What explains power laws in language typology and language change?

Gerhard Jäger
gerhard.jaeger@uni-tuebingen.de

UTübingen

Freiburg, January 19, 2011

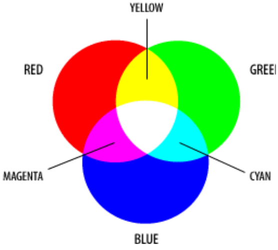
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Overview

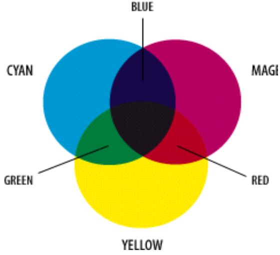
- Typological distribution of color naming systems
- Power laws
- Computer simulations

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

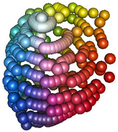


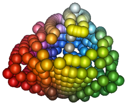
subtractive color space

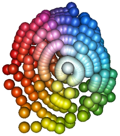
The psychological color space

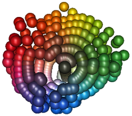
- psychologically correct color space should not only correctly represent the topology of, but also the distances between colors
- distance is inverse function of perceived similarity
- L*a*b* color space has this property
- three axes:
  - black — white
  - red — green
  - blue — yellow
- irregularly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

The color solid



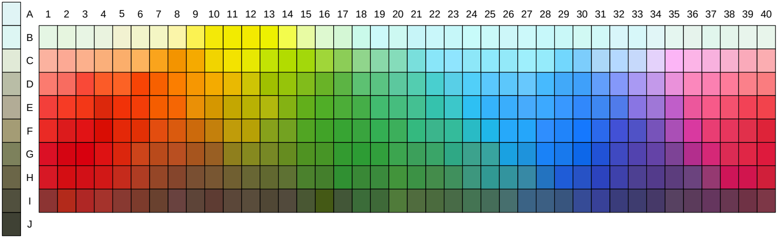






The Munsell chart

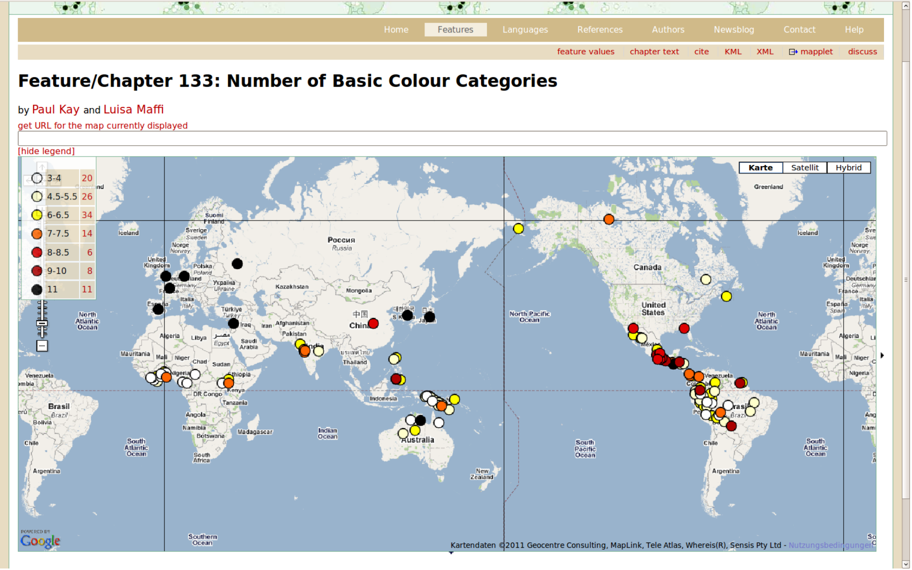
- for psychological investigations, the *Munsell chart* is being used
- 2d-rendering of the surface of the color solid
  - 8 levels of lightness
  - 40 hues
- plus: black–white axis with 8 shaded of grey in between
- neighboring chips differ in the minimally perceivable way



The World Color Survey

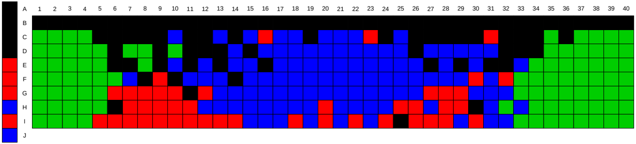
- started by Paul Kay and co-workers; traces back to Berlin & Kay 1969
- investigation of color vocabulary of 110 non-written languages from around the world
- around 25 informants per language
- two tasks:
  - the 330 Munsell chips were presented to each test person one after the other in random order; 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/

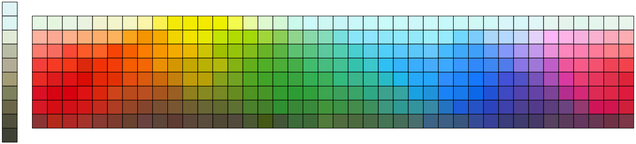
The World Color Survey



Raw data

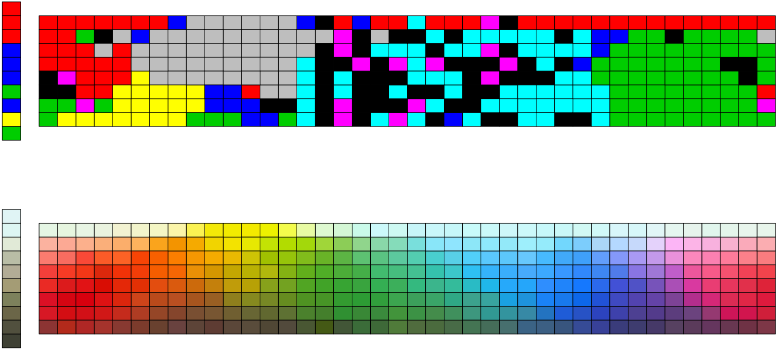
- are irregular and noisy
- example: randomly picked test person (native language: Piraha)
- 1,771 such data points in total





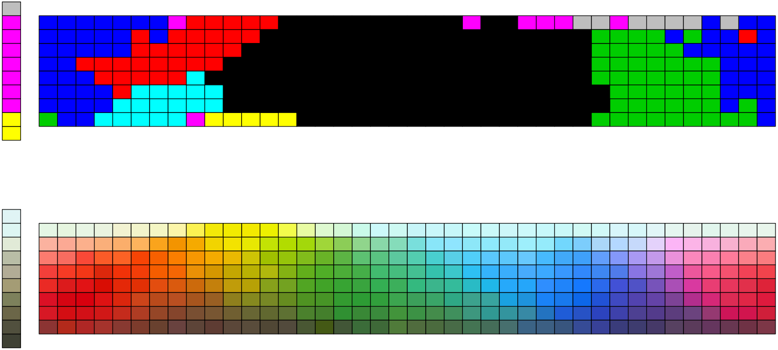
Raw data

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



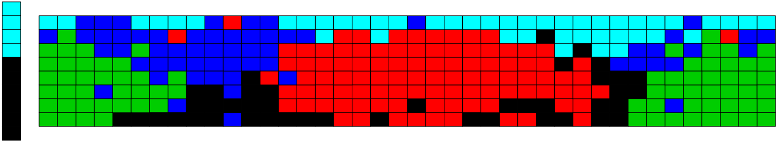
Raw data

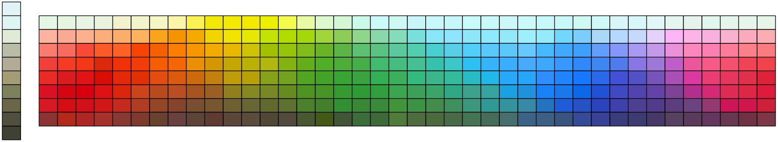
- partition of a randomly chosen informant from a randomly chosen language



