Color naming universals: a statistical approach

Gerhard Jäger
gerhard.jaeger@uni-tuebingen.de

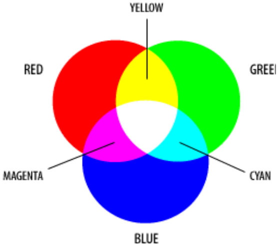
March 29, 2011

QITL 4

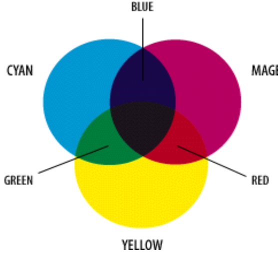
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

The psychological color space

- physical color space has infinite dimensionality — every wavelength within the visible spectrum is one dimension
- psychological color space is only 3-dimensional
- this fact is employed in technical devices like computer screens (additive color space) or color printers (subtractive color space)



additive color space

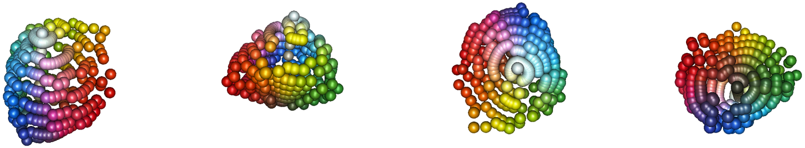


subtractive color space

The psychological color space

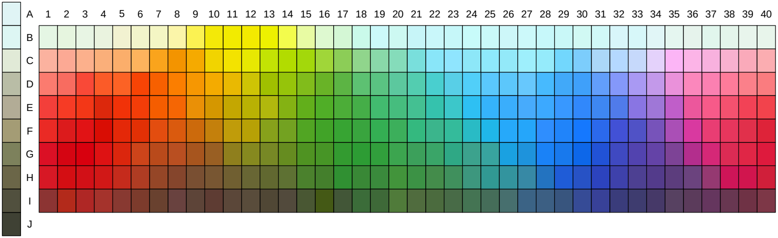
- psychologically correct color space should not only correctly represent the topology of, but also the distances between colors
- distance is inverse function of perceived similarity
- L*a*b* color space has this property
- three axes:
  - black — white
  - red — green
  - blue — yellow
- irregularly shaped 3d **color solid**

The color solid



The Munsell chart

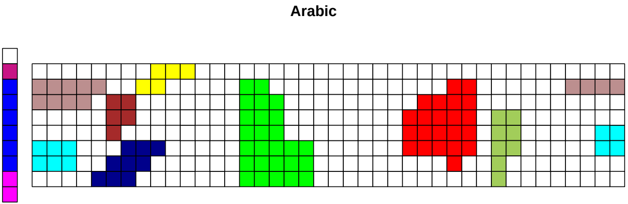
- for psychological investigations, the *Munsell chart* is being used
- 2d-rendering of the surface of the color solid
  - 8 levels of lightness
  - 40 hues
- plus: black–white axis with 8 shaded of grey in between
- neighboring chips differ in the minimally perceivable way

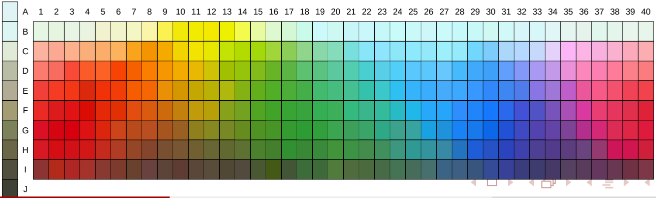


Berlin and Kay 1969

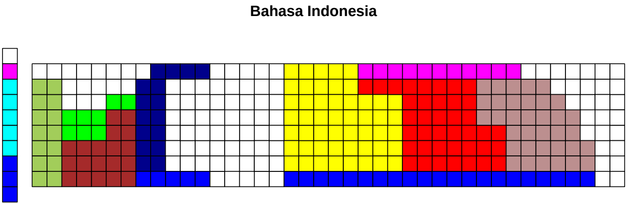
- pilot study how different languages carve up the color space into categories
- informants: speakers of 20 typologically distant languages (who happened to be around the Bay area at the time)
- questions (using the Munsell chart):
  - What are the basic color terms of your native language?
  - What is the extension of these terms?
  - What are the prototypical instances of these terms?
- results are not random
- indicate that there are universal tendencies in color naming systems

