A "Rank vs. Cluster" Conflict: Or, Is It Just an Artifact?

N. Sugino, K. Yamakawa, M. Nakano, H. Ohba, Y. Shimizu and K. Shojima

This study reports on conflicting outcomes of two analytical procedures, viz., NTT and SOM, on a same set of data from a grammaticality judgement task on the unaccusative/unergative distinction by Japanese EFL learners. Based on the findings, discussed in the presentation are:

- The distinctions among the learner "ranks," and the pre-determined "classes" and "categories," were not clearly observed in the SOM outcome.
- Although the target features can be classified into "classes" and "categories" and the learners into "ranks,"
  - those categories are derived from the theory, and
  - those distinctions among learners are dependent on the robustness of the analytical procedure employed.
- This implies that researchers are seeing what they would rather see.
  - Thus, claims entailing such subtle differences need to be made with due reservation on the subtlety.

PAAL 2010 Annual Conference: August 18, 2010, Hanyang Women's University, Seoul, Korea

The present study is supported by Grant-in-Aid for Scientific Research (B) (22320114) from Japan Society for the Promotion of Science.

Unergatives and Unaccusatives

- The Unaccusative Hypothesis (Perlmutter, 1978, Burzio, 1986)
  - Two types of intransitive verbs
    - Unergative Verbs (e.g., *dance, cry*)
      - Mary <u>danced.</u>
      - [Mary [$_{VP}$ dance]]
    - Unaccusative Verbs (e.g., *happen, break*)
      - The accident happened 15 years ago.
      - [empty [$_{VP}$ happen [the accident]]]

- Overpassivization errors of unaccusatives in SLA
  - * Most people <u>are fallen</u> in love ... (Zobl, 1989, p. 204)
- Unaccusative Trap Hypothesis (Oshita, 2001)
  - U-shaped development in the acquisition of unaccusatives

Two Accounts on L2 Errors with Passivized Unaccusatives

- NP-movement
  - [empty [VP be written [the song]]]

    ↓ wrongly applied
  - *[empty [VP be happened [the accident]]]
- Lexical causativization
  - *Something happened the accident.

    ↓
  - *The accident was happened.

Data

- Participants: (369) University-level Japanese EFL learners
- Two classes of English verbs
  - Unaccusatives: *appear, arrive, die, exist, fall, happen*
  - Unergatives: *cry, dance, laugh, play, sing, work*
- Grammaticality judgment in 3 sentence patterns; 5-point scale

(1) NP+V

- A funny thing *happened* in the office today. [Category A]
- We all *laughed* when we saw his face. [Category B]

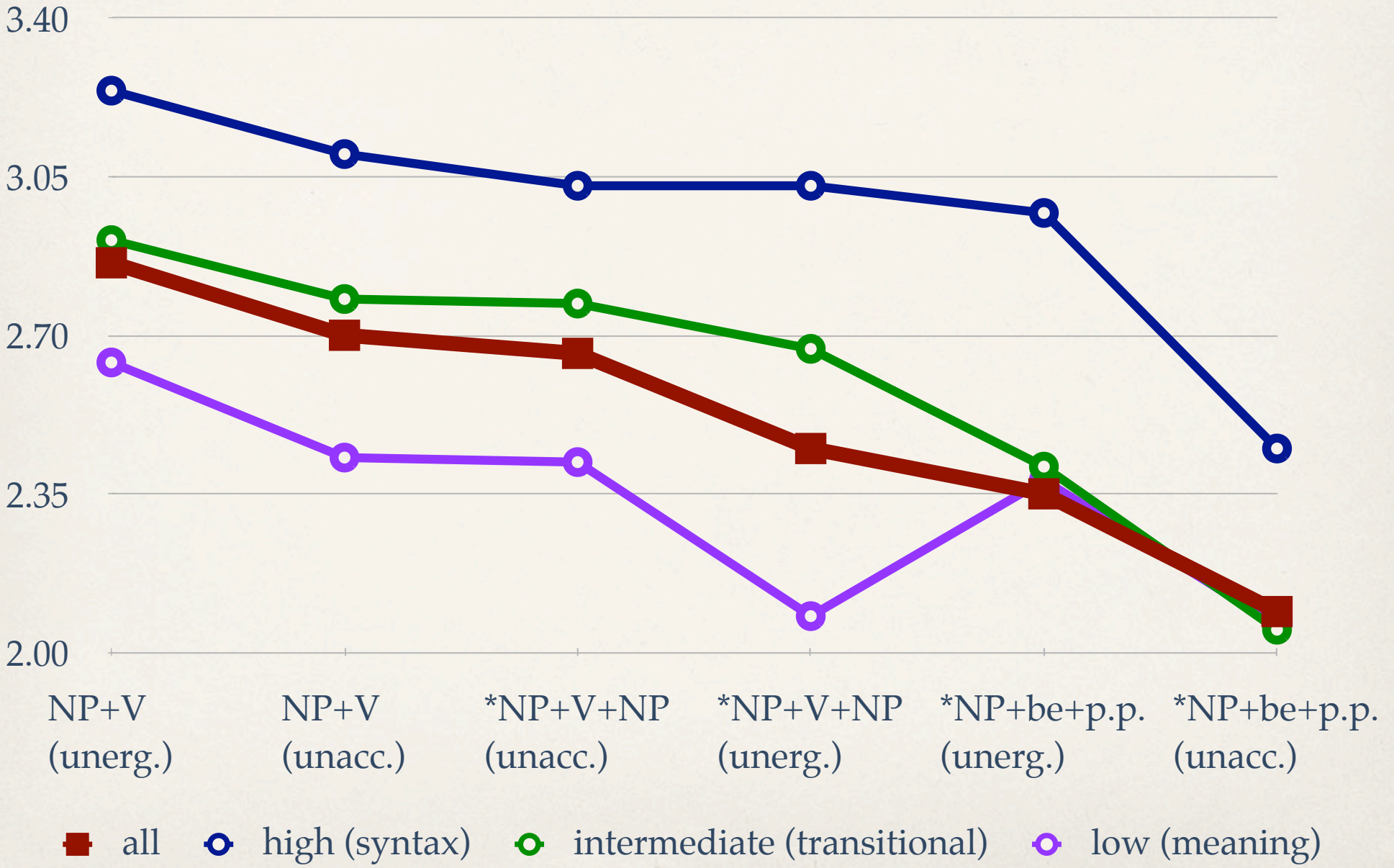
(2) *NP+be+Vp.p.