Raw data

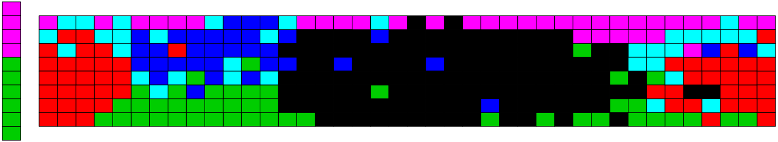
- partition of a randomly chosen informant from a randomly chosen language

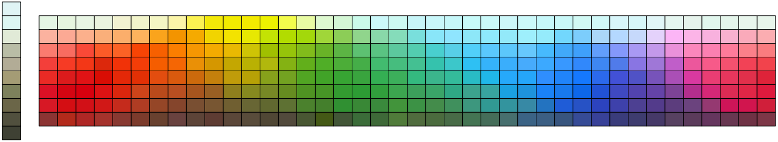




Raw data

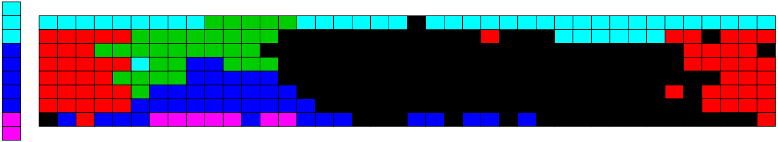
- partition of a randomly chosen informant from a randomly chosen language

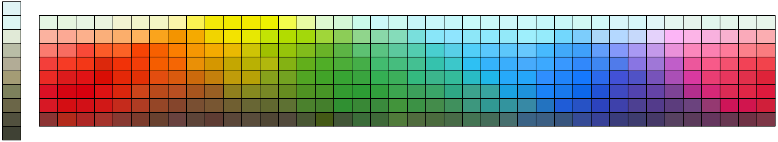




Raw data

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





Statistical feature extraction

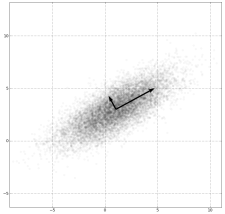
- first step: representation of raw data in *contingency matrix*
  - rows: color terms from various languages
  - columns: Munsell chips
  - cells: number of test persons who used the row-term for the column-chip

	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

Statistical feature extraction: PCA

- 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:

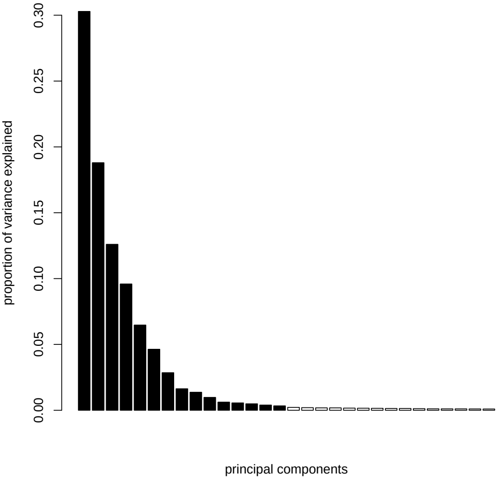
- choose a suitable $m < n$
- project the data on those m new coordinates where the data have the highest variance

Statistical feature extraction: PCA

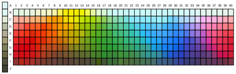
- alternative formulation:
  - choose an m-dimensional linear sub-manifold of your n-dimensional space
  - project your data onto this manifold
  - when doing so, pick your sub-manifold such that the average squared distance of the data points from the sub-manifold is minimized
- intuition behind this 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

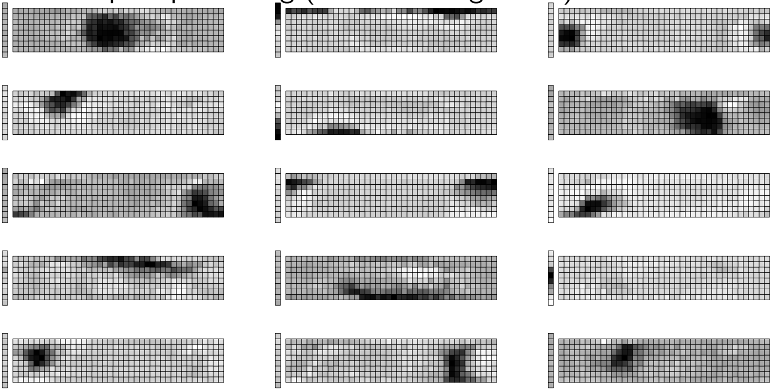
- first 15 principal components jointly explain 91.6% of the total variance
- choice of $m = 15$ is determined by using "Kaiser's stopping rule"



Statistical feature extraction: PCA



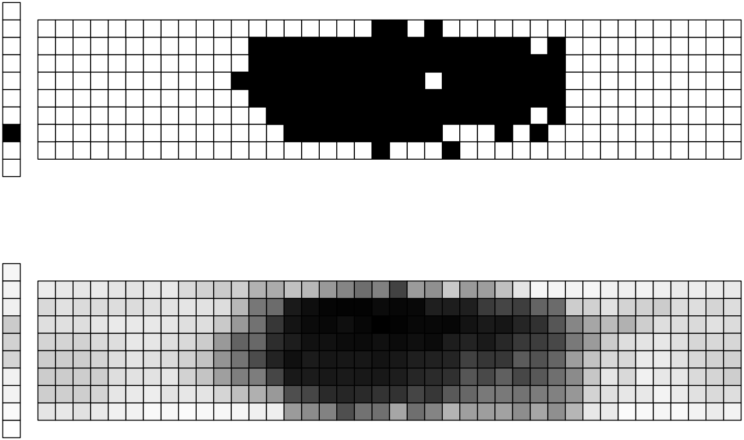
after some post-processing ("varimax" algorithm):



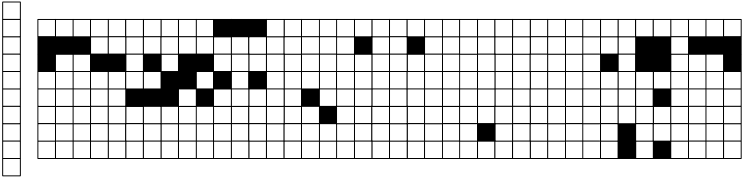
Projecting observed data on lower-dimensional-manifold

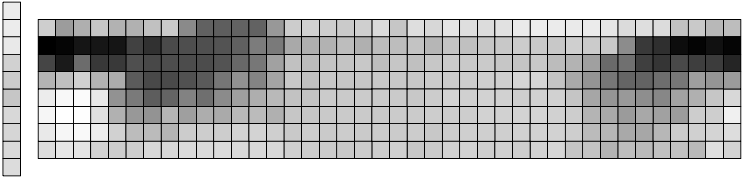
- noise removal: project observed data onto the lower-dimensional submanifold that was obtained via PCA
- in our case: noisy binary categories are mapped to smoothed fuzzy categories (= probability distributions over Munsell chips)
- some examples:

Projecting observed data on lower-dimensional-manifold

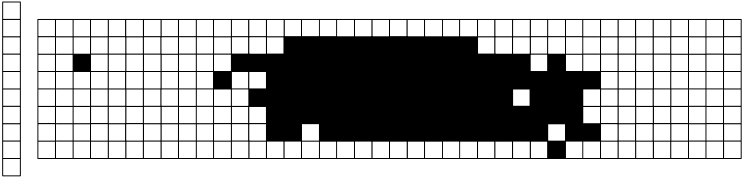


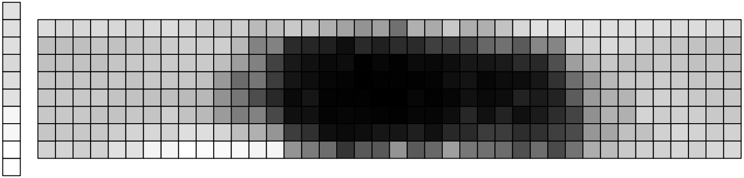
Projecting observed data on lower-dimensional-manifold