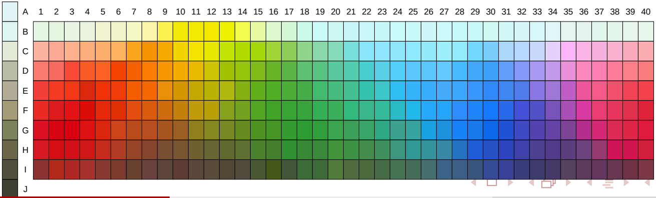
- extensions


Arabic


A
B
C
D
E
F
G
H
I
J
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

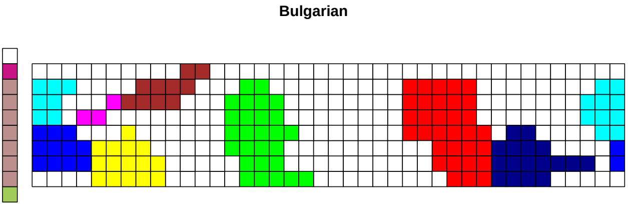
- extensions

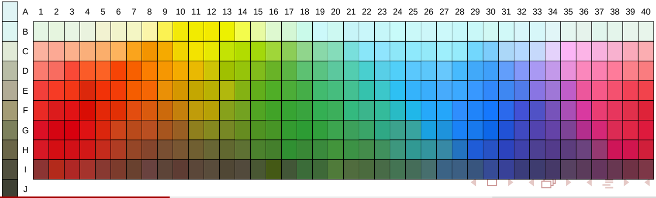

Bahasa Indonesia


A
B
C
D
E
F
G
H
I
J
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

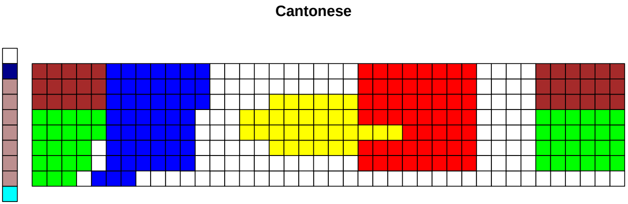
Berlin and Kay 1969

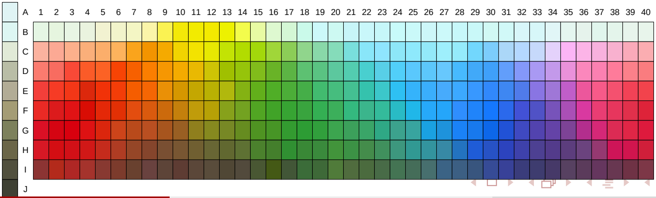
- extensions





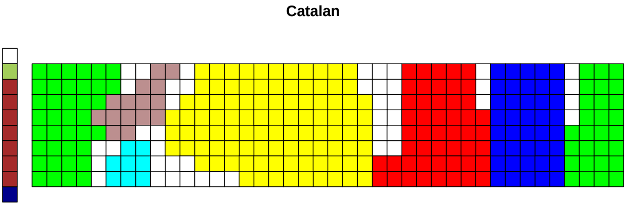
- extensions

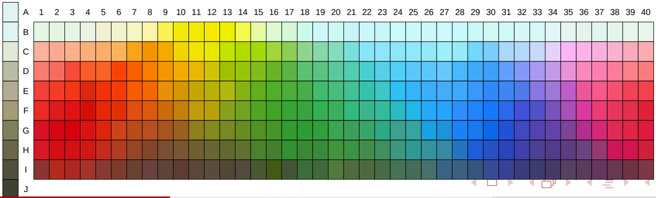

Cantonese


A
B
C
D
E
F
G
H
I
J
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

Berlin and Kay 1969

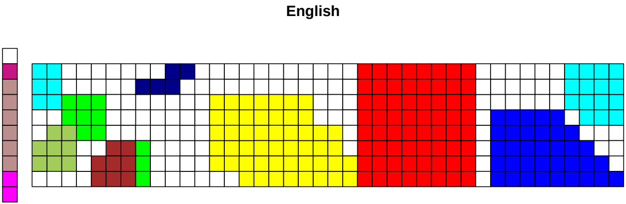
- extensions

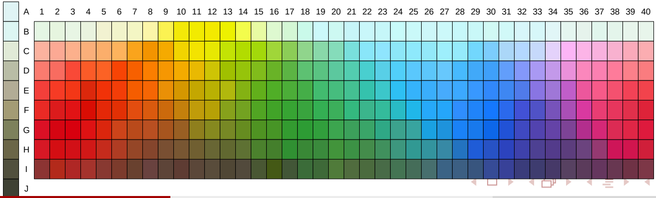




Berlin and Kay 1969

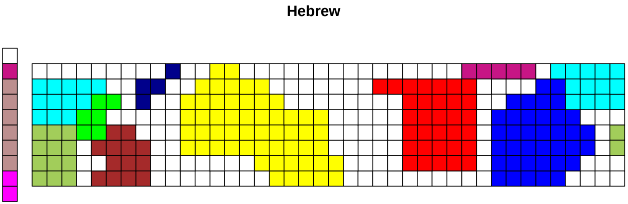
- extensions

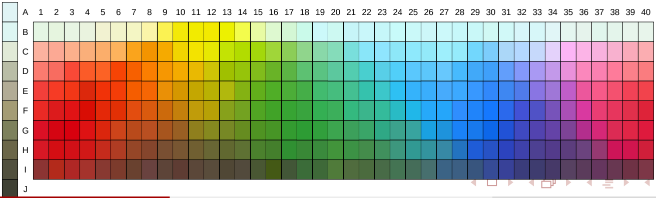




Berlin and Kay 1969

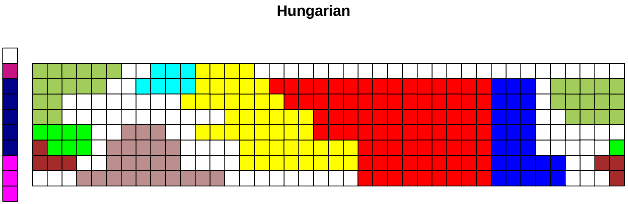
- extensions

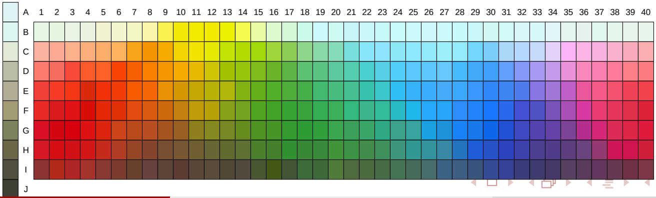




Berlin and Kay 1969

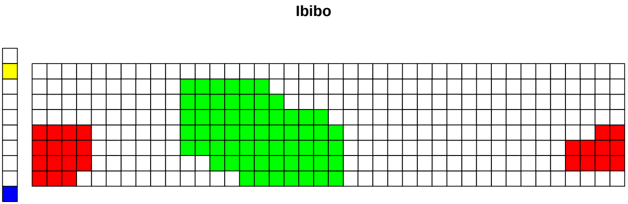
- extensions

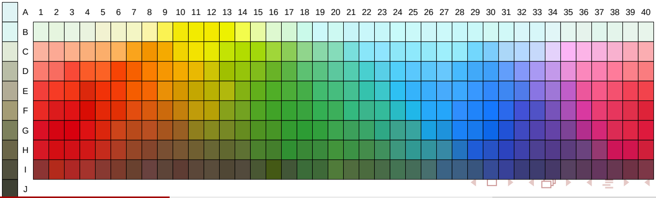




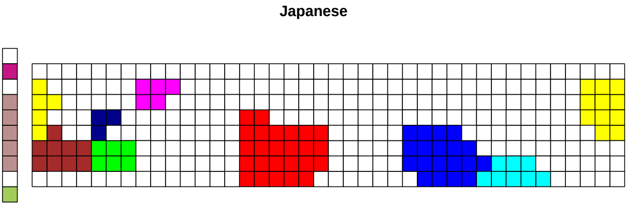
Berlin and Kay 1969

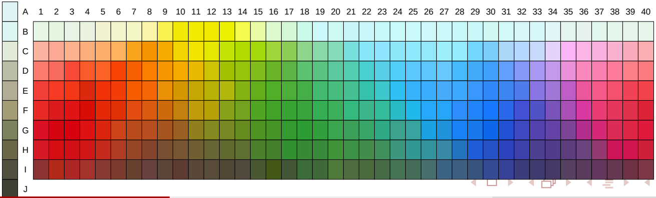
- extensions





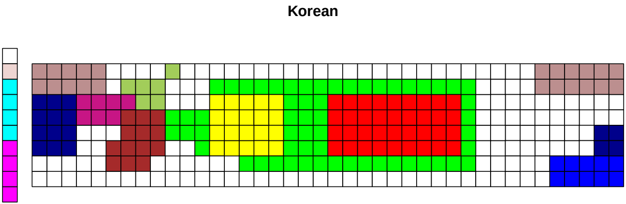
- extensions

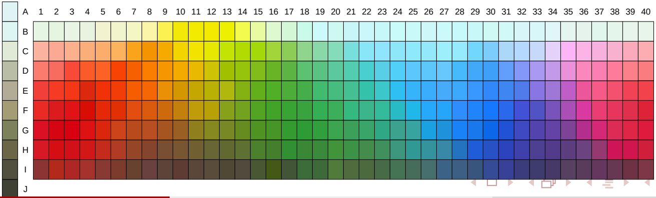

Japanese


A
B
C
D
E
F
G
H
I
J
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

Berlin and Kay 1969

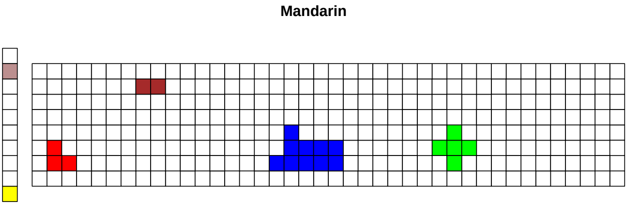
- extensions

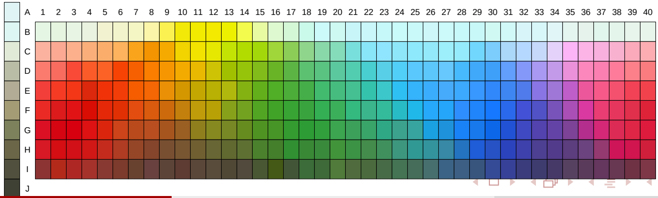




Berlin and Kay 1969

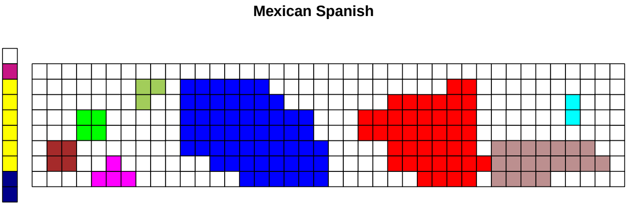
- extensions

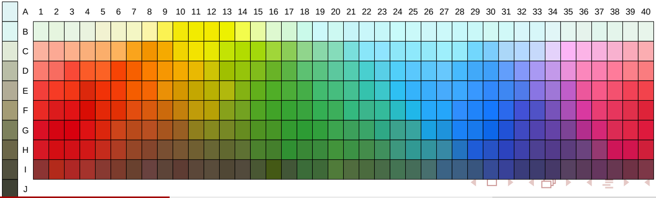




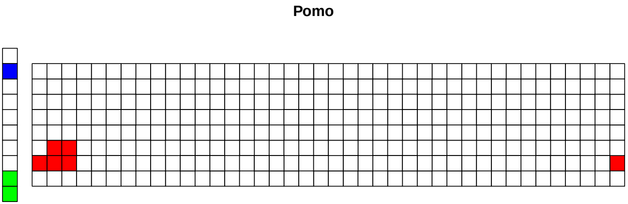
Berlin and Kay 1969

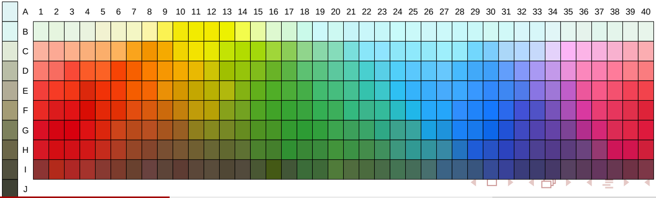
- extensions





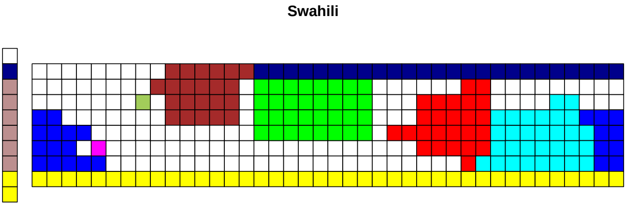
- extensions

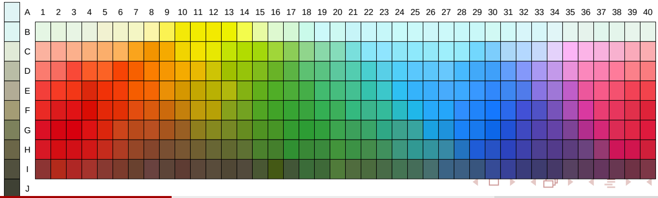

Pomo


A
B
C
D
E
F
G
H
I
J
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

Berlin and Kay 1969

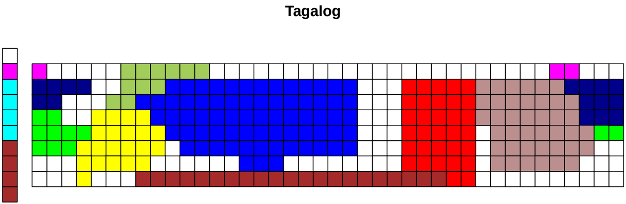
- extensions

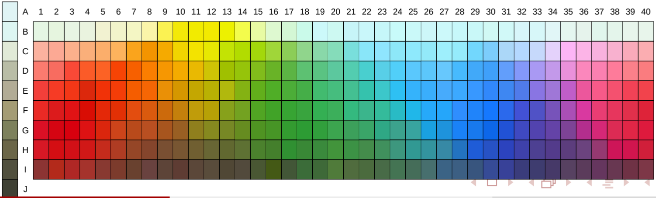




Berlin and Kay 1969

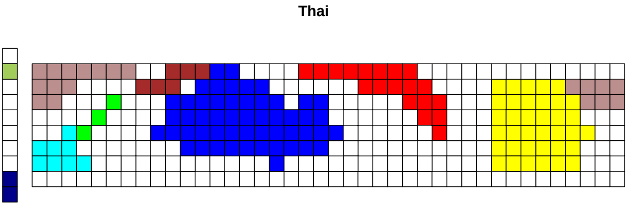
- extensions

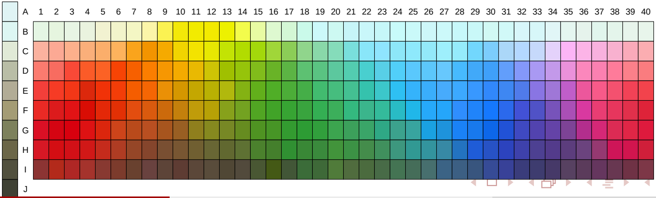




Berlin and Kay 1969

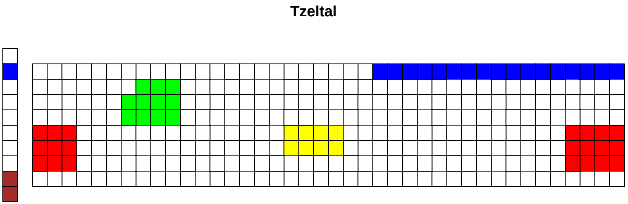
- extensions

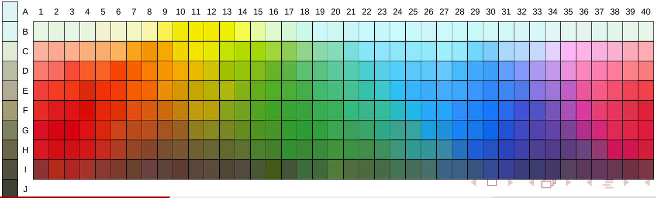




Berlin and Kay 1969

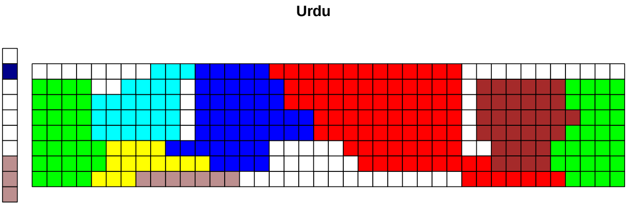
- extensions

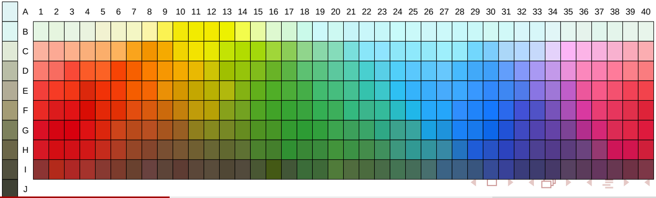




Berlin and Kay 1969

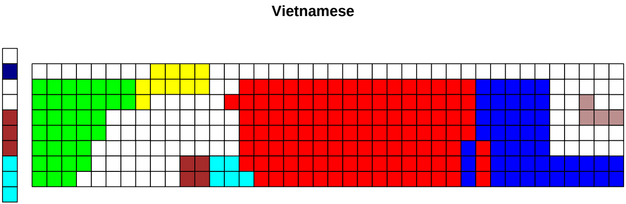
- extensions

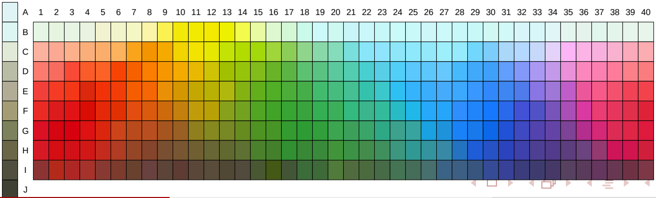




Berlin and Kay 1969

- extensions





Berlin and Kay 1969

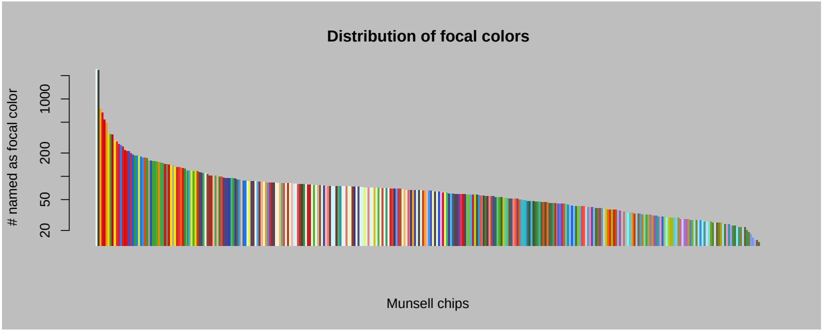
- identification of absolute and implicational universals, like
  - all languages have words for *black* and *white*
  - if a language has a word for *yellow*, it has a word for *red*
  - if a language has a word for *pink*, it has a word for *blue*
  - ...