- *I was not there when the accident *was happened*. [Category C]
- *They *were laughed* when she told a funny joke. [Category D]

(3) *NP+V+NP

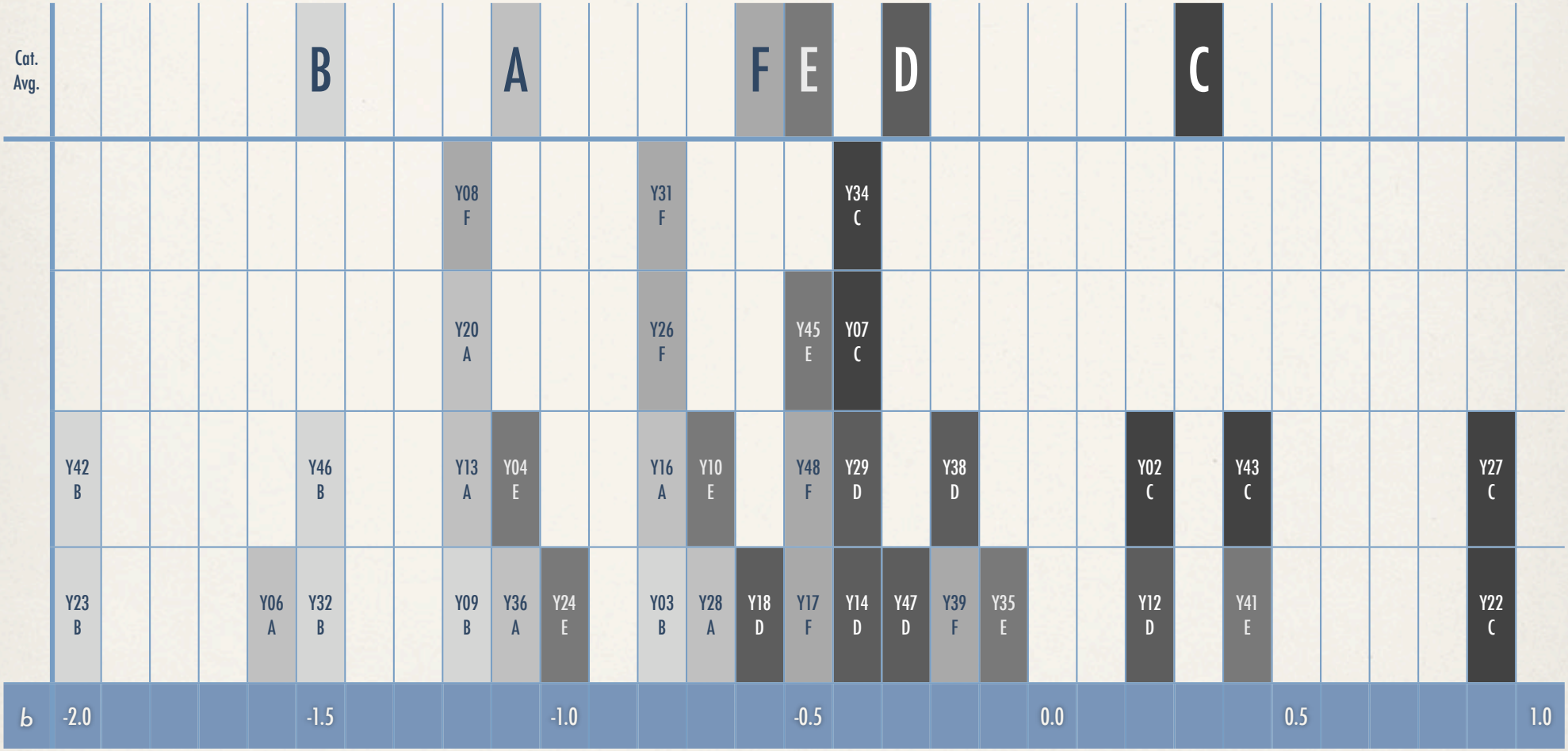
- *Jimmy *happens* a lot of trouble to me. [Category E]
- Bob is so funny. *He always *laughs* me. [Category F]

Conventional Analysis (Yamakawa et al., 2003)



IRT-based Analysis (Yamakawa et al., 2008)

b-parameter values of each item



← Easy

Difficult→

From our previous studies...

