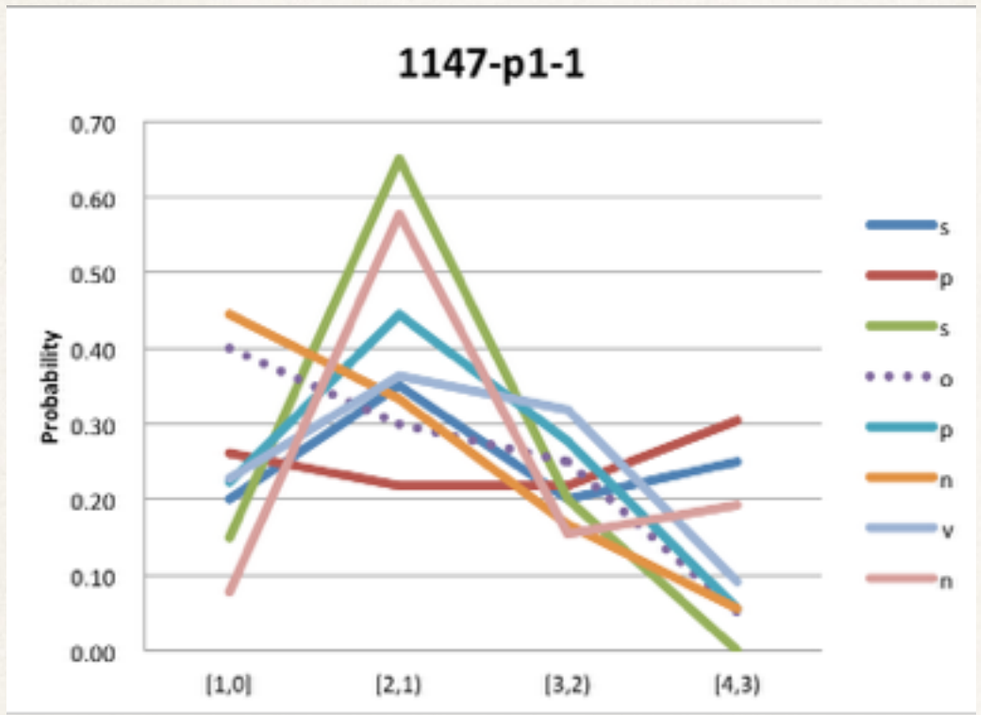
Individual responses ID=0377 (V= *tsutawaru* (spread, travel, get transmitted))