The World Color Survey

- B&K was criticized for methodological reasons
- in response, in 1976 Kay and co-workers launched the *world color survey*
- investigation of 110 non-written languages from around the world
- around 25 informants per language
- two tasks:
  - the 330 Munsell chips were presented to each test person one after the other in random order; they had to assign each chip to some basic color term from their native language
  - for each native basic color term, each informant identified the prototypical instance(s)
- data are publicly available under http://www.icsi.berkeley.edu/wcs/

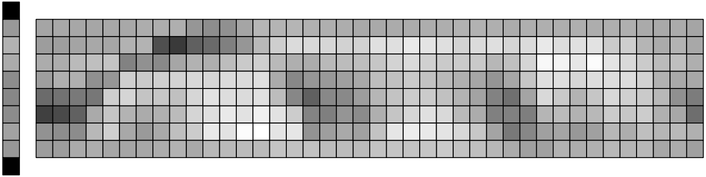
Data digging in the WCS

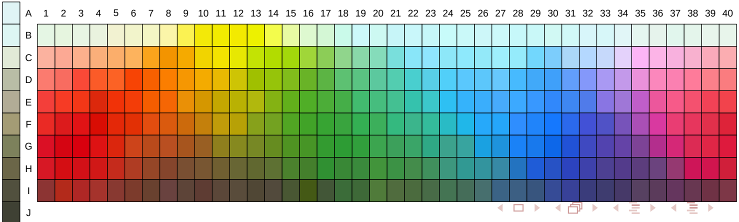
- distribution of focal colors across all informants:



Data digging in the WCS

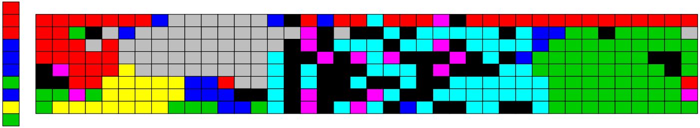
- distribution of focal colors across all informants:

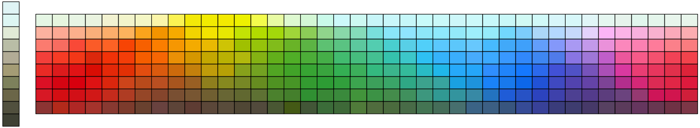




Data digging in the WCS

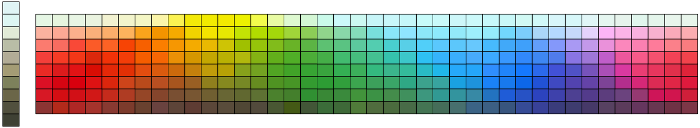
- partition of a randomly chosen informant from a randomly chosen language





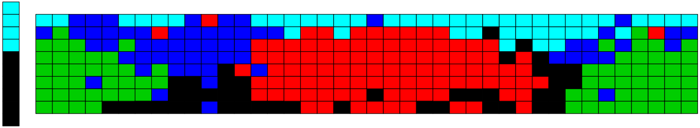
Data digging in the WCS

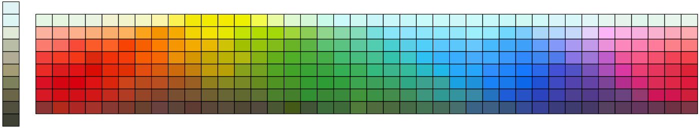
- partition of a randomly chosen informant from a randomly chosen language



Data digging in the WCS

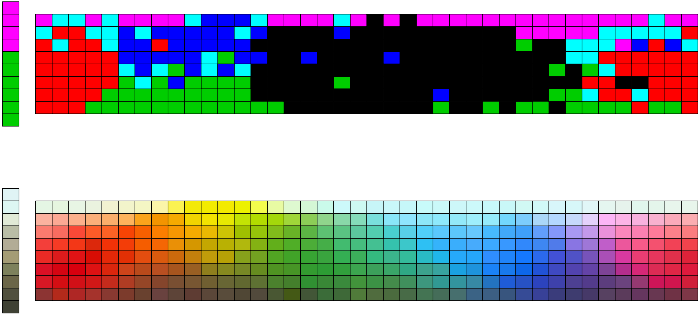
- partition of a randomly chosen informant from a randomly chosen language





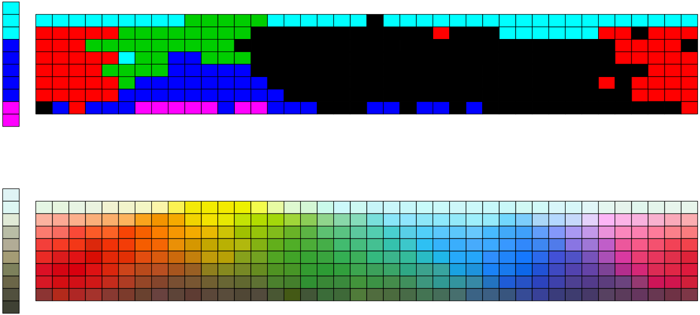
Data digging in the WCS

- partition of a randomly chosen informant from a randomly chosen language

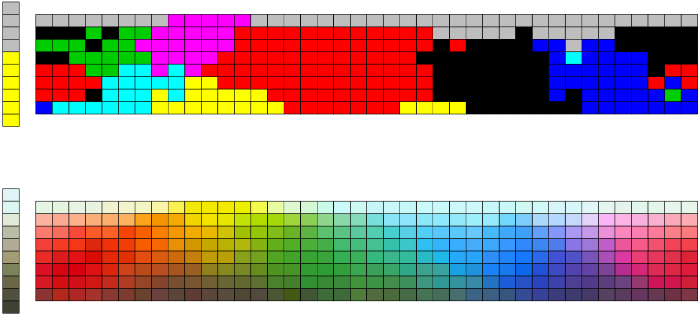


Data digging in the WCS

- partition of a randomly chosen informant from a randomly chosen language

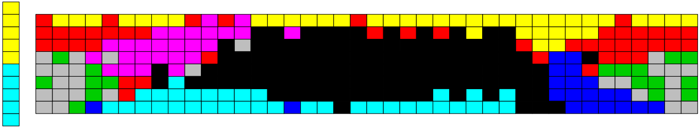


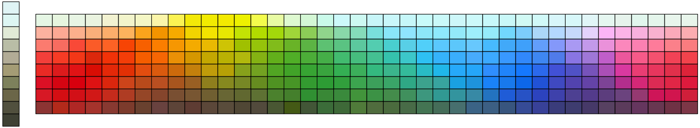
- partition of a randomly chosen informant from a randomly chosen language



Data digging in the WCS

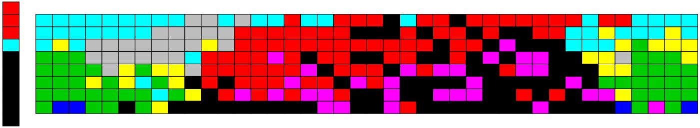
- partition of a randomly chosen informant from a randomly chosen language

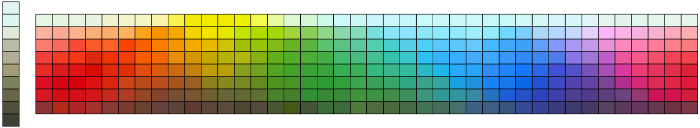




Data digging in the WCS

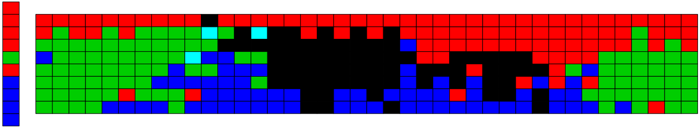
- partition of a randomly chosen informant from a randomly chosen language

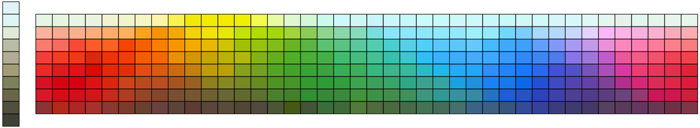




Data digging in the WCS

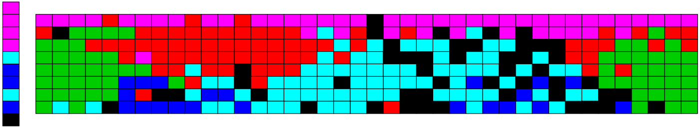
- partition of a randomly chosen informant from a randomly chosen language

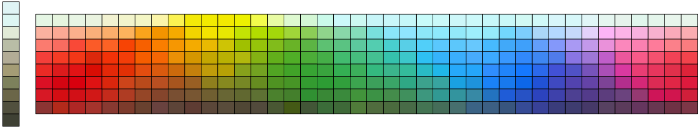




Data digging in the WCS

- partition of a randomly chosen informant from a randomly chosen language





What is the extension of categories?

- data from individual informants are extremely noisy
- averaging over all informants from a language helps, but there is still noise, plus dialectal variation
- desirable: distinction between "genuine" variation and noise

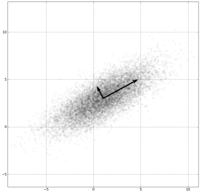
Statistical feature extraction

- first step: representation of raw data in *contingency matrix*
  - rows: color terms from various languages
  - columns: Munsell chips
  - cells: number of test persons who used the row-term for the column-chip

	A0	B0	B1	B2	⋯	I38	I39	I40	J0
red	0	0	0	0	⋯	0	0	2	0
green	0	0	0	0	⋯	0	0	0	0
blue	0	0	0	0	⋯	0	0	0	0
black	0	0	0	0	⋯	18	23	21	25
white	25	25	22	23	⋯	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
rot	0	0	0	0	⋯	1	0	0	0
grün	0	0	0	0	⋯	0	0	0	0
gelb	0	0	0	1	⋯	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
rouge	0	0	0	0	⋯	0	0	0	0
vert	0	0	0	0	⋯	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- further processing:
  - divide each row by the number n of test persons using the corresponding term
  - duplicate each row n times

Principal Component Analysis



- technique to reduce dimensionality of data
- input: set of vectors in an n-dimensional space

first step:

- rotate the coordinate system, such that
  - the new n coordinates are orthogonal to each other
  - the variations of the data along the new coordinates are stochastically independent

second step:

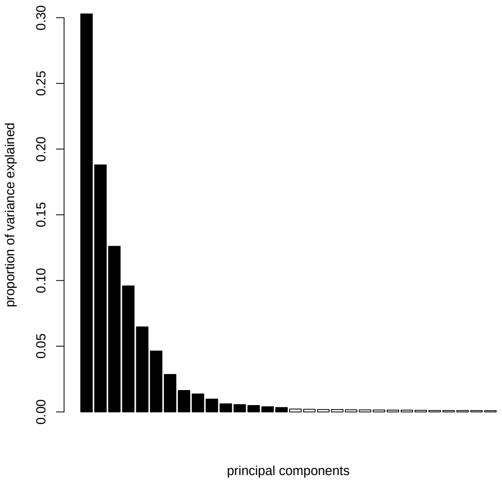
- choose a suitable $m < n$
- project the data on those m new coordinates where the data have the highest variance

Principal Component Analysis

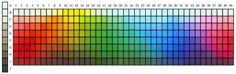
- alternative formulation:
  - choose an m-dimensional linear sub-manifold of your n-dimensional space
  - project your data onto this manifold
  - when doing so, pick your sub-manifold such that the average squared distance of the data points from the sub-manifold is minimized
- intuition behind this formulation:
  - data are "actually" generated in an m-dimensional space
  - observations are disturbed by n-dimensional noise
  - PCA is a way to reconstruct the underlying data distribution
- applications: picture recognition, latent semantic analysis, statistical data analysis in general, data visualization, ...