- The learners have different degrees of difficulties
  - between the classes (viz., the unaccusatives and the unergatives), and
  - among the categories (viz., Categories A - F) we prepared.
- The learners with different cue dependencies display different response patterns.
  - This suggests the connection between the cue dependency, a performance factor, and the grammaticality judgment.

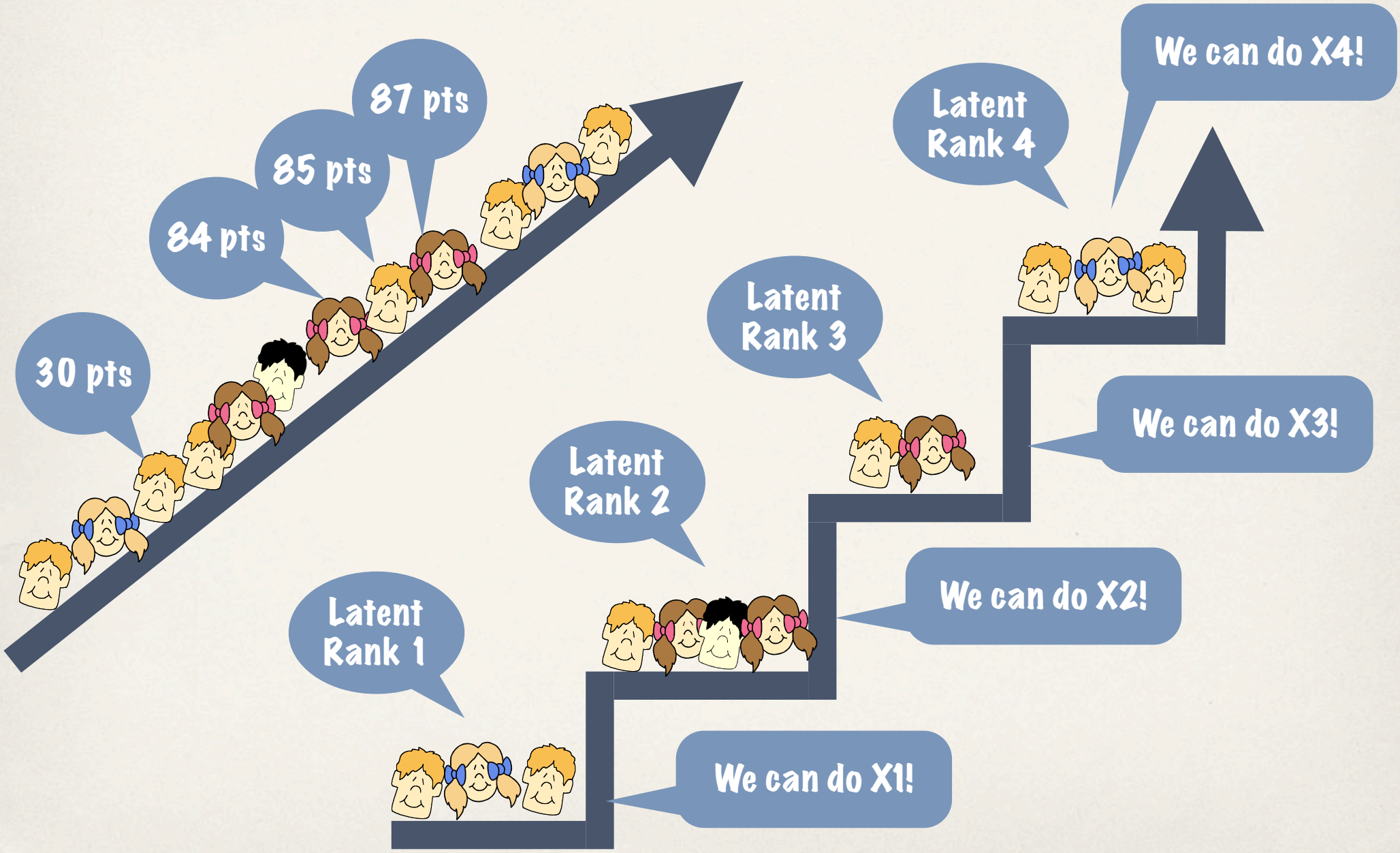
- In order to establish the connection, is there any better way to classify the learners?

➡ Neural Test Theory (NTT; a.k.a. Latent Rank Theory) (Shojima, 2009)

Neural Test Theory

- A test theory designed for evaluating achievements on an ordinal scale.
  - Tests do not have enough resolution to continuously evaluate human ability. Tests are at best capable of ranking test takers into 5 to 20 ranks.
- Each level is called a “latent rank”.
  - Latent ranks are based on the response patterns of the test takers.
- Advantages
  - Easy to explain the relationship between scores and abilities
  - A researcher can decide on the number of latent ranks.
    - Goodness-of-fit indices are used in deciding the number of latent ranks.