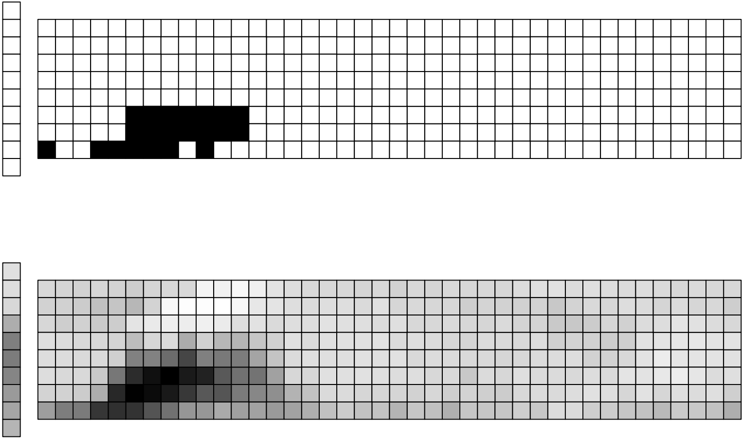


Projecting observed data on lower-dimensional-manifold

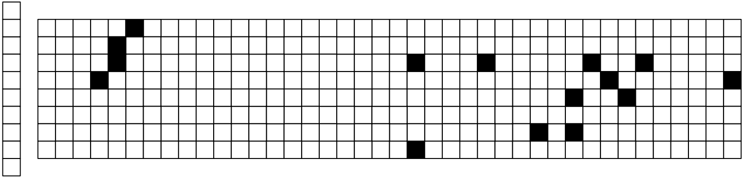


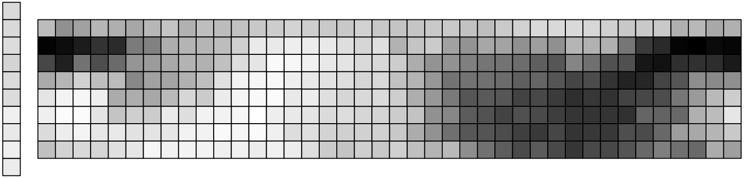


Projecting observed data on lower-dimensional-manifold

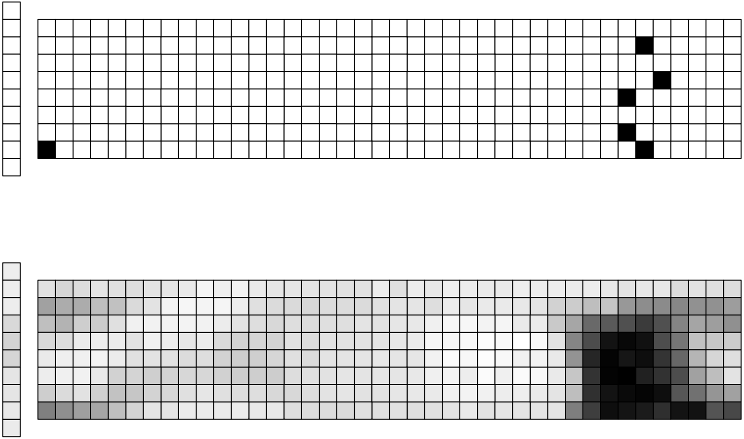


Projecting observed data on lower-dimensional-manifold

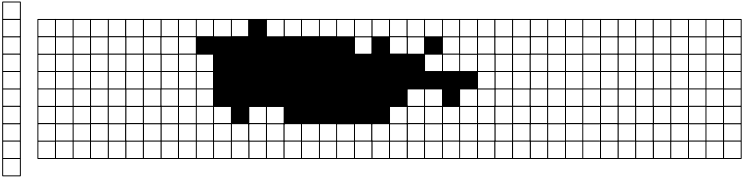


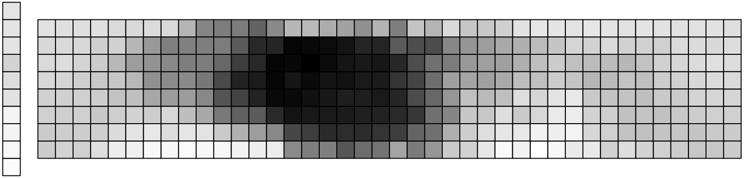


Projecting observed data on lower-dimensional-manifold

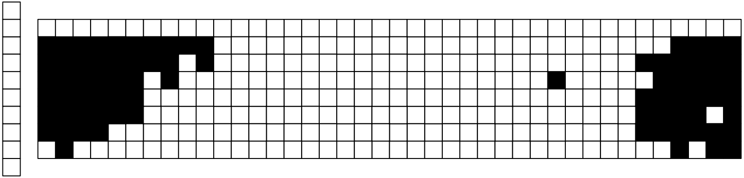


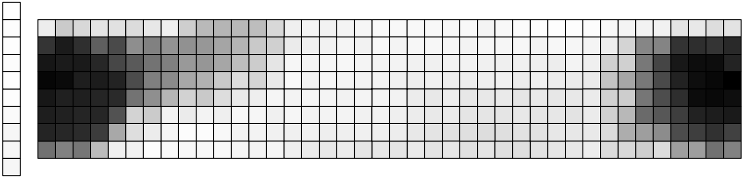
Projecting observed data on lower-dimensional-manifold



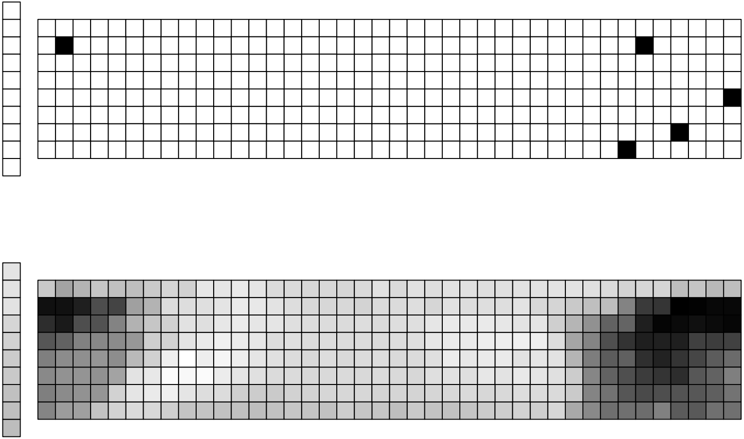


Projecting observed data on lower-dimensional-manifold

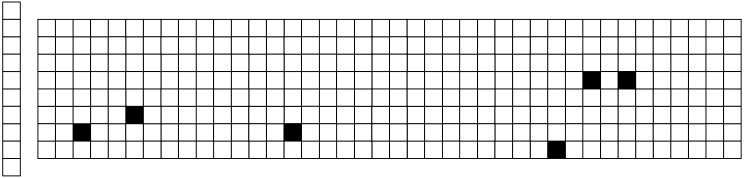


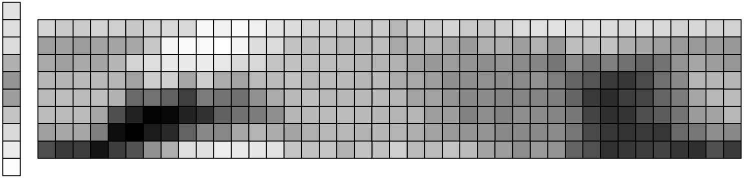


Projecting observed data on lower-dimensional-manifold

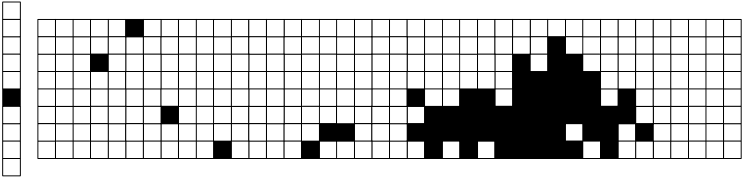


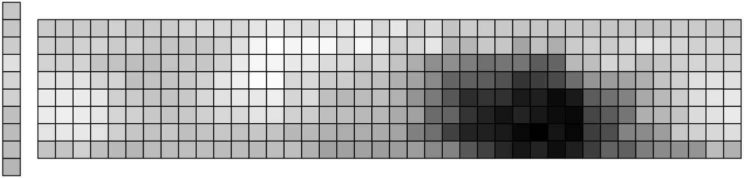
Projecting observed data on lower-dimensional-manifold