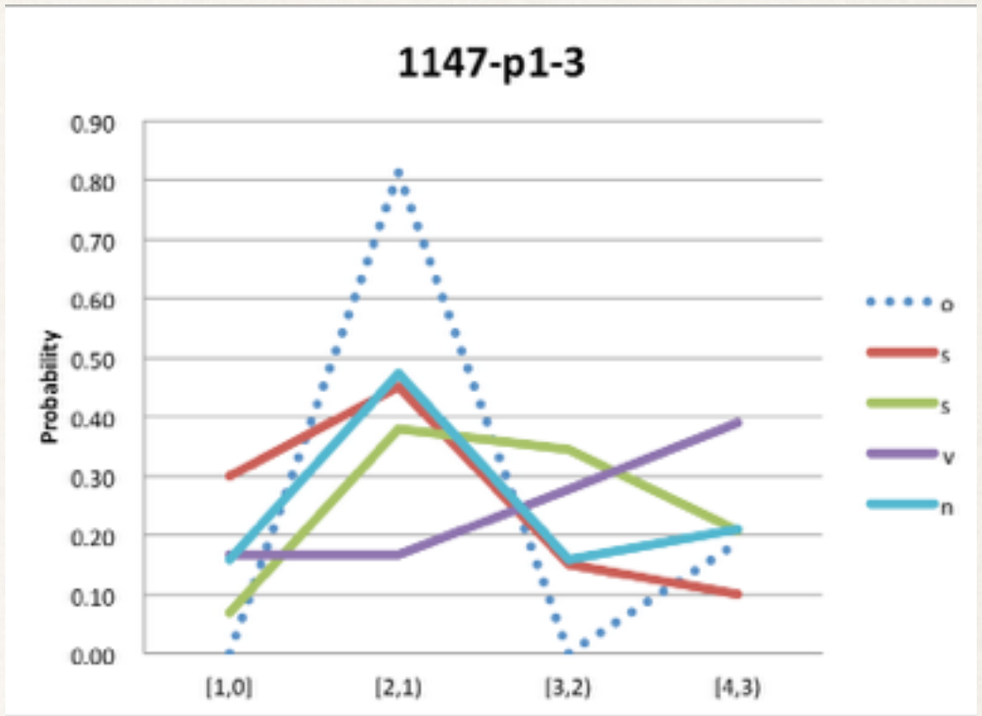


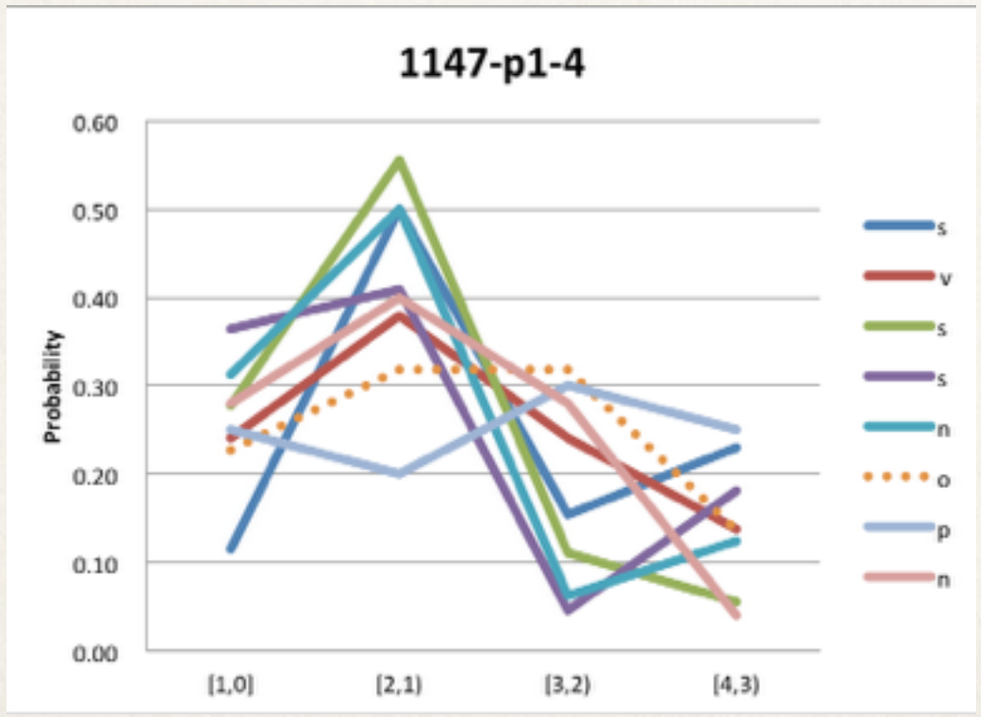


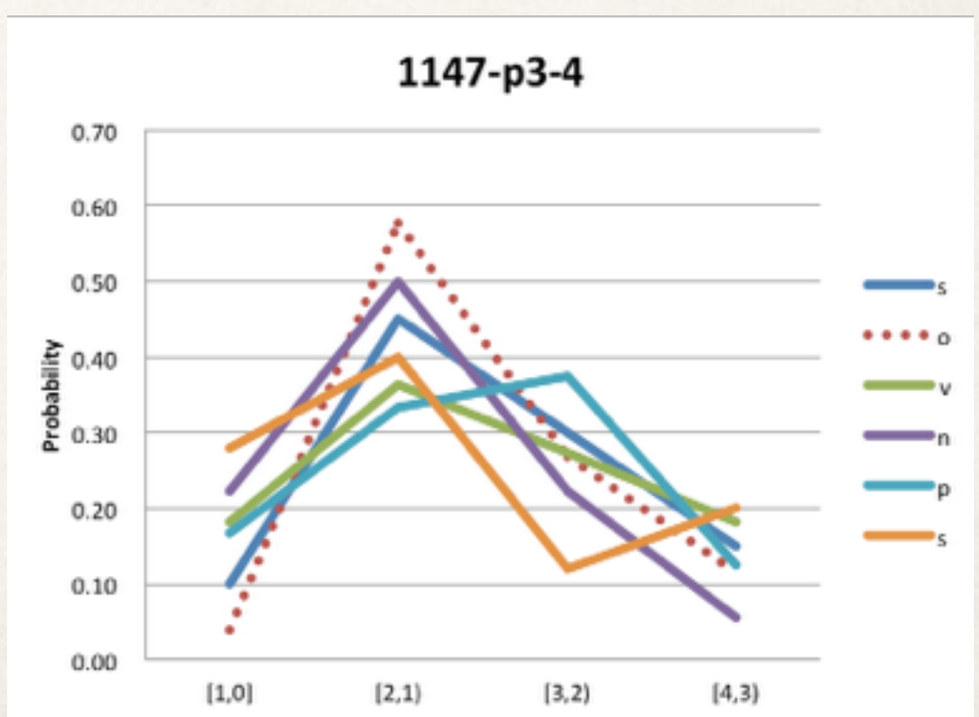


Individual responses ID=1147 (V= *shiri-au* (know each other))