Continuous/Ordinal Scales



Results: NTT

- Learners classified into 4 latent ranks.

Item	Category	Verb	Item Response Profile			
			Rank 1	Rank 2	Rank 3	Rank4
			n = 88	*n* = 99	*n* = 95	*n* = 87
Y46	B	sing	2.991	3.365	3.672	3.822
Y42	B	cry	2.890	3.232	3.672	3.592
Y32	B	dance	2.830	3.203	3.672	3.517
Y06	A	die	2.949	3.086	3.672	3.460
Y23	B	laugh	2.817	3.123	3.672	3.832
Y04	E	arrive	2.813	3.013	3.672	3.282
Y36	A	happen	2.674	2.946	3.672	3.278
Y26	F	work	2.549	2.865	3.672	3.236
Y09	B	work	2.621	2.696	2.808	3.010
Y24	E	die	2.571	2.706	2.915	*2.931
Y03	B	play	2.405	2.554	2.888	3.169
Y16	A	appear	*2.539	2.607	2.775	3.048
Y18	D	sing	2.414	2.661	2.819	3.005
Y28	A	exist	2.324	*2.477	2.846	3.142
Y08	F	sing	*2.566	2.637	2.601	2.745
Y31	F	cry	2.416	2.525	2.688	2.907
Y20	A	arrive	2.421	2.613	2.805	2.751
Y13	A	fall	2.409	2.583	2.675	2.799
Y10	E	happen	2.345	2.524	2.755	2.818
Y17	F	dance	2.348	2.423	2.678	2.936
Y07	C	exist	2.358	*2.582	2.694	2.804
Y48	F	play	2.268	2.395	2.643	2.858
Y34	C	arrive	2.243	2.328	2.627	2.866
Y45	E	appear	2.310	*2.521	2.612	2.661
Y35	E	exist	2.126	2.380	2.594	2.841
Y29	D	play	2.125	2.293	2.627	2.874
Y14	D	cry	2.209	2.315	2.561	2.588
Y47	D	work	2.206	2.287	2.442	2.665
Y38	D	dance	2.152	2.214	2.367	2.581
Y39	F	laugh	2.270	2.347	2.316	2.300
Y12	D	laugh	1.937	1.982	2.073	2.171
Y02	C	appear	1.822	1.899	2.041	2.343
Y41	E	fall	1.933	1.938	1.873	2.085
Y43	C	die	1.790	1.780	1.916	2.116
Y27	C	fall	1.928	1.831	1.841	1.847
y22	C	happen	1.736	1.623	1.685	1.848

Results: NTT

	Rank 1	Rank 2	Rank 3	Rank 4
B	**2.759**	**3.029**	**3.265**	**3.416**
A	2.555	**2.767**	**2.923**	**3.080**
F	2.370	2.532	**2.678**	**2.830**
E	2.350	2.512	**2.661**	**2.737**
D	2.174	2.292	2.481	**2.647**
C	1.979	1.892	2.134	2.304

- Among the categories:
  - B>A>F>E>D>C
- Learners at each rank can correctly judge ...
  - Rank 4: up to Category D
  - Rank 3: up to Category E
  - Rank 2: up to Category A
  - Rank 1: up to Category B

Item	Category	Verb	Item Response Profile			
			Rank 1	Rank 2	Rank 3	Rank4
			n = 88	*n* = 99	*n* = 95	*n* = 87
Y46	B	sing	2.991	3.365	3.672	3.822
Y42	B	cry	2.890	3.232	3.672	3.592
Y32	B	dance	2.830	3.203	3.672	3.517
Y23	B	laugh	2.817	3.123	3.672	3.832
Y09	B	work	2.621	2.696	2.808	3.010
Y03	B	play	2.405	2.554	2.888	3.169
Y06	A	die	2.949	3.086	3.672	3.460
Y36	A	happen	2.674	2.946	3.672	3.278
Y28	A	exist	2.324	*2.477	2.846	3.142
Y16	A	appear	*2.539	2.607	2.775	3.048
Y13	A	fall	2.409	2.583	2.675	2.799
Y20	A	arrive	2.421	2.613	2.805	2.751
Y26	F	work	2.549	2.865	3.672	3.236
Y31	F	cry	2.416	2.525	2.688	2.907
Y08	F	sing	*2.566	2.637	2.601	2.745
Y17	F	dance	2.348	2.423	2.678	2.936
Y48	F	play	2.268	2.395	2.643	2.858
Y39	F	laugh	2.270	2.347	2.316	2.300
Y04	E	arrive	2.813	3.013	3.672	3.282
Y24	E	die	2.571	2.706	2.915	*2.931
Y10	E	happen	2.345	2.524	2.755	2.818
Y45	E	appear	2.310	*2.521	2.612	2.661
Y35	E	exist	2.126	2.380	2.594	2.841
Y41	E	fall	1.933	1.938	1.873	2.085
Y18	D	sing	2.414	2.661	2.819	3.005
Y29	D	play	2.125	2.293	2.627	2.874
Y14	D	cry	2.209	2.315	2.561	2.588
Y47	D	work	2.206	2.287	2.442	2.665
Y38	D	dance	2.152	2.214	2.367	2.581
Y12	D	laugh	1.937	1.982	2.073	2.171
Y07	C	exist	2.358	*2.582	2.694	2.804
Y34	C	arrive	2.243	2.328	2.627	2.866
Y02	C	appear	1.822	1.899	2.041	2.343
Y43	C	die	1.790	1.780	1.916	2.116
Y27	C	fall	1.928	1.831	1.841	1.847
y22	C	happen	1.736	1.623	1.685	1.848