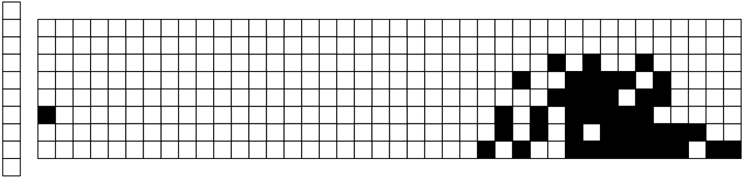


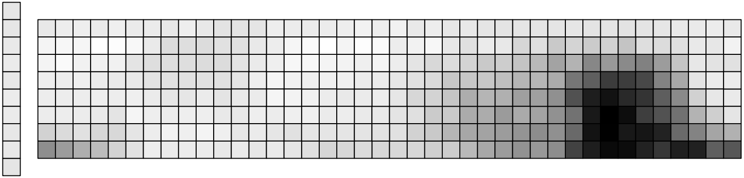
Projecting observed data on lower-dimensional-manifold



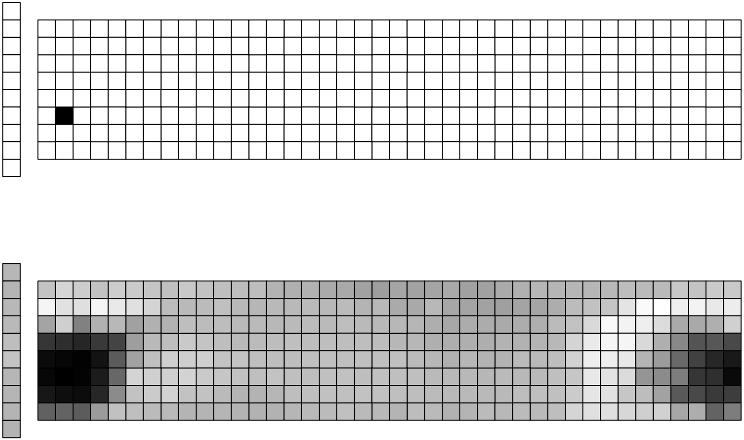


Projecting observed data on lower-dimensional-manifold

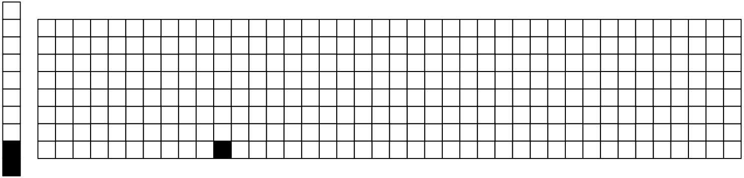


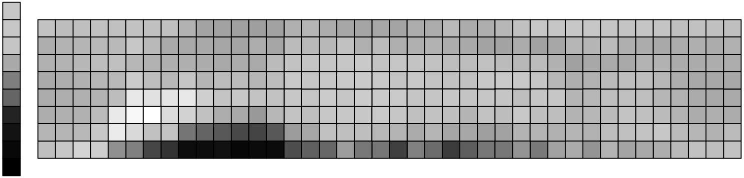


Projecting observed data on lower-dimensional-manifold

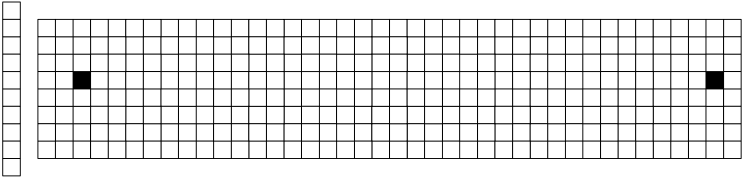


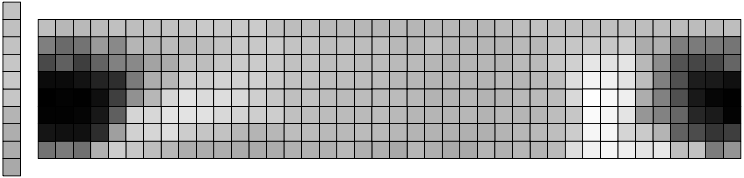
Projecting observed data on lower-dimensional-manifold



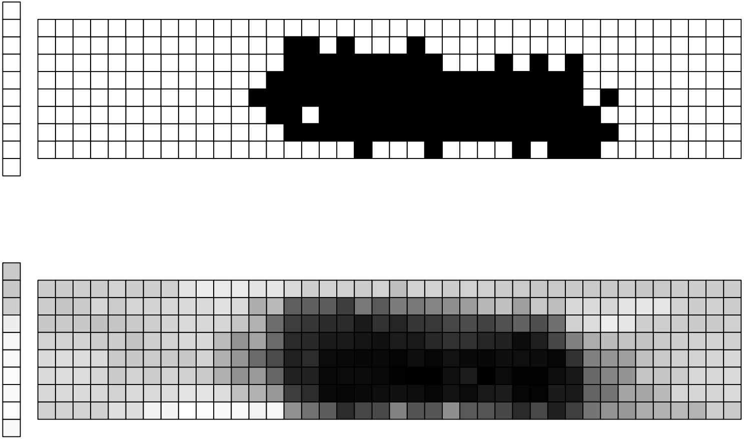


Projecting observed data on lower-dimensional-manifold

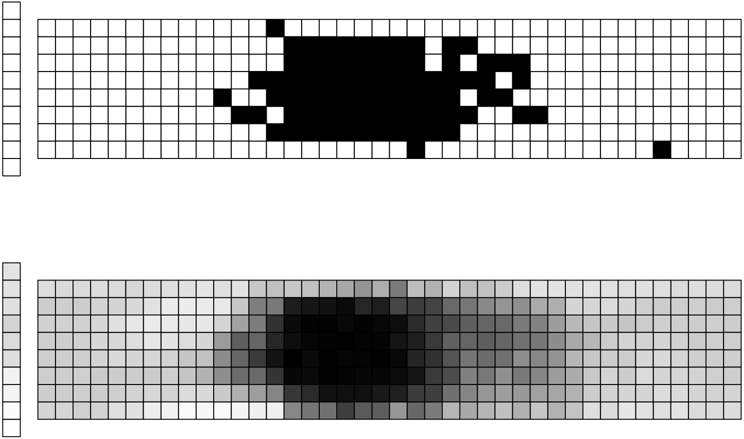




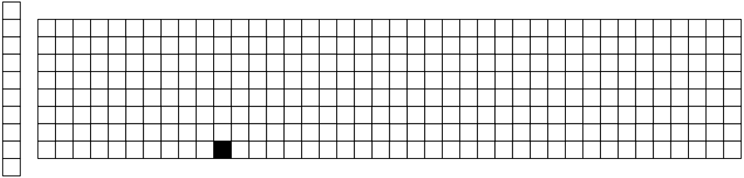
Projecting observed data on lower-dimensional-manifold