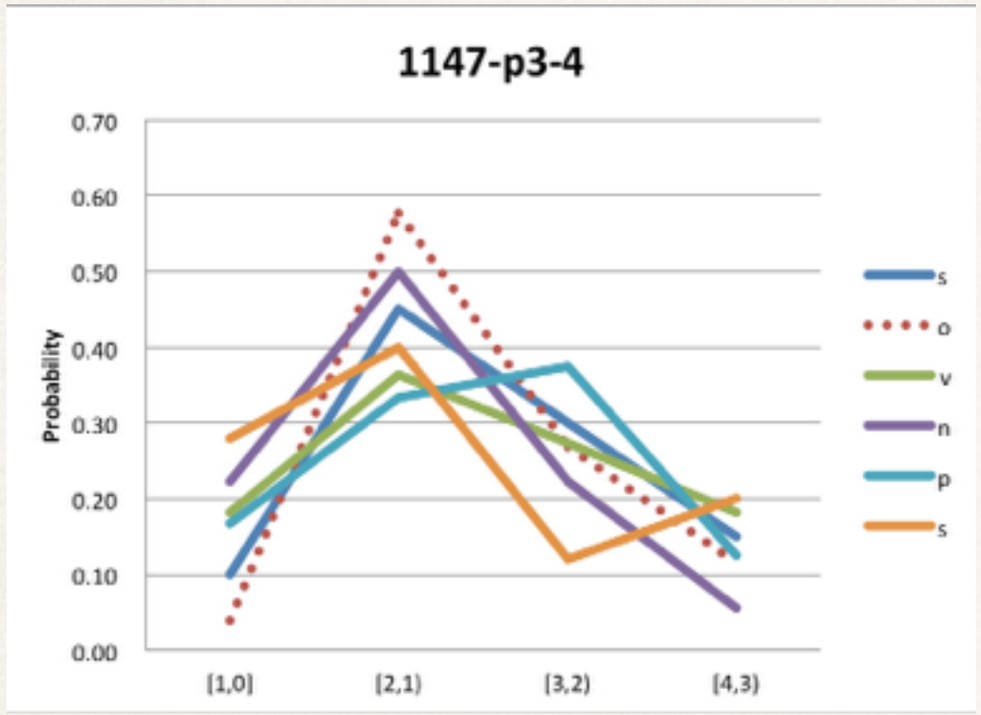


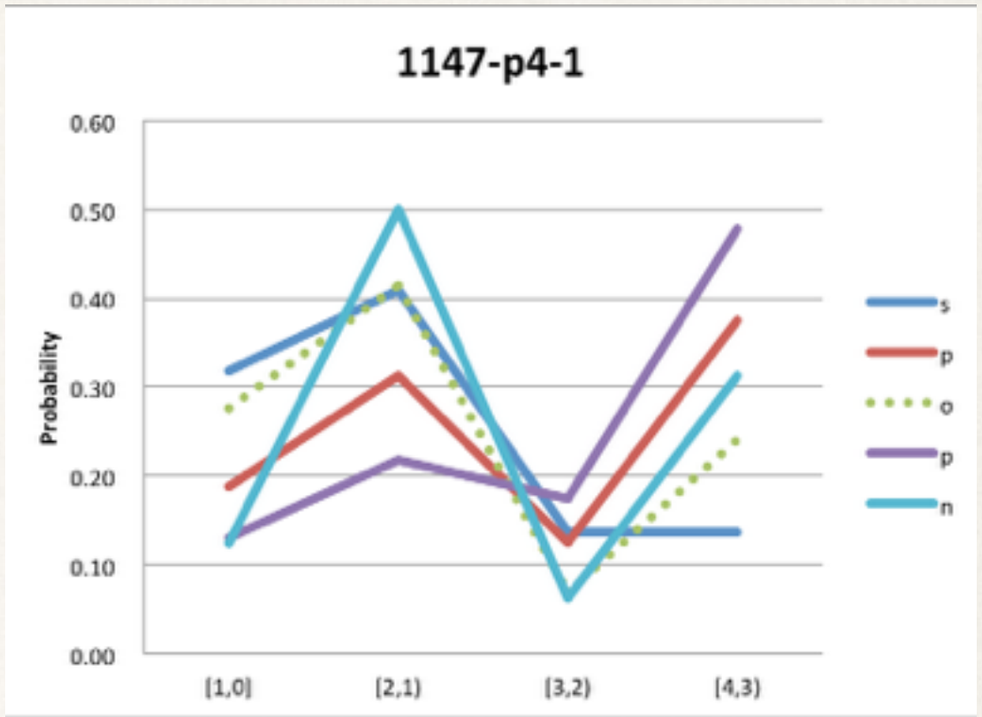




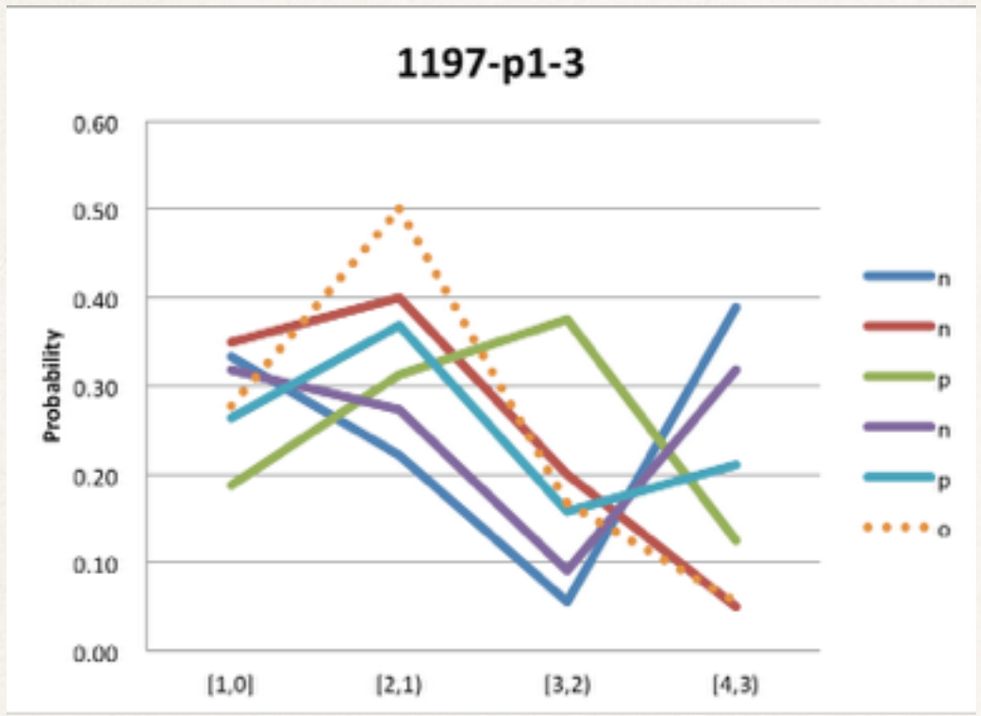


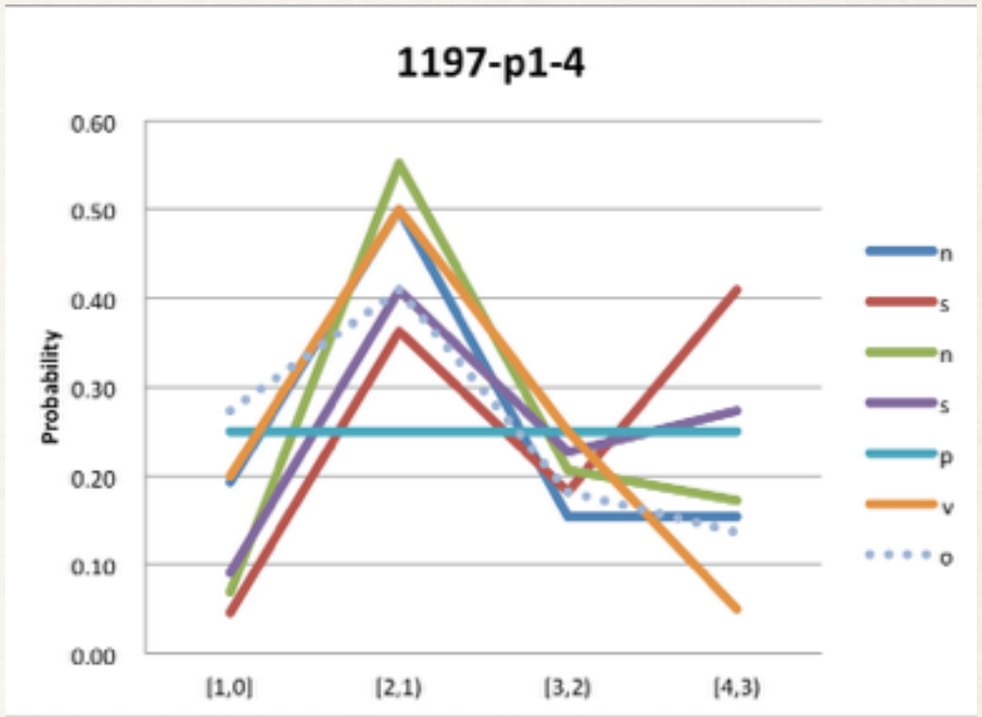
Individual responses ID=1147

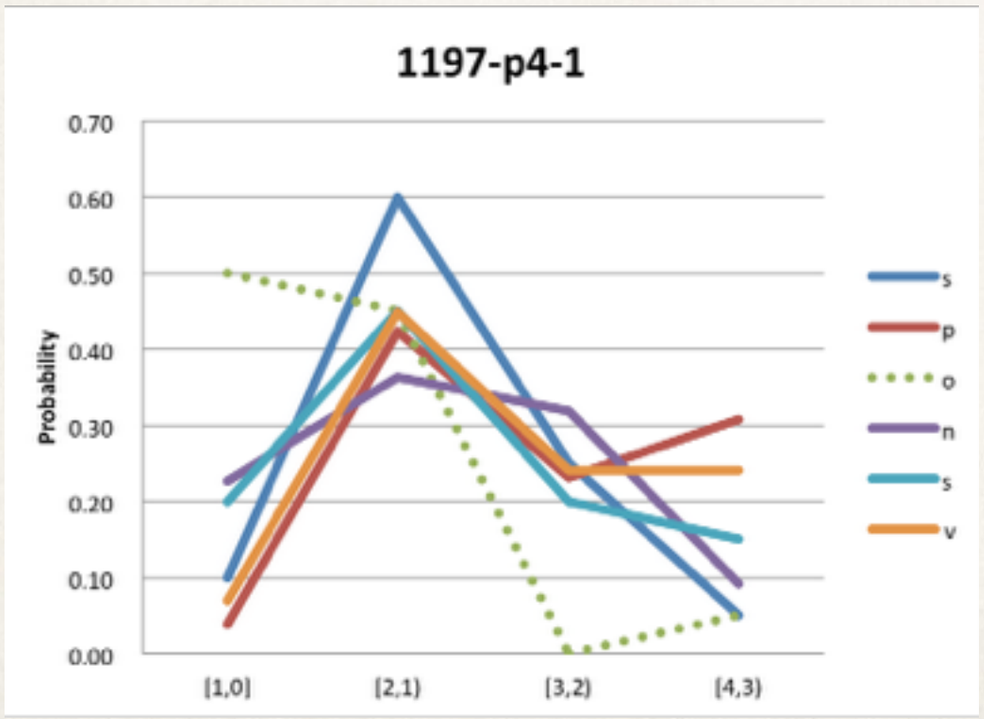




Individual responses ID=1197 (V= *kansen-suru* (contract))







(Someting like) Conclusions

- I can't give you here what people can normally call "conclusions."
- But I believe I provided pieces of evidence to argue that
  - human responses to systematically generated stimuli on a larger scale are more complicated than linguists thought to be,
  - they deny easy explanations.
    - The two classes of deviant stimuli didn't correspond linguistically conceivable classes of deviance.
  - they call for effective method of analysis and interpretation.

Future work

- For this pilot study, I will make
  - investigation of how responses vary for particular verbs.
  - investigation of how responses vary for particular patterns/ constructions.
  - investigation of what classes raters belong to, trying to find social attributes that affect it.
- And of course I will run the main survey with expanded scope of variabilities

Additional Note

- Our submission to LREC 2018 on these results was rejected.
- I was surprised to know that
  - I was expected to explain what *acceptability* is.
  - I was expected to explain why even professional linguists are not well suited to produce deviant sentences systematically.
    - If these are really the cases, give me more space to talk about it.

Thank you