Some "naïve" questions

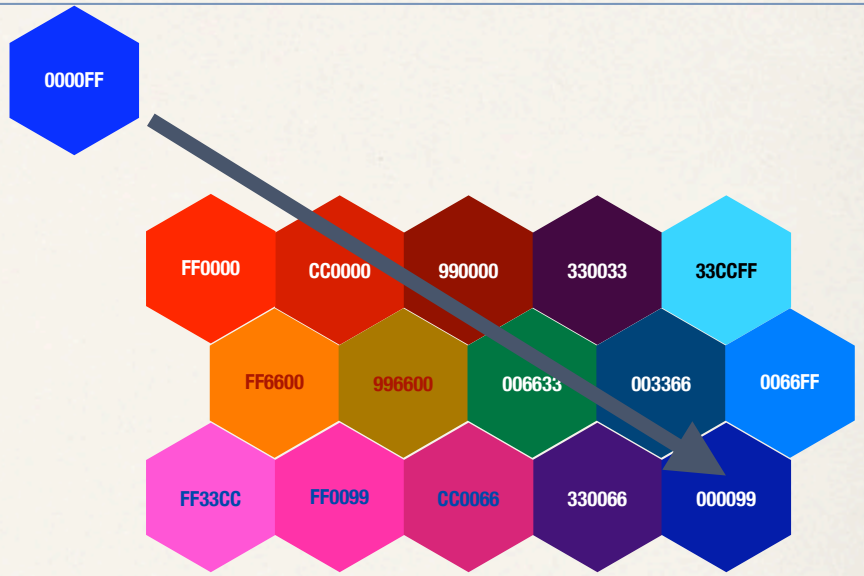
- Conventional research procedure involves comparisons of mean scores
  - in theoretically-determined "categories" and
  - of "learner groups" divided, in many cases, according to their (overall) language proficiency.
- This procedure seems to presuppose that ...
  - The member-items in a particular category are acquired and represented in the same way, and
  - The learners in a group, i.e., those at the same proficiency level, share similar representations of the target linguistic features.
- But is this really the case?

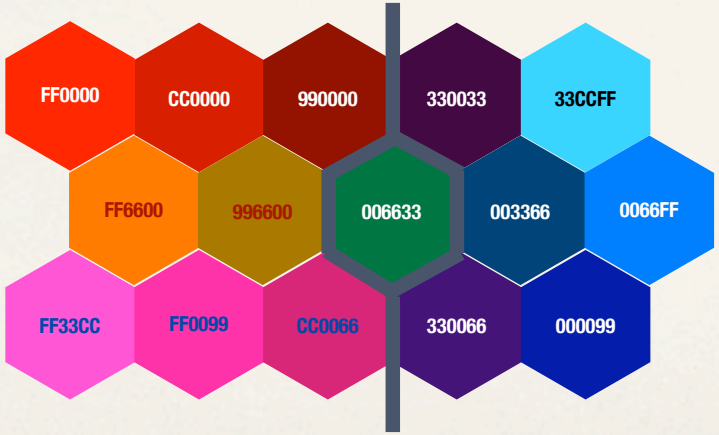
➡ Self-organizing map (SOM) (Kohonenn, 1997)

Self Organizing Map

- A kind of cluster analysis
  - Each sample is processed as an input bearing a unique set of weight vectors, and is positioned onto a two-dimensional map as a “node.”
- Advantages
  - No need to manipulate the raw data.
    - Loss of information can be avoided.
  - The data are categorized into clusters based on their similarities.
    - The clusters obtained are not pre-determined by a theory or a model.
  - A researcher has more freedom in interpreting the results.