Statistical feature extraction: PCA

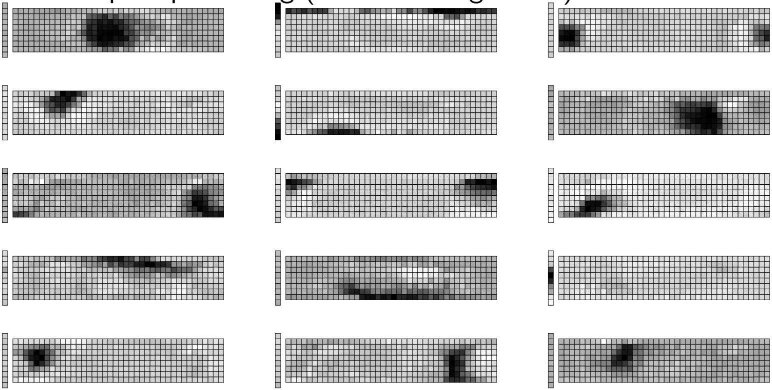
- first 15 principal components jointly explain 91.6% of the total variance
- choice of $m = 15$ is determined by using "Kaiser's stopping rule"



Statistical feature extraction: PCA



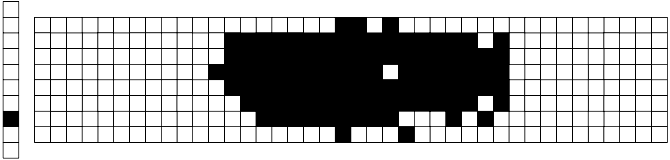
after some post-processing ("varimax" algorithm):

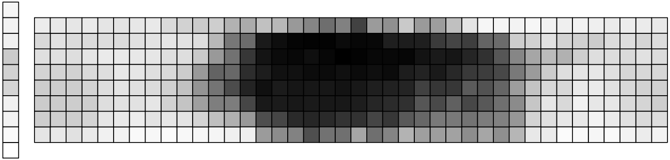


Projecting observed data on lower-dimensional-manifold

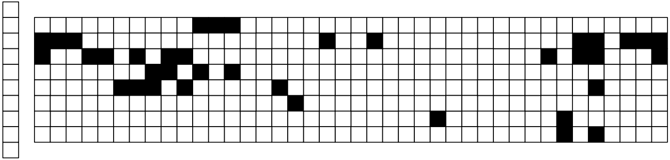
- noise removal: project observed data onto the lower-dimensional submanifold that was obtained via PCA
- in our case: noisy binary categories are mapped to smoothed fuzzy categories (= probability distributions over Munsell chips)
- some examples:

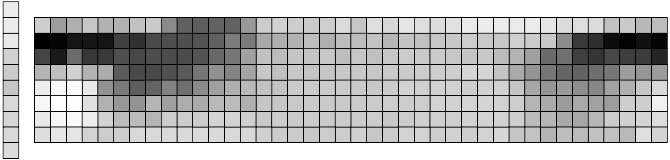
Projecting observed data on lower-dimensional-manifold



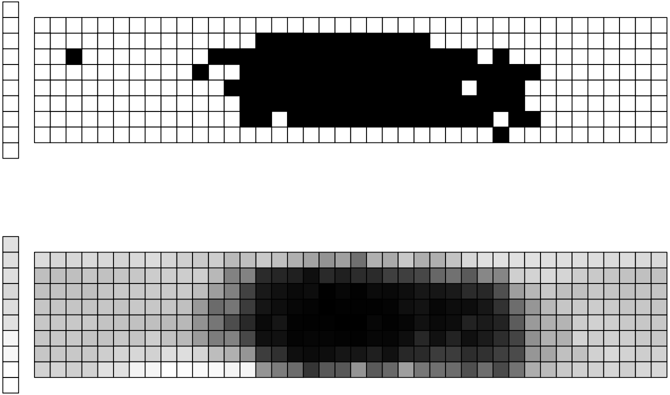


Projecting observed data on lower-dimensional-manifold

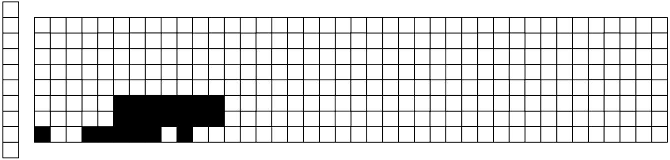


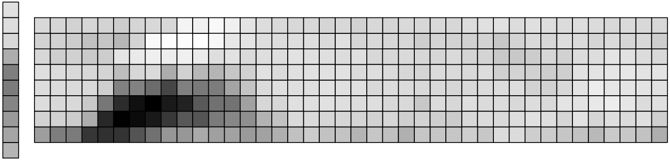


Projecting observed data on lower-dimensional-manifold

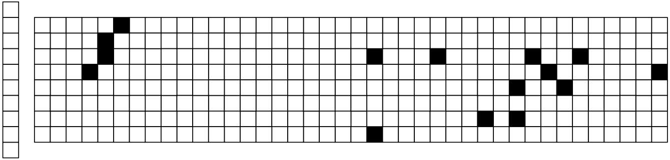


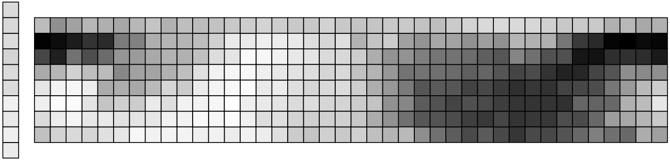
Projecting observed data on lower-dimensional-manifold



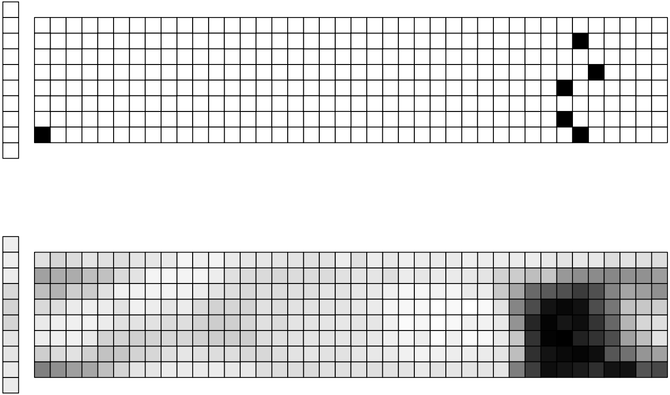


Projecting observed data on lower-dimensional-manifold

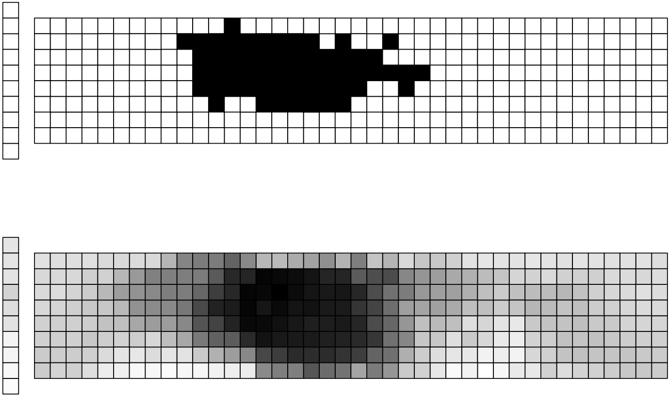




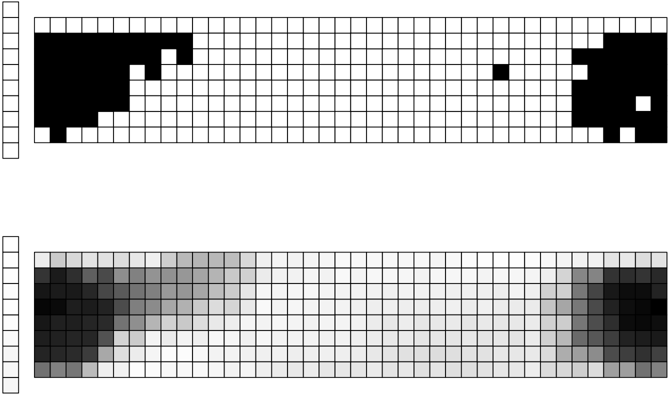
Projecting observed data on lower-dimensional-manifold



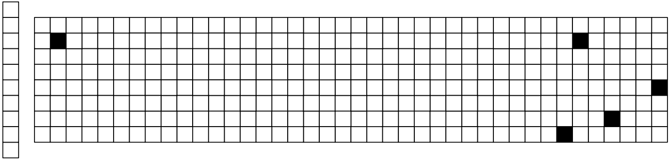
Projecting observed data on lower-dimensional-manifold

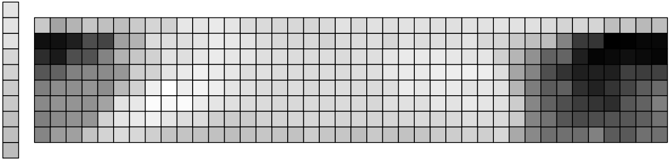


Projecting observed data on lower-dimensional-manifold

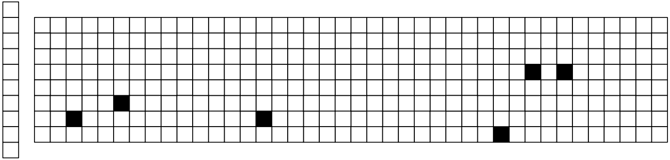


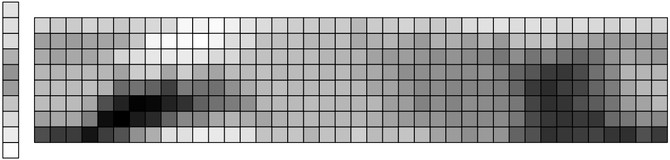
Projecting observed data on lower-dimensional-manifold



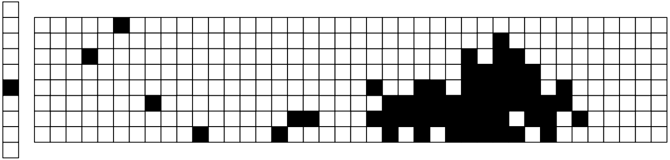


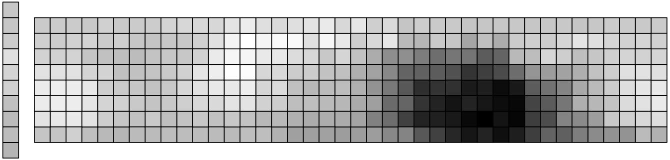
Projecting observed data on lower-dimensional-manifold