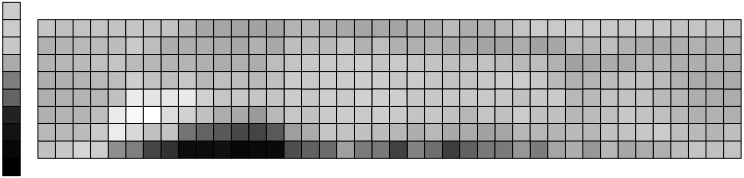


Projecting observed data on lower-dimensional-manifold

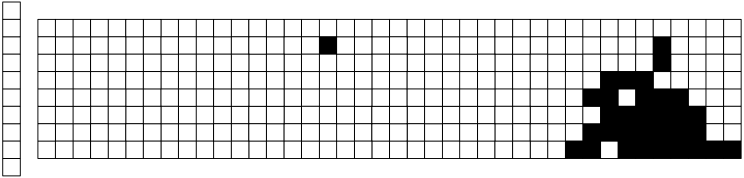


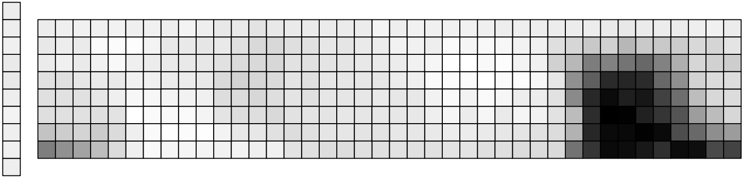
Projecting observed data on lower-dimensional-manifold



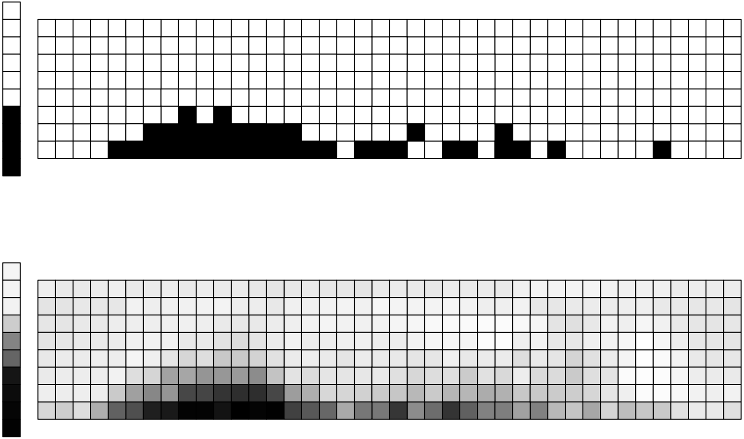


Projecting observed data on lower-dimensional-manifold





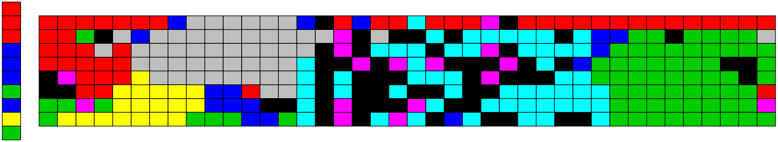
Projecting observed data on lower-dimensional-manifold

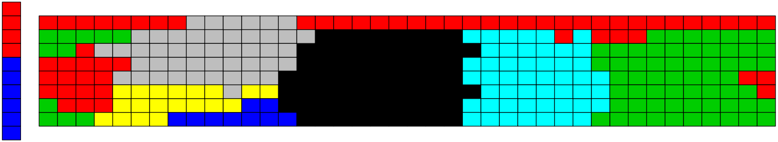


Smoothing the partitions

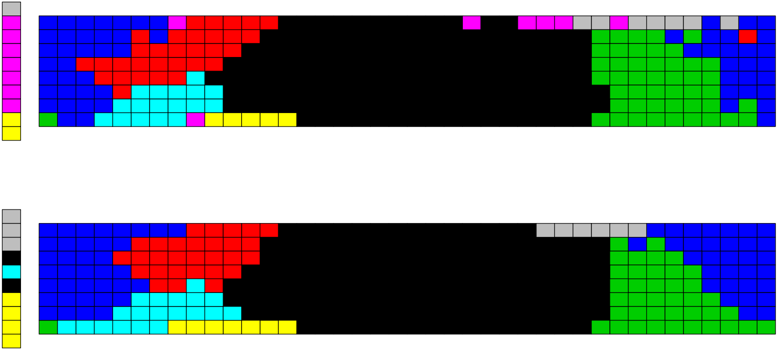
- from smoothed extensions we can recover smoothed partitions
- each pixel is assigned to category in which it has the highest degree of membership