Self Organizing Map





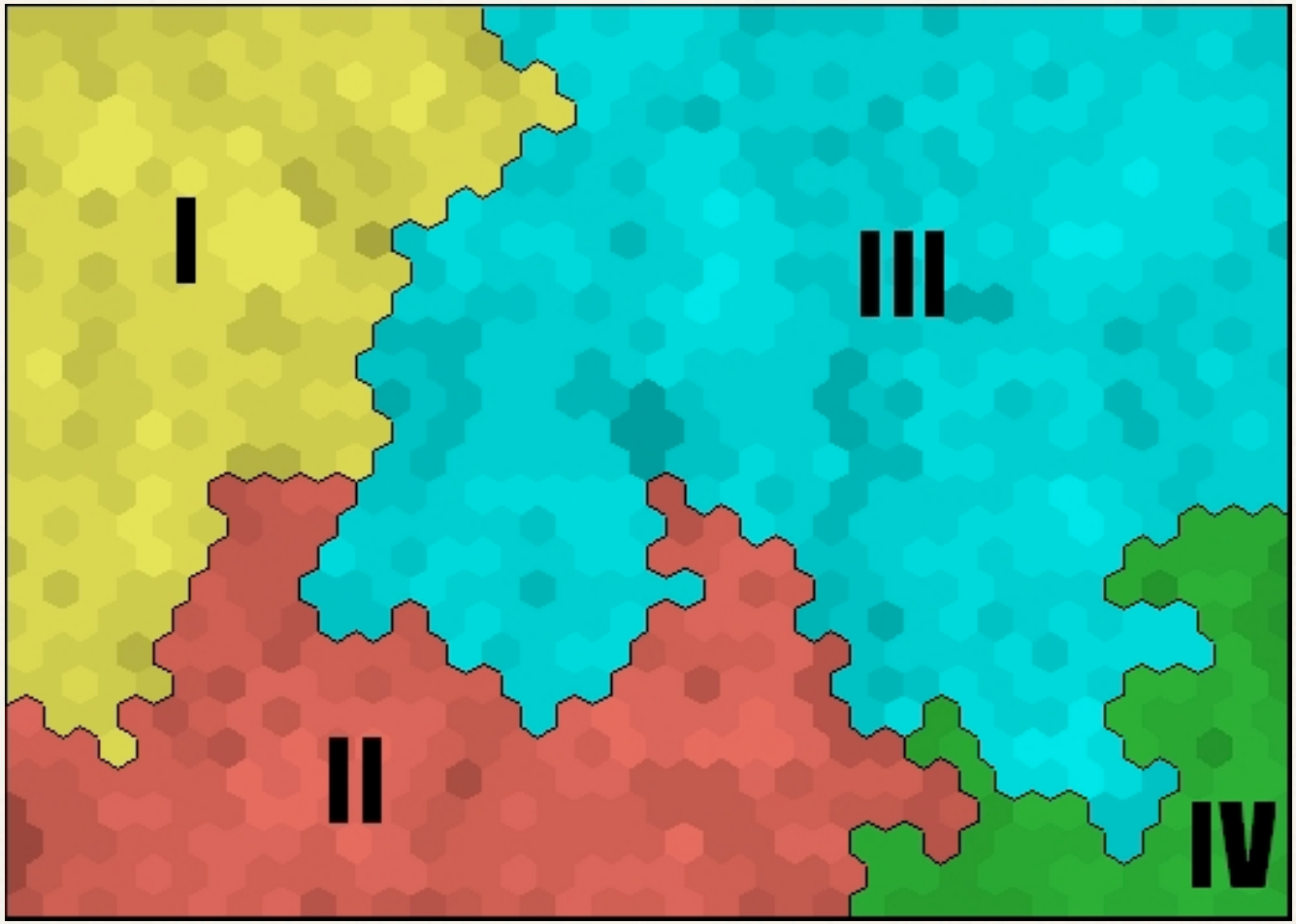
(1) Referential nodes, each with a set of "weight vectors"(e.g., "FFFFFF"), are randomly aligned, forming, in this case, a 3 x 5 rectangle. Awaits an input.
(2) A new input with a particular set of weight vectors is given.
(3) The new input is placed onto the referential node that has the identical or similar set of weight vectors. This node is called a *winner*.
(4) The neighboring referential nodes are to be affected; the weight vectors of the neighboring referential nodes are adjusted so that they will more closely resemble the winner node.
(5) The process is repeated until all the data are placed onto the map.
(6) Clusters are formed based on the similarities between data and displayed topographically.

Clustering the Learners

	unaccusative					unergative			
	appear			*arrive*	…	*sing*	*work*		
	NP+V	NP+be +Vp.p.	NP+V+NP	NP+V	…	NP+V+NP	NP+V	NP+be +Vp.p.	NP+V+NP
S1	5	1	2	4	…	1	4	3	2
S2	5	5	1	5	…	1	5	1	1
…	…	…	…	…	…	…	…	…	…