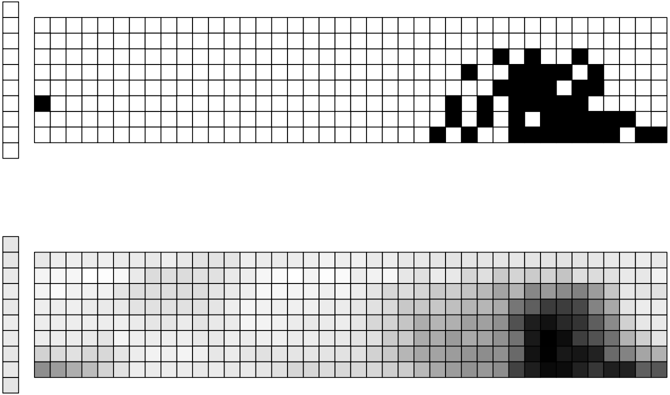


Projecting observed data on lower-dimensional-manifold

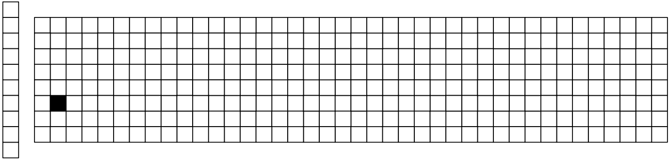


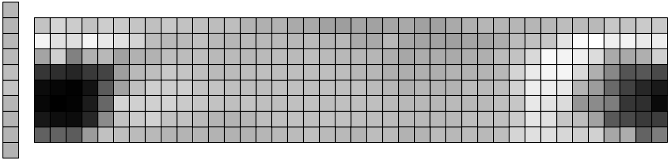


Projecting observed data on lower-dimensional-manifold

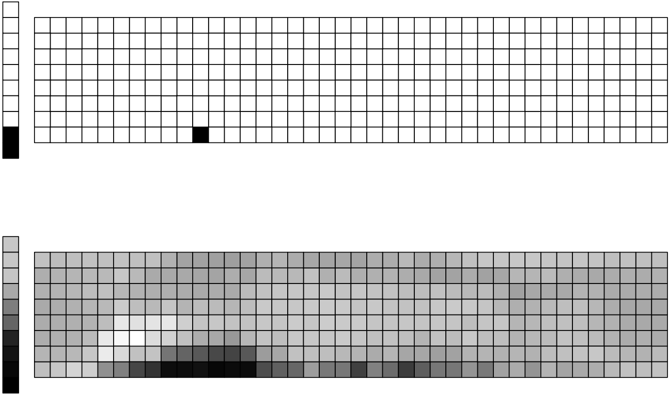


Projecting observed data on lower-dimensional-manifold

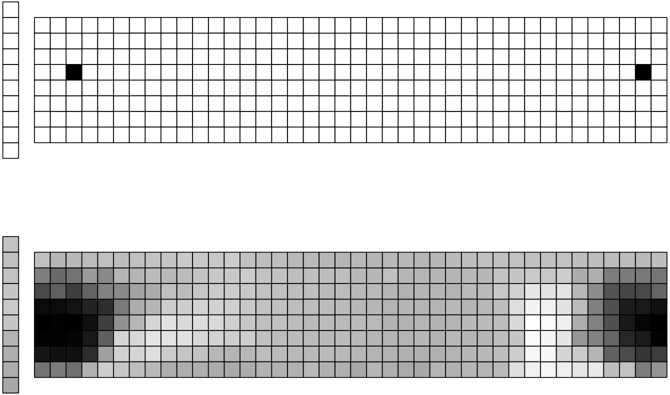




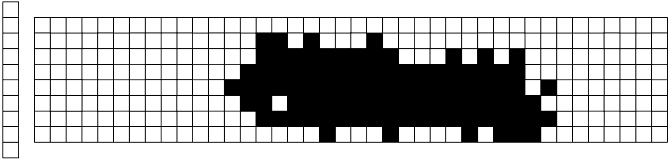
Projecting observed data on lower-dimensional-manifold

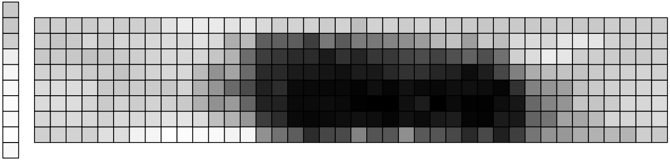


Projecting observed data on lower-dimensional-manifold

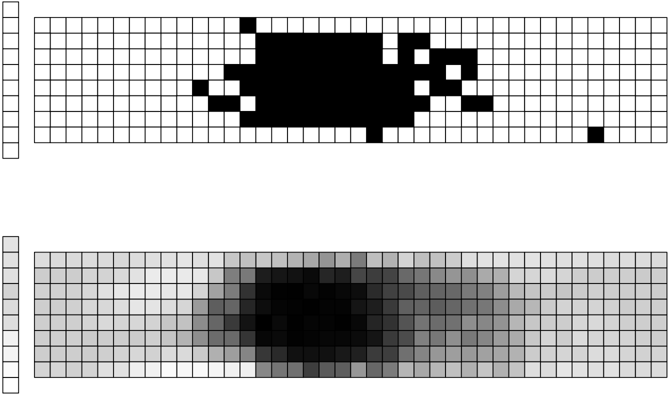


Projecting observed data on lower-dimensional-manifold

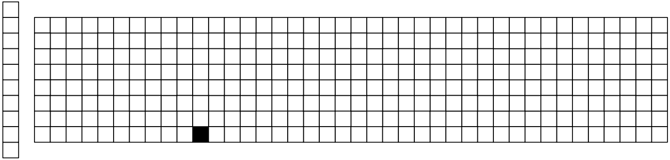


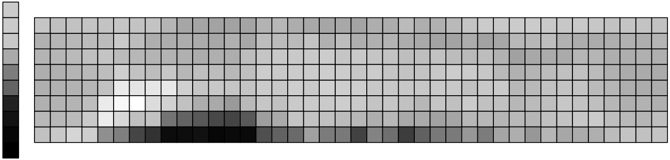


Projecting observed data on lower-dimensional-manifold

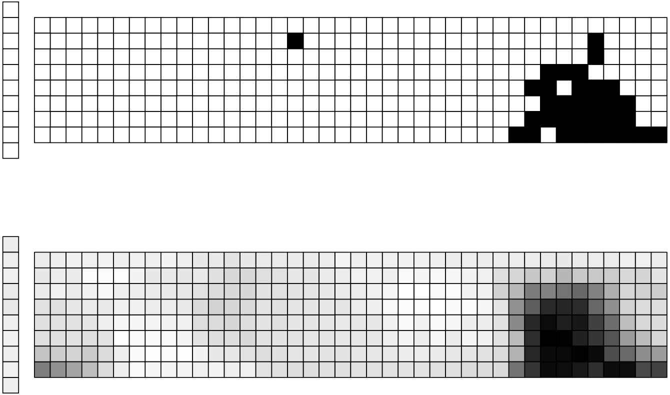


Projecting observed data on lower-dimensional-manifold

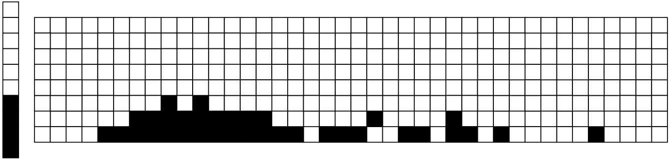


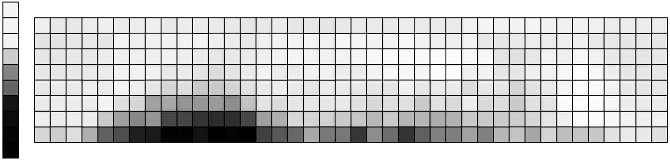


Projecting observed data on lower-dimensional-manifold



Projecting observed data on lower-dimensional-manifold





Implicative universals

- first six features correspond nicely to the six primary colors *white, black, red, green, blue, yellow*
- according to Kay et al. (1997) (and many other authors) simple system of **implicative universals** regarding possible partitions of the primary colors

Implicative universals

I	II	III	IV	V
		white red/yellow green/blue black	white red yellow green/blue black	
white/red/yellow black/green/blue	white red/yellow black/green/blue	white red/yellow green black/blue		white red yellow green blue black
		white red yellow black/green/blue	white red yellow green black/blue	
		white red yellow/green/blue black	white red yellow/green blue black	
		white red yellow/green black/blue		

source: Kay et al. (1997)

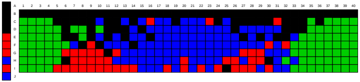
Partition of the primary colors

- each speaker/term pair can be projected to a 15-dimensional vector
- primary colors correspond to first 6 entries
- each primary color is assigned to the term for which it has the highest value
- defines for each speaker a partition over the primary colors

Partition of the primary colors

- for instance: sample speaker (from Piraha):
- extracted partition:

$$\begin{bmatrix} \text{white/yellow} \\ \text{red} \\ \text{green/blue} \\ \text{black} \end{bmatrix}$$



- supposedly impossible, but occurs 61 times in the database

Partition of primary colors

- most frequent partition types:
  1. {white}, {red}, {yellow}, {green, blue}, {black} (41.9%)
  2. {white}, {red}, {yellow}, {green}, {blue}, {black} (25.2%)
  3. {white}, {red, yellow}, {green, blue, black} (6.3%)
  4. {white}, {red}, {yellow}, {green}, {black, blue} (4.2%)
  5. {white, yellow}, {red}, {green, blue}, {black} (3.4%)
  6. {white}, {red}, {yellow}, {green, blue, black} (3.2%)
  7. {white}, {red, yellow}, {green, blue}, {black} (2.6%)
  8. {white, yellow}, {red}, {green, blue, black} (2.0%)
  9. {white}, {red}, {yellow}, {green, blue, black} (1.6%)
  10. {white}, {red}, {green, yellow}, {blue, black} (1.2%)

Partition of primay colors

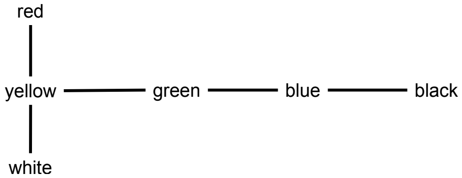
- 87.1% of all speaker partitions obey Kay et al.'s universals
- the ten partitions that confirm to the universals occupy ranks 1, 2, 3, 4, 6, 7, 9, 10, 16, 18
- decision what counts as an exception seems somewhat arbitrary on the basis of these counts

The semantic map of primary colors

- Manual inspection of the frequently occurring patterns shows that:
  - most speakers lump *green* and *blue* into one category ($\approx 63.2\%$)
  - many speakers lump *black* and *blue* into one category ($\approx 19.3\%$)
  - a fair amount of speakers lumps *red* and *yellow* into one category ($\approx 9.8\%$)
  - some speakers lump *white* and *yellow* into one category ($\approx 7.6\%$)
  - a few speakers even lump *green* and *yellow* into one category ($\approx 4.6\%$)

The semantic map of primary colors

- leads to a graph structure (a reviewer pointed out that this is a kind of semantic map):



(1)
a. All partition cells are continuous subgraphs of the connection graph.
b. No partition cell has more than three elements.
c. *Red* and *white* only occur in cells with at most two elements.