Smoothed partitions of the color space

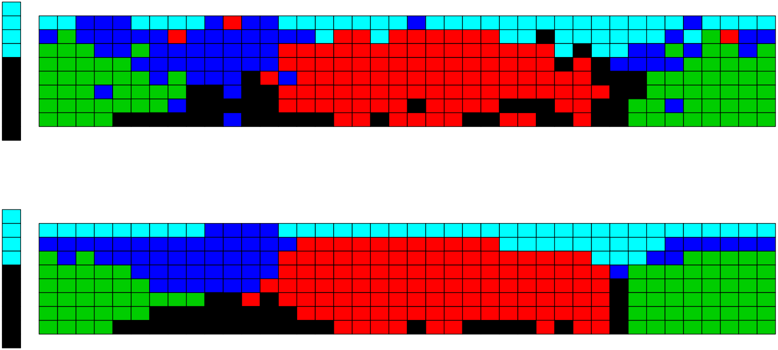




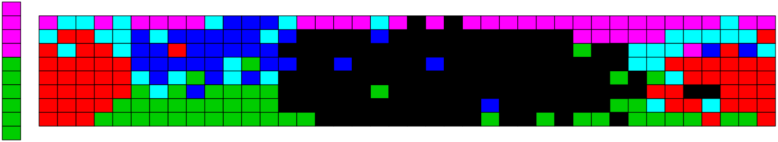
Smoothed partitions of the color space

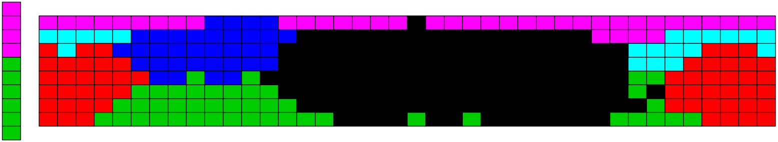


Smoothed partitions of the color space

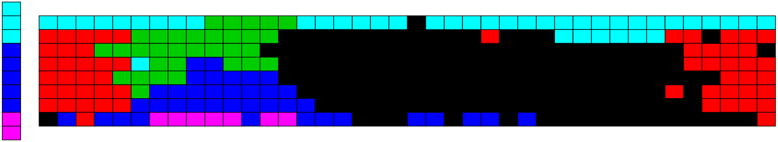


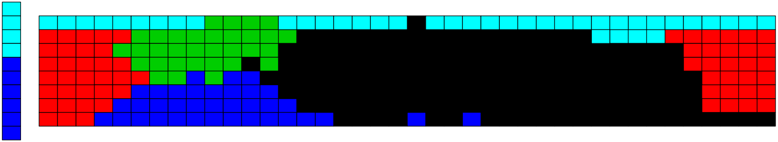
Smoothed partitions of the color space



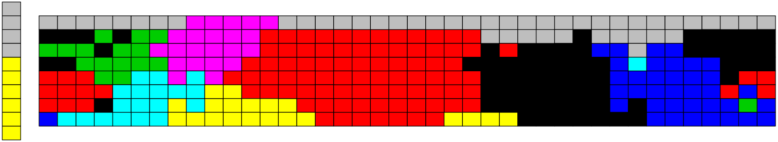


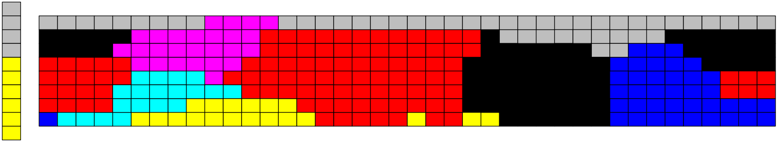
Smoothed partitions of the color space



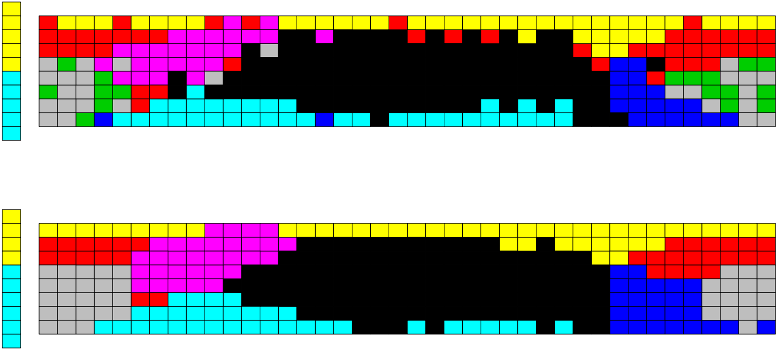


Smoothed partitions of the color space

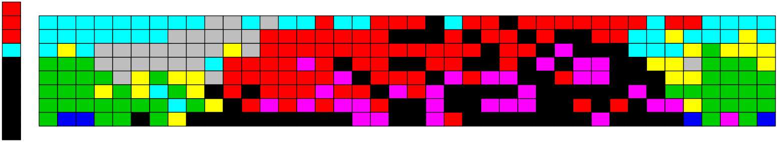


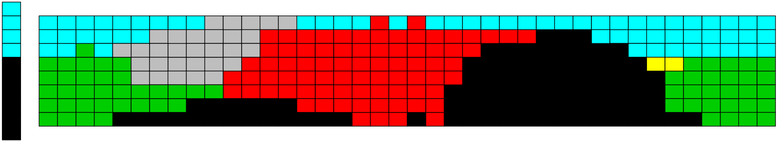


Smoothed partitions of the color space

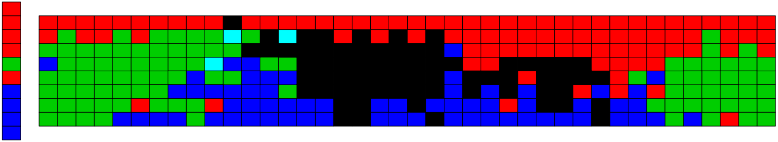


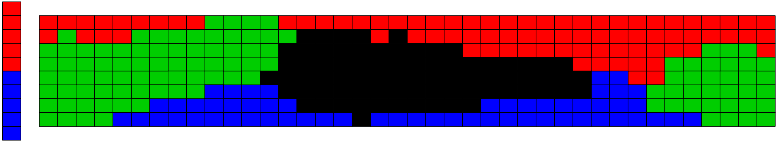
Smoothed partitions of the color space



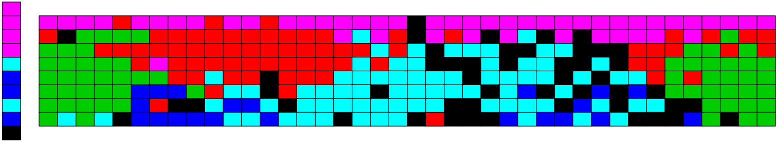


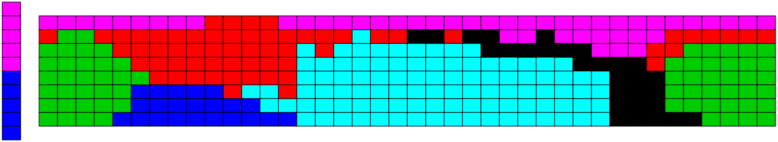
Smoothed partitions of the color space





Smoothed partitions of the color space



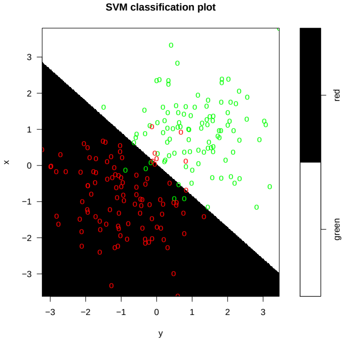


Convexity

- 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

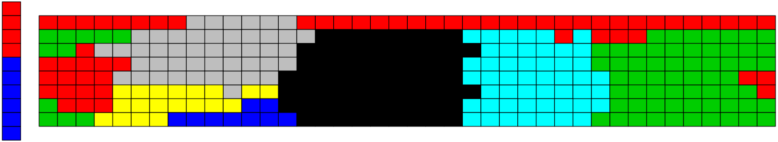
- supervised learning technique
- smart algorithm to classify data in a high-dimensional space by a (for instance) linear boundary
- minimizes number of mis-classifications if the training data are not linearly separable

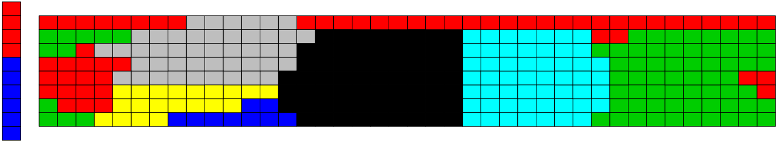


Convex partitions

- 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)