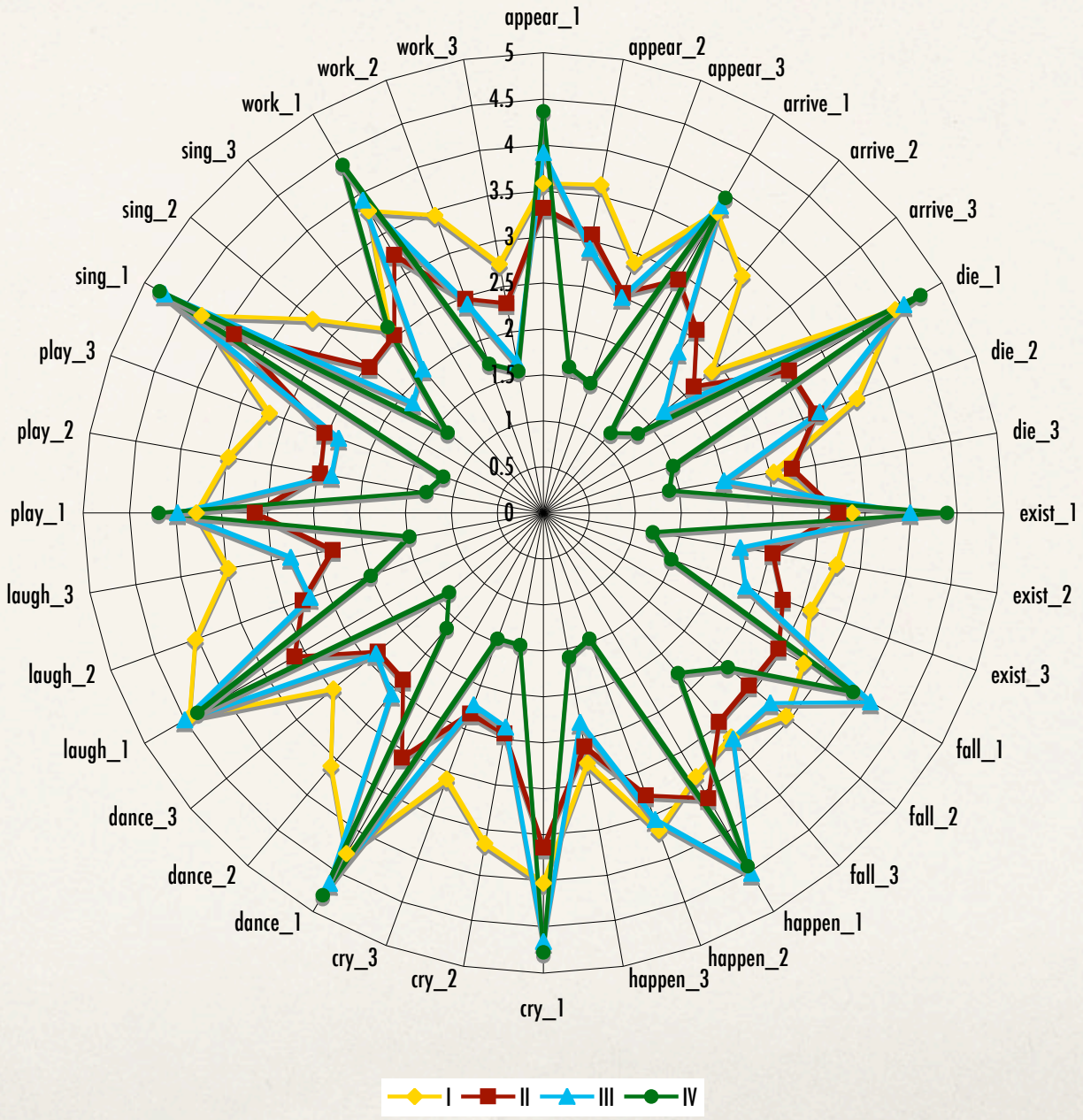
- Prediction
  - There should be some differences in response patterns between the unaccusatives and the unergatives.
  - There should be a cluster of learners who responded wrongly to the "*NP+be+Vp.p." sentences, esp. with the unaccusatives.