The semantic map of primary colors

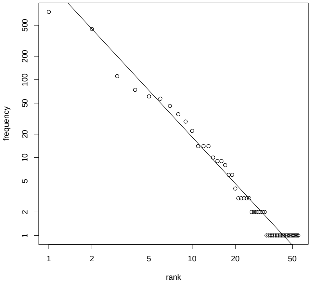
- three more partition types obey this constraint, which all occur in the data:
  - {green}, {white/yellow}, {red}, {black/blue} (14 occurrences)
  - {green}, {white/yellow}, {red}, {black}, {blue} (8 occurrences)
  - {green}, {white}, {red/yellow}, {black}, {blue} (2 occurrences)
- all predicted partition types occur in the data
- about 94% of the data fit to the model
- adding further links to the graph (*green-black*, *black-white*) improves the precision but reduces the recall

Power Laws

- more fundamental problem:
  - partition frequencies are distributed according to **power law**

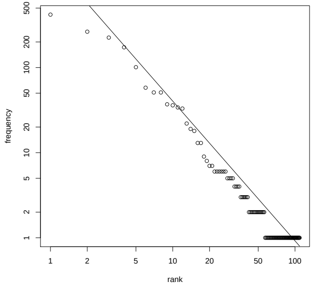
$$frequency \sim rank^{-1.99}$$

- no natural cutoff point to distinguish regular from exceptional partitions



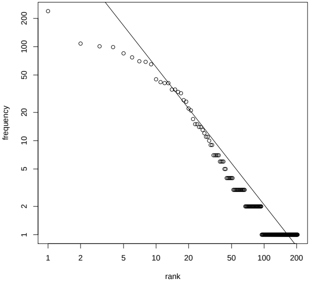
Partition of seven most important colors

$$frequency \sim rank^{-1.64}$$



Partition of eight most important colors

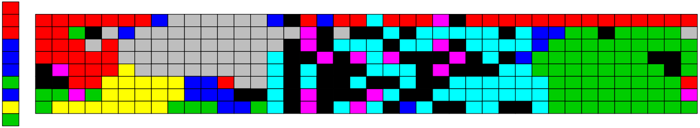
$$frequency \sim rank^{-1.46}$$

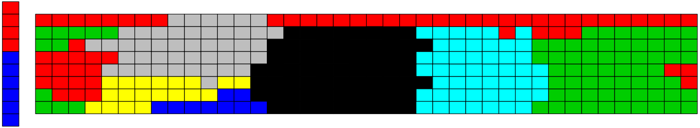


Smoothing the partitions

- from smoothed extensions we can recover smoothed partitions
- each pixel is assigned to category in which it has the highest degree of membership

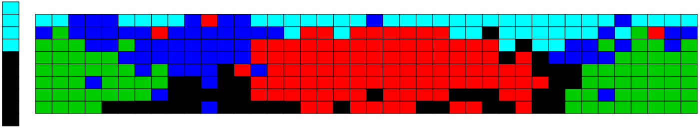
Smoothed partitions of the color space

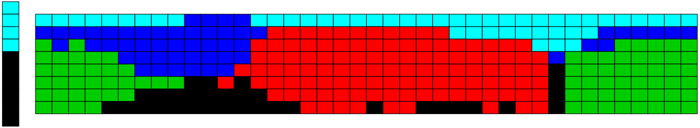




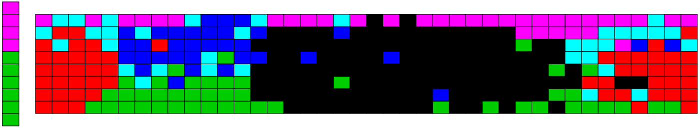
Smoothed partitions of the color space

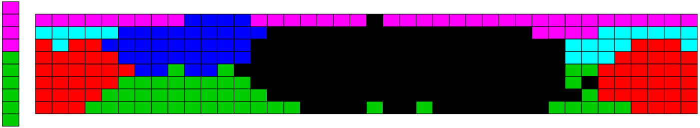
Smoothed partitions of the color space



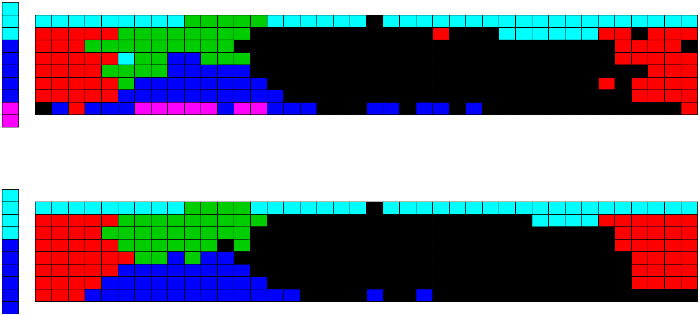


Smoothed partitions of the color space

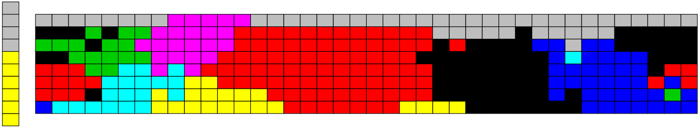


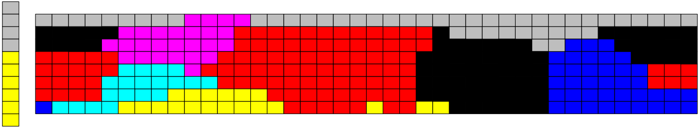


Smoothed partitions of the color space

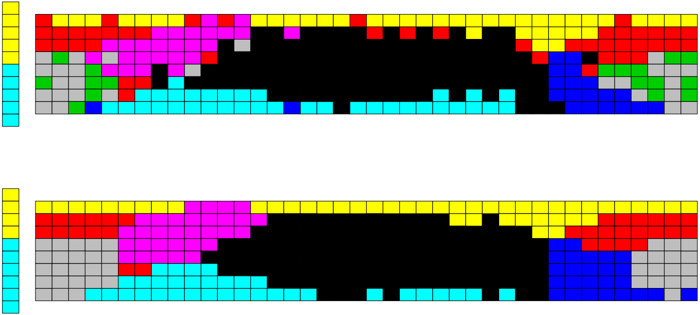


Smoothed partitions of the color space

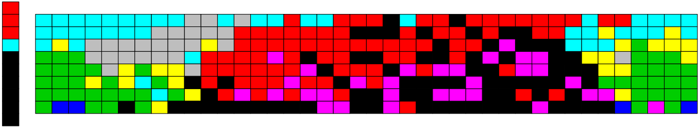


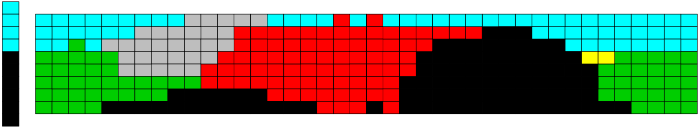


Smoothed partitions of the color space

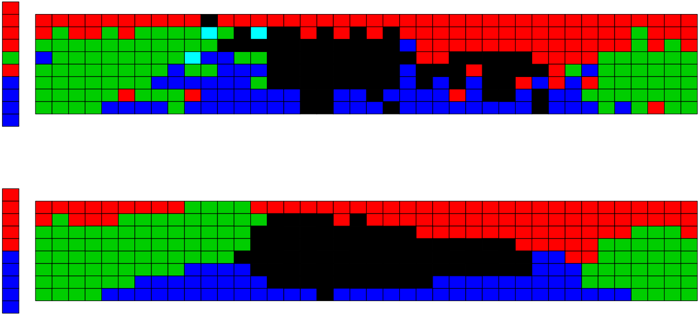


Smoothed partitions of the color space

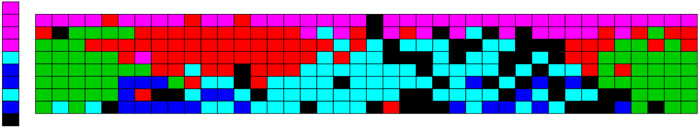


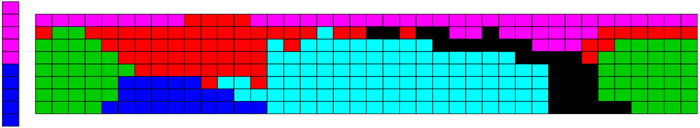


Smoothed partitions of the color space



Smoothed partitions of the color space



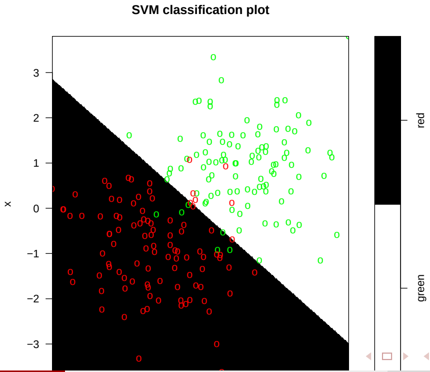


Convexity

- note: so far, we only used information from the WCS
- the location of the 330 Munsell chips in L*a*b* space played no role so far
- still, apparently partition cells always form continuous clusters in L*a*b* space
- Hypothesis (Gärdenfors): extension of color terms always form **convex** regions of L*a*b* space

Support Vector Machines

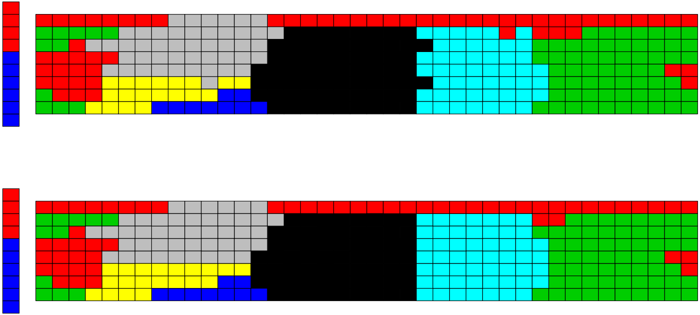
- supervised learning technique
- smart algorithm to classify data in a high-dimensional space by a (for instance) linear boundary
- minimizes number of mis-classifications if the training data are not linearly separable



Convex partitions

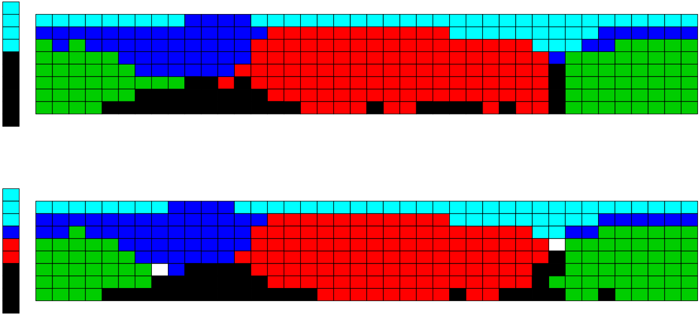
- a binary linear classifier divides an n-dimensional space into two **convex** half-spaces
- intersection of two convex set is itself convex
- hence: intersection of k binary classifications leads to convex sets
- procedure: if a language partitions the Munsell space into m categories, train $\frac{m(m-1)}{2}$ many binary SVMs, one for each pair of categories **in L*a*b* space**
- leads to m convex sets (which need not split the L*a*b* space exhaustively)