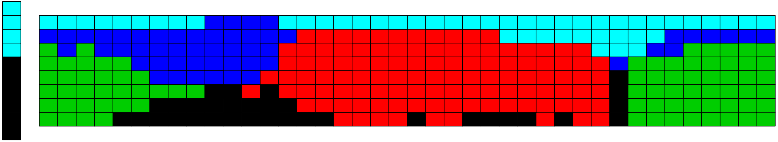
Convex approximation

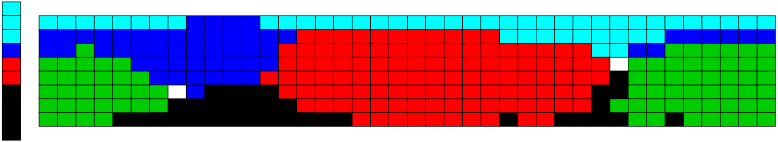




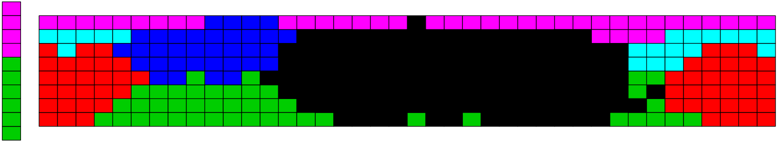
Convex approximation

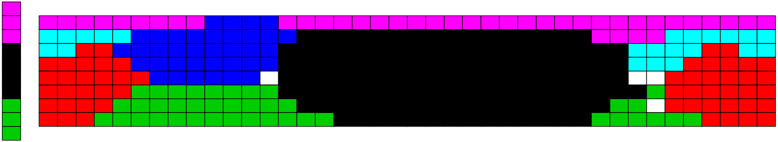
Convex approximation



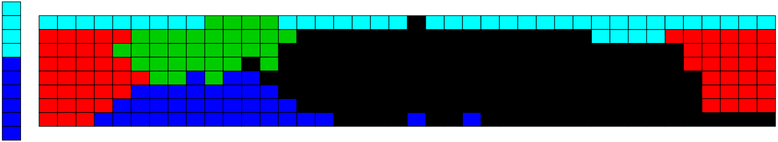


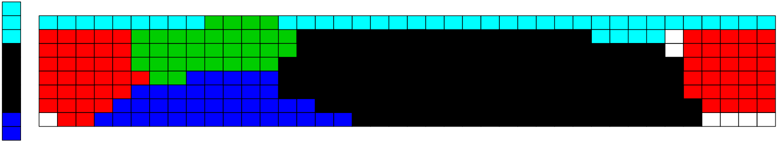
Convex approximation



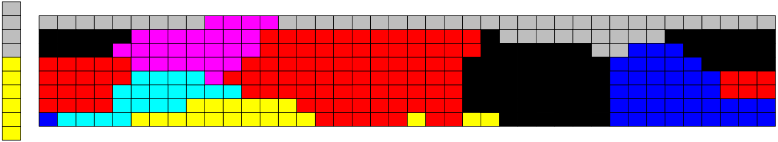


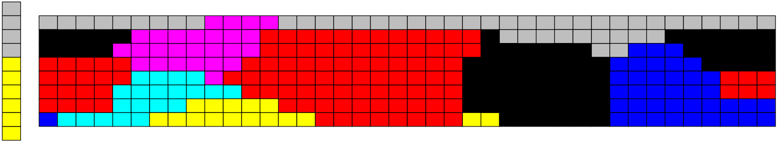
Convex approximation



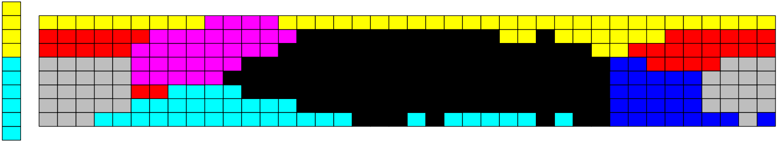


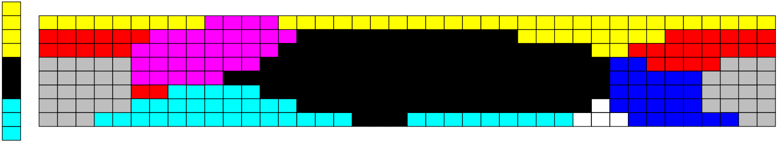
Convex approximation



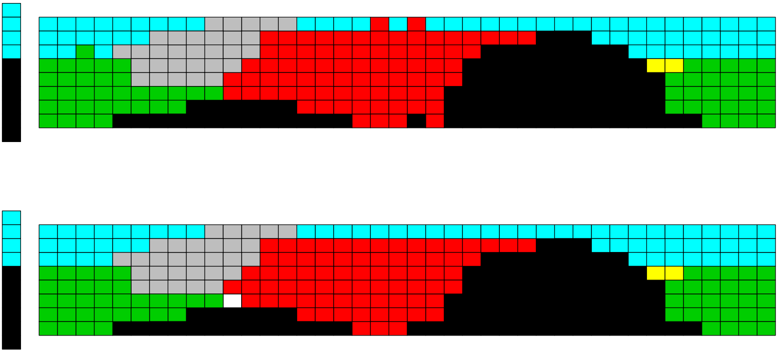


Convex approximation

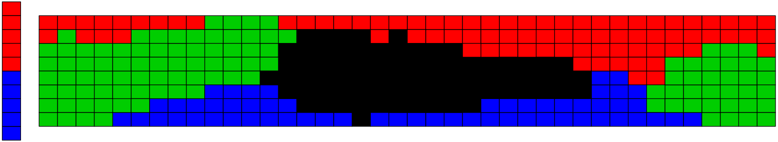


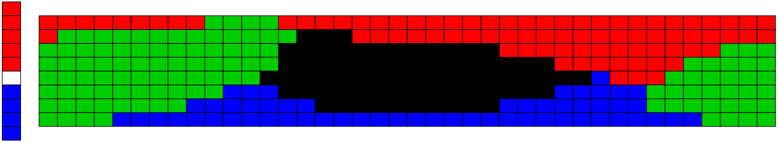


Convex approximation

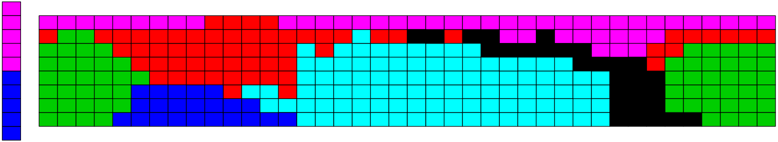


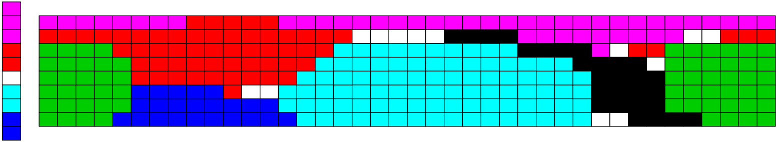
Convex approximation





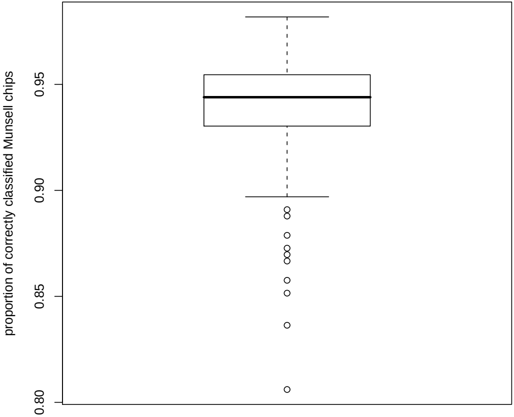
Convex approximation





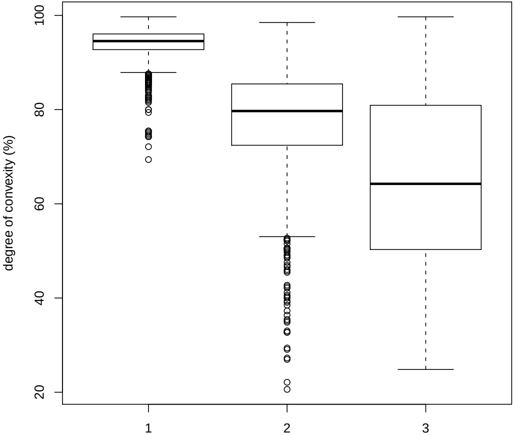
Convex approximation

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



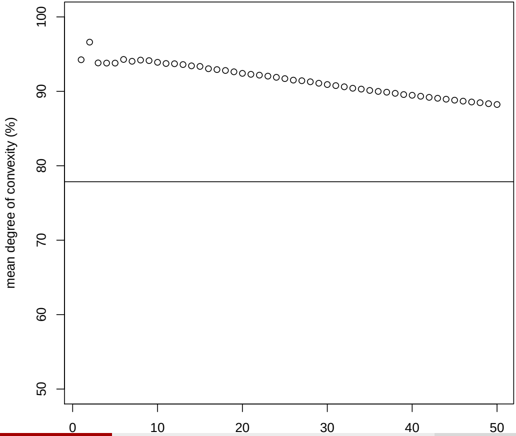
Convex approximation

- 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



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

[white/red/yellow, black/green/blue]

II

[white, red/yellow, black/green/blue]

III

[white, red/yellow, green/blue, black]

[white, red/yellow, green, black/blue]

[white, red, yellow, black/green/blue]

[white, red, yellow/green/blue, black]

[white, red, yellow/green, black/blue]

IV

[white, red, yellow, green/blue, black]

[white, red, yellow, green, black/blue]

[white, red, yellow/green, blue, black]

V

[white, red, yellow, green, blue, black]

source: Kay et al. (1997)

Partition of the primary colors

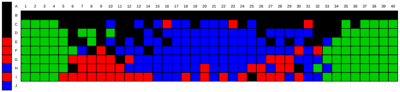
- each speaker/term pair can be projected to a 15-dimensional vector
- primary colors correspond to first 6 entries
- each primary color is assigned to the term for which it has the highest value
- defines for each speaker a partition over the primary colors

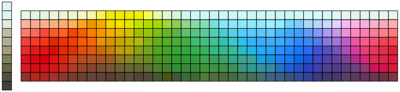
Partition of the primary colors

- for instance: sample speaker from Piraha (see above):
- 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

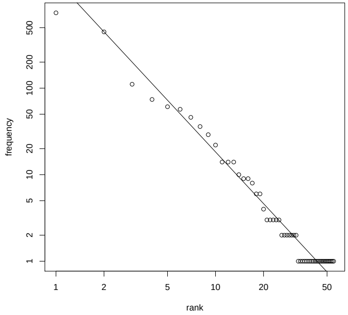
- 87.1% of all speaker partitions obey Kay et al.'s universals
- the ten partitions that confirm to the universals occupy ranks 1, 2, 3, 4, 6, 7, 9, 10, 16, 18
- decision what counts as an exception seems somewhat arbitrary on the basis of these counts