Learner Clusters

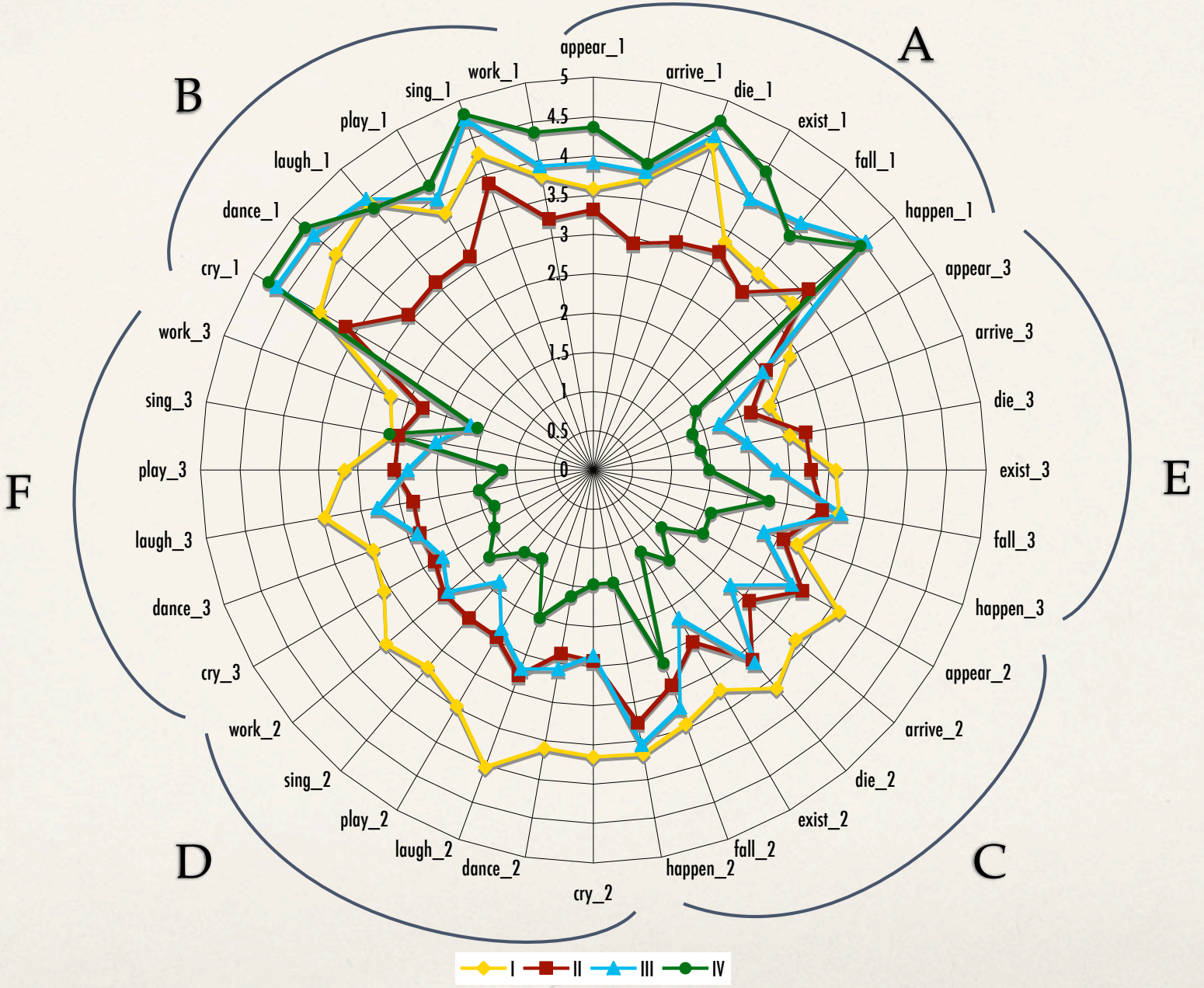


Four clusters of the learners

Mean Scores of the Four Clusters



Mean Scores of the Four Clusters



Characteristics of the Clusters

- Cluster I (*n*=151; 40.9%): Weakly "trapped"
  - Tendency to accept "*NP+be+Vp.p." sentences
  - Regardless of the target verb subclasses
- Cluster II (*n*=83; 22.5%): Uncertain with the grammatical sentences
  - Relatively low acceptance of "NP+V" sentences
  - *cry* and *sing* exceptions
- Cluster III (*n*=91; 24.7%): Transitional
  - "NP+V" > "NP+be+Vp.p." > "NP+V+NP" for most of the unaccusatives
  - "NP+V" > "NP+be+Vp.p." = "NP+V+NP" for most of the unergatives
- Cluster IV (*n*=44; 11.9%): Correctly judging
  - "NP+V" > "NP+be+Vp.p." = "NP+V+NP" for most of the verbs
  - *fall* an exception

In sum

- When the learners are classified according to the similarities in their response patterns ...
  - The distinction between the unergatives and the unaccusatives is not as clear cut as in the previous studies.
  - The difficulty order among the categories is not observed.

➡ So, why not?

Discussion

- Differences in Analytical Procedures
  - ANOVA and IRT
    - Hypothesis-testing procedures
    - Robust against violation of the assumptions
  - NTT
    - Both hypothesis-testing and exploratory
  - SOM
    - Hypothesis-forming and exploratory
    - Strong descriptive power
    - Sensitive to the subtle differences among the responses

Discussion

- Assumptions and hypotheses behind ANOVA, IRT, and NTT analyses
  - The unergative/unaccusative distinction
  - Learners' sensitivity to the distinction
  - That the unaccusatives are atypical would impose more difficulty.
- The findings, though consistent, were obtained when the researchers set about analyzing the data with the above assumptions and testing the hypotheses and 'appropriate' analytical procedures were employed.

Discussion

- When the same set of data are analyzed without such assumptions and general tendencies are exploratory sought for, quite different 'clusters' were obtained.
- This poses question on the significance of the unergative/unaccusative distinction for the Japanese EFL learners.
  - If the distinction is important in SLA, the significance of the distinction needs to be proved not only in its own light but also in relation to the other linguistic features.

References (selected)

Nakano, M., Sugino, N., Yamakawa, K., Ohba, H., & Shimizu, Y. (2007). A study of grammar development among Japanese university students: Intransitive verbs, transitive verbs, ditransitive verbs and logical subjects in Xcomps - Part 1. *Proceedings of the12th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 264-267).

Oshita, H. (2000). *What is happened* may not be what appears to be happening: A corpus study of "passive" unaccusatives in L2 English. *Second Language Research*, *16*, 293-324.

Oshita, H. (2001). The unaccusative trap in second language acquisition. *Studies in Second Language Acquisition*, *23*, 279-304.

Perlmutter, D.M. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the Berkeley Linguistic Society*, *4*, 157-189.

Shojima, K. (2009). Neural test theory. In K. Shigemasu et al. (eds.), *New Trends in Psychometrics* (pp. 407-416), Universal Academy Press.

Yamakawa, K., Sugino, N., Ohba, H., Nakano, M. & Shimizu, Y. (2008). Acquisition of English grammatical features by adult Japanese EFL learners: The application of Item Response Theory in SLA research. *Electronic Journal of Foreign Language Teaching (e-FLT)*, *5* (1), 13-40.

Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S.M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 203-211). Cambridge: Cambridge University Press.