Convex approximation

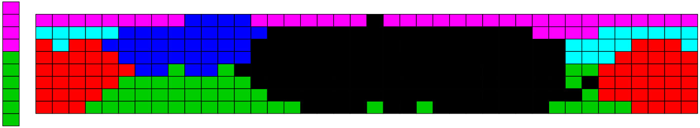


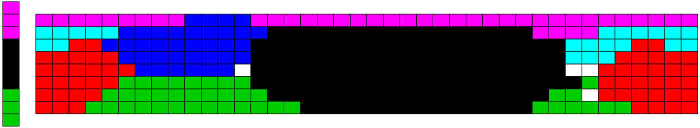
Convex approximation

Convex approximation

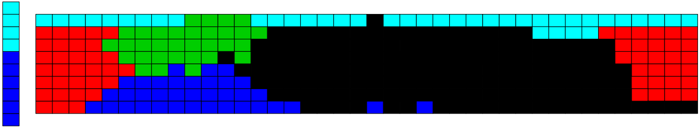


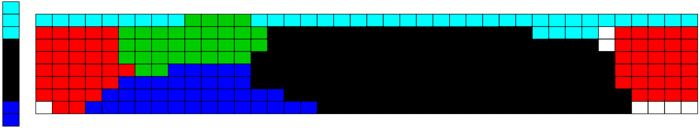
Convex approximation



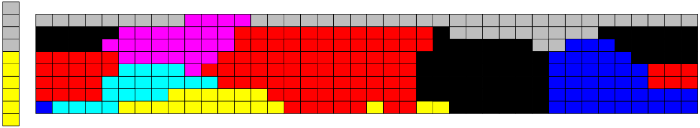


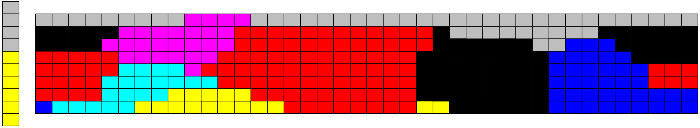
Convex approximation



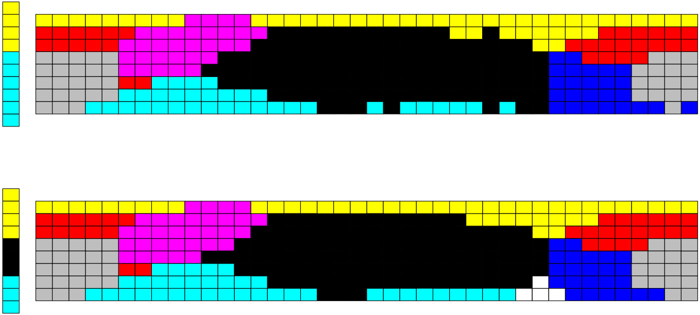


Convex approximation

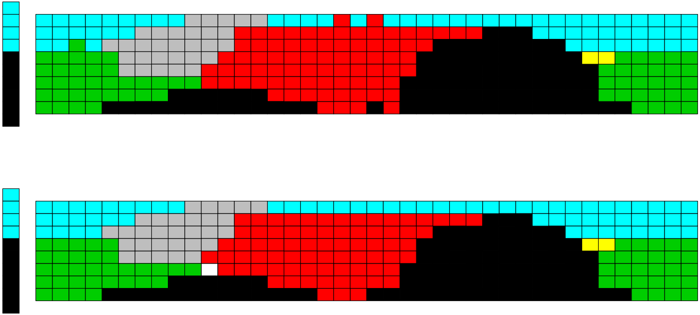




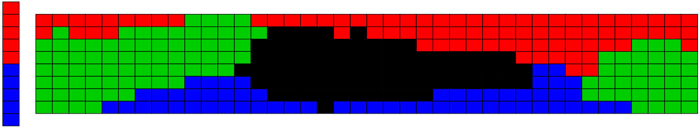
Convex approximation

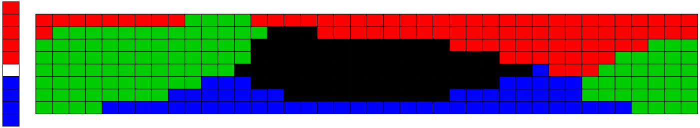


Convex approximation

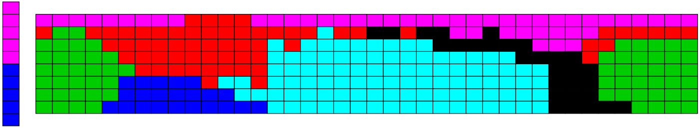


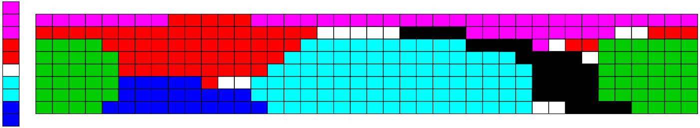
Convex approximation





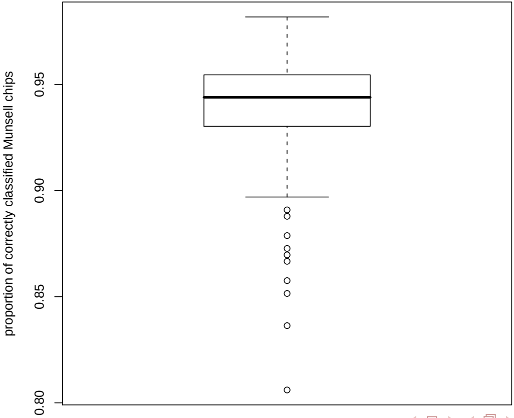
Convex approximation





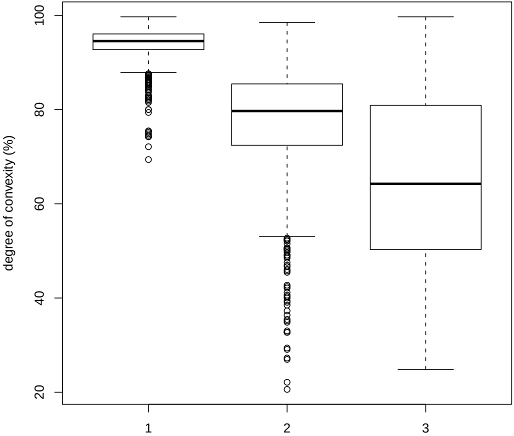
Convex approximation

- on average, 93.7% of all Munsell chips are correctly classified by convex approximation



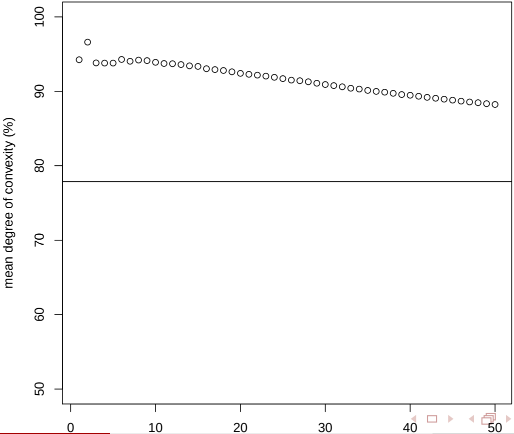
Convex approximation

- compare to the outcome of the same procedure without PCA, and with PCA but using a random permutation of the Munsell chips

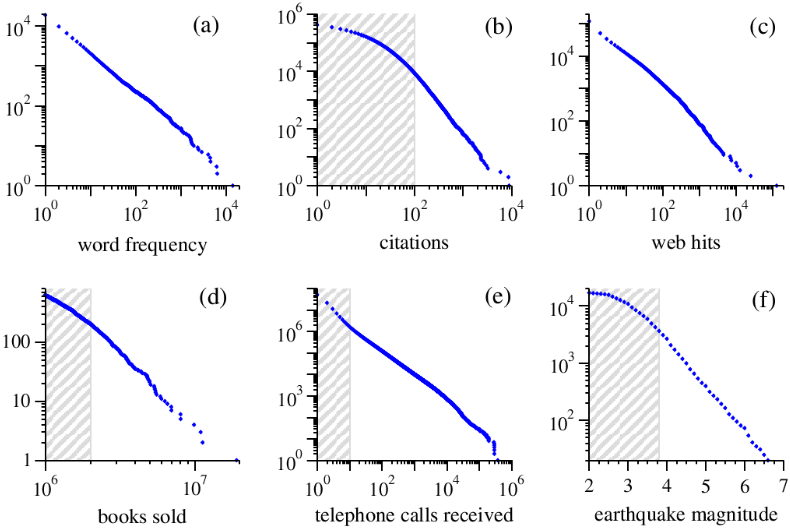


Convex approximation

- choice of $m = 10$ is somewhat arbitrary
- outcome does not depend very much on this choice though

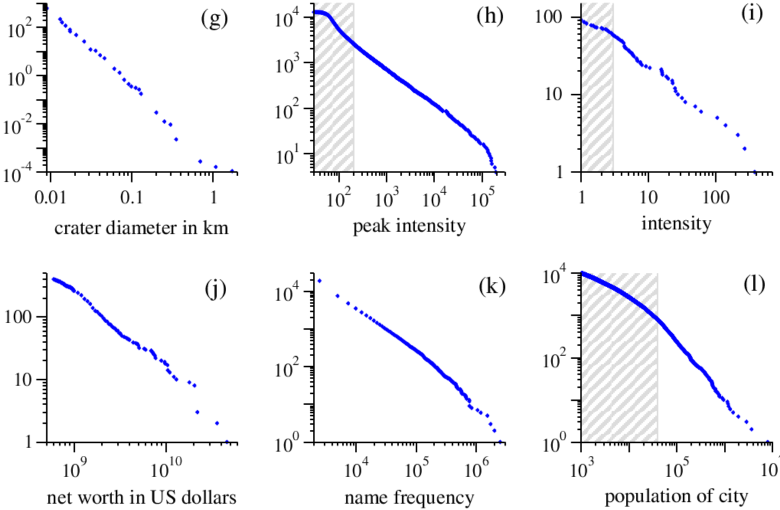


Power laws



(a)
10^4
10^2
10^0
10^0
10^2
10^4
word frequency
(b)
10^6
10^4
10^2
10^0
10^0
10^2
10^4
citations
(c)
10^4
10^2
10^0
10^0
10^2
10^4
web hits
(d)
100
10
1
10^6
10^7
books sold
(e)
10^6
10^3
10^0
10^0
10^2
10^4
10^6
telephone calls received
(f)
10^4
10^3
10^2
2
3
4
5
6
7
earthquake magnitude

Power laws



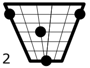
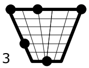
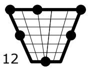
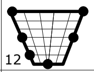
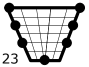
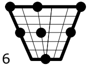
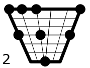
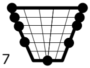
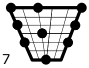
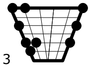
(g)
crater diameter in km
(h)
peak intensity
(i)
intensity
(j)
net worth in US dollars
(k)
name frequency
(l)
population of city

Power laws

FIG. 4 Cumulative distributions or "rank/frequency plots" of twelve quantities reputed to follow power laws. The distributions were computed as described in Appendix A. Data in the shaded regions were excluded from the calculations of the exponents in Table I. Source references for the data are given in the text. (a) Numbers of occurrences of words in the novel *Moby Dick* by Hermann Melville. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (c) Numbers of hits on web sites by 60 000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of copies of bestselling books sold in the US between 1895 and 1965. (e) Number of calls received by AT&T telephone customers in the US for a single day. (f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear. (g) Diameter of craters on the moon. Vertical axis is measured per square kilometre. (h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989. (i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries. (j) Aggregate net worth in dollars of the richest individuals in the US in October 2003. (k) Frequency of occurrence of family names in the US in the year 1990. (l) Populations of US cities in the year 2000.

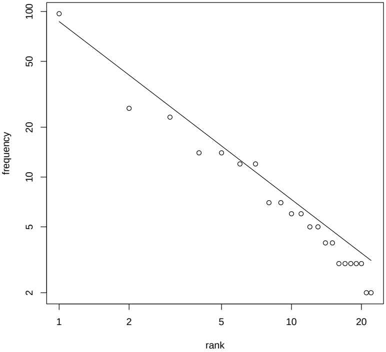
from Newman 2006

Other linguistic power law distributions

number of vowels	vowel systems and their frequency of occurrence				
3	 14				
4	 14	 5	 4	 2	
5	 97	 3			
6	 26	 12	 12		
7	 23	 6	 5	 4	 3
8	 6	 3	 3	 2	
9	 7	 7	 3		

Other linguistic power law distributions

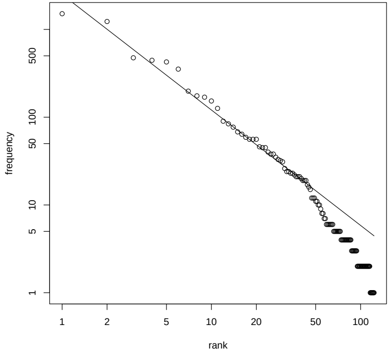
$$frequency \sim rank^{-1.06}$$



Other linguistic power law distributions

- size of language families
- source: Ethnologue

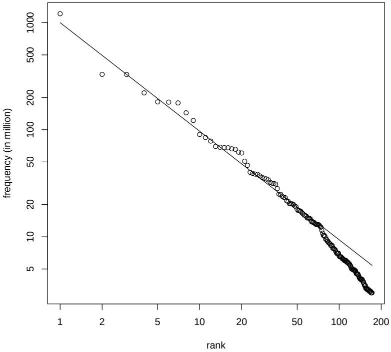
$$frequency \sim rank^{-1.32}$$



Other linguistic power law distributions

- number of speakers per language
- source: Ethnologue

$$frequency \sim rank^{-1.01}$$



The World Atlas of Language Structures

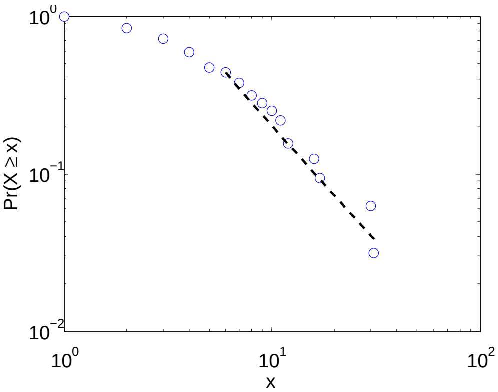
- large scale typological database, conducted mainly by the MPI EVA, Leipzig
- 2,650 languages in total are used
- 142 features, with between 120 and 1,370 languages per feature
- available online

The World Atlas of Language Structures

- question: are frequency of feature values powerlaw distributed?
- problem: number of feature values usually too small for statistic evaluation
- solution:
  - cross-classification of two (randomly chosen) features
  - only such feature pairs are considered that lead to at least 30 non-empty feature value combinations
- pilot study with 10 such feature pairs

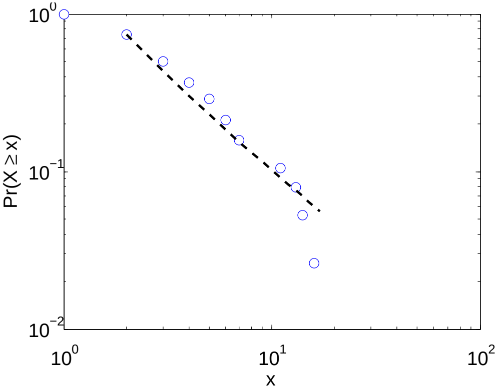
The World Atlas of Language Structures

- Feature 1: Consonant-Vowel Ratio
- Feature 2: Subtypes of Asymmetric Standard Negation
- Kolmogorov-Smirnov test: positive



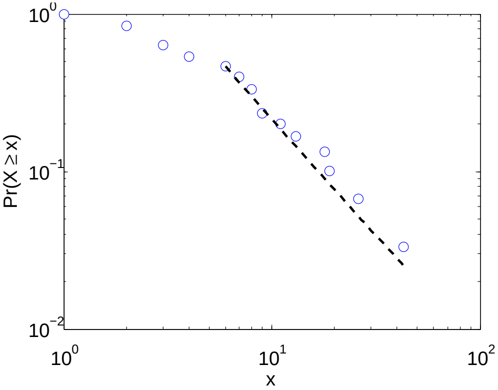
The World Atlas of Language Structures

- Feature 1: Weight Factors in Weight-Sensitive Stress Systems
- Feature 2: Ordinal Numerals
- Kolmogorov-Smirnov test: positive



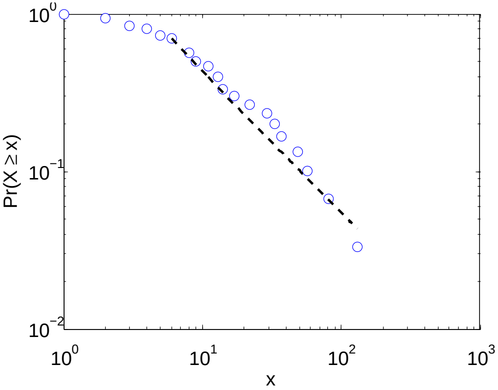
The World Atlas of Language Structures

- Feature 1: Third Person Zero of Verbal Person Marking
- Feature 2: Subtypes of Asymmetric Standard Negation
- Kolmogorov-Smirnov test: positive



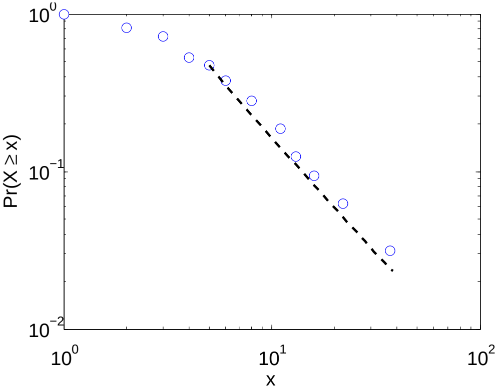
The World Atlas of Language Structures

- Feature 1: Relationship between the Order of Object and Verb and the Order of Adjective and Noun
- Feature 2: Expression of Pronominal Subjects
- Kolmogorov-Smirnov test: positive



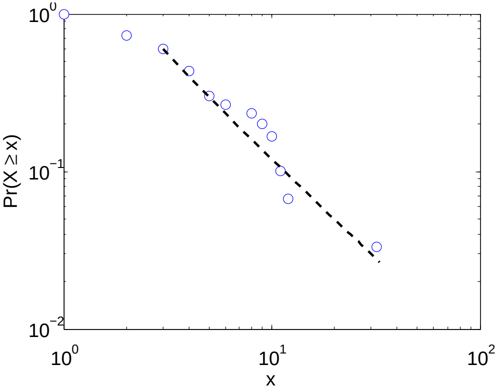
The World Atlas of Language Structures

- Feature 1: Plurality in Independent Personal Pronouns
- Feature 2: Asymmetrical Case-Marking
- Kolmogorov-Smirnov test: positive



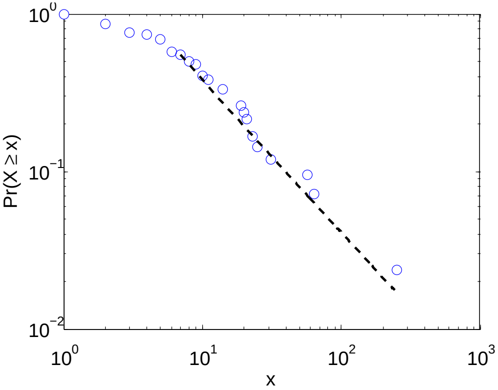
The World Atlas of Language Structures

- Feature 1: Locus of Marking: Whole-language Typology
- Feature 2: Number of Cases
- Kolmogorov-Smirnov test: positive



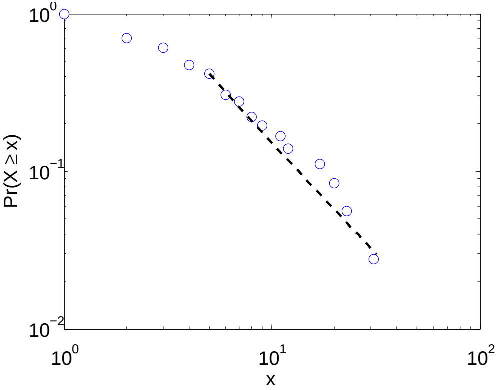
The World Atlas of Language Structures

- Feature 1: Prefixing vs. Suffixing in Inflectional Morphology
- Feature 2: Coding of Nominal Plurality
- Kolmogorov-Smirnov test: positive



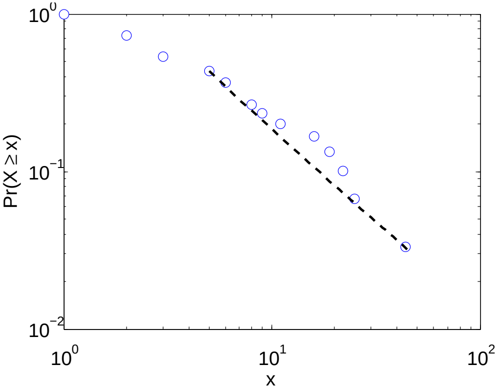
The World Atlas of Language Structures

- Feature 1: Prefixing vs. Suffixing in Inflectional Morphology
- Feature 2: Ordinal Numerals
- Kolmogorov-Smirnov test: positive



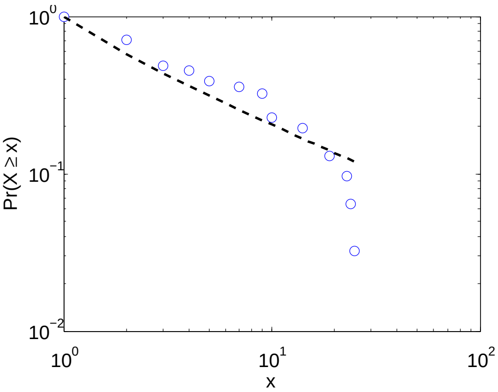
The World Atlas of Language Structures

- Feature 1: Coding of Nominal Plurality
- Feature 2: Asymmetrical Case-Marking
- Kolmogorov-Smirnov test: positive



The World Atlas of Language Structures

- Feature 1: Position of Case Affixes
- Feature 2: Ordinal Numerals
- Kolmogorov-Smirnov test: negative



Why power laws?

- critical states
- self-organized criticality
- preferential attachment
- random walks
- ...

Preferential attachment

- items are stochastically added to bins
- probability to end up in bin n is linear in number of items that are already in bin n

(Wide) Open questions

- Preferential attachment explains power law distribution *if there are no a priori biases for particular types*
- first simulations suggest that preferential attachment + biased type assignment does **not** lead to power law
- negative message: uneven typological frequency distribution does not prove that frequent types are inherently preferred linguistically/cognitively/socially
- unsettling questions:
  - Are there linguistic/cognitive/social biases in favor of certain types?
  - If yes, can statistical typology supply information about this?
  - If power law distributions are the norm, is their any content to the notion of *statistical universal* in a Greenbergian sense?