Partition of primary colors

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

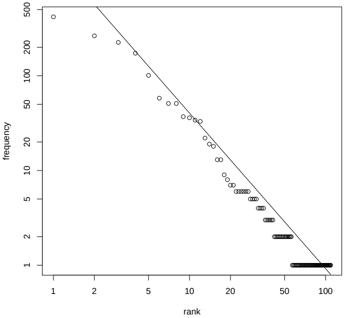
$$frequency \sim rank^{-1.99}$$

- no natural cutoff point to distinguish regular from exceptional partitions



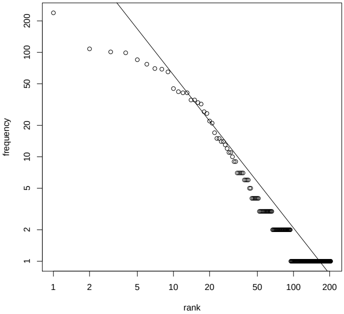
Partition of seven most important colors

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

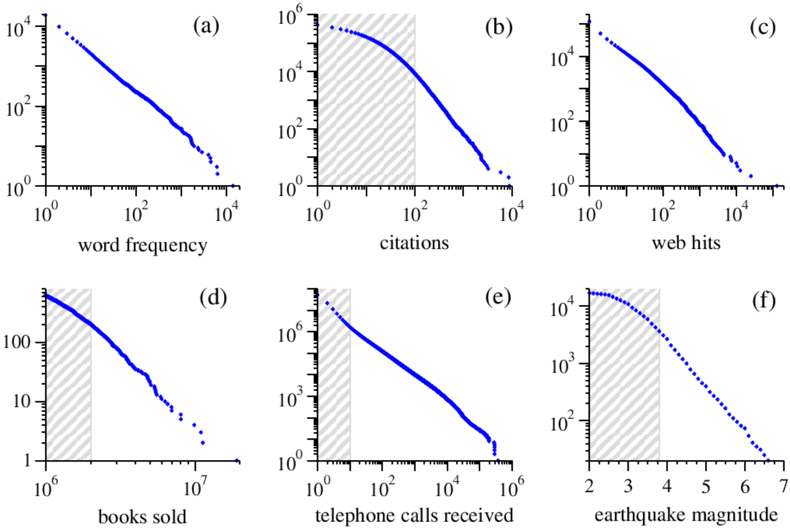


Partition of eight most important colors

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

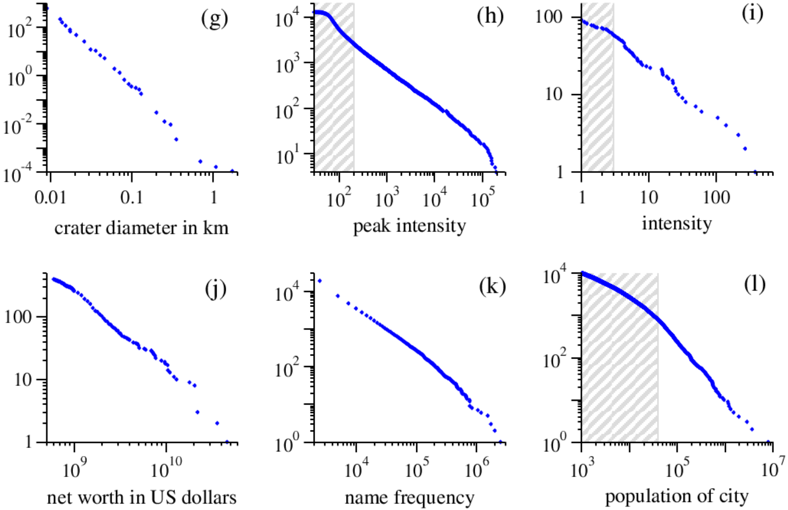


Power laws



(a)
word frequency
(b)
citations
(c)
web hits
(d)
books sold
(e)
telephone calls received
(f)
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

Power laws are **not** everywhere

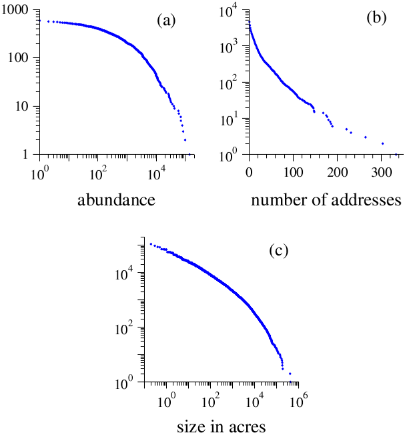


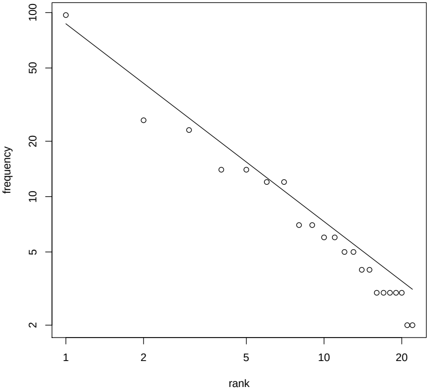
FIG. 5 Cumulative distributions of some quantities whose distributions span several orders of magnitude but that nonetheless do not follow power laws. (a) The number of sightings of 591 species of birds in the North American Breeding Bird Survey 2003. (b) The number of addresses in the email address books of 16 881 users of a large university computer system [33]. (c) The size in acres of all wildfires occurring on US federal land between 1986 and 1996 (National Fire Occurrence Database, USDA Forest Service and Department of the Interior). Note that the horizontal axis is logarithmic in frames (a) and (c) but linear in frame (b).

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

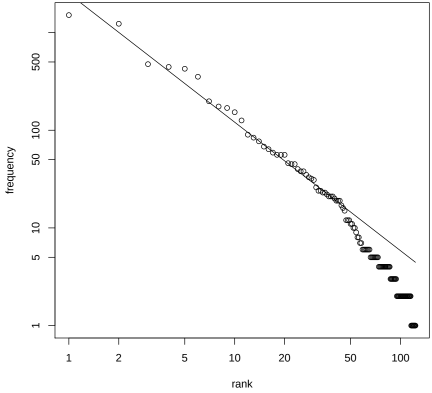
$$frequency \sim rank^{-1.06}$$



Other linguistic power law distributions

- size of language families
- source: Ethnologue

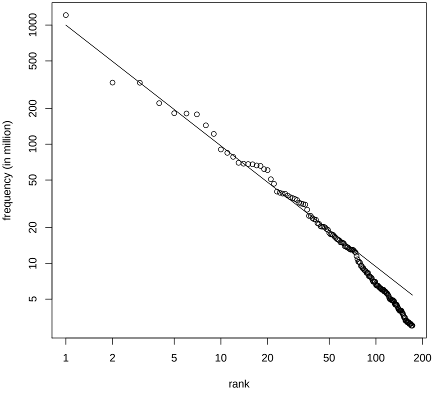
$$frequency \sim rank^{-1.32}$$



Other linguistic power law distributions

- number of speakers per language
- source: Ethnologue

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

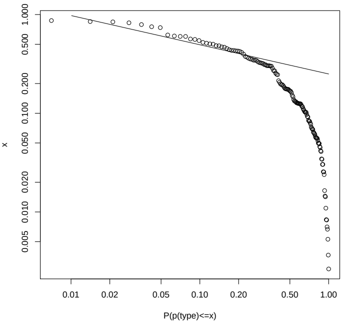


The World Atlas of Language Structures

- 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

- Maslova 2008, "Meta-typological distributions"
- hypothesis:
  - pick a random value for each feature
  - estimate the probability that a random language has this value
  - the likelihood that an arbitrarily chosen feature value has a probability x is proportional to a power of x
- only holds for the most frequent 30% of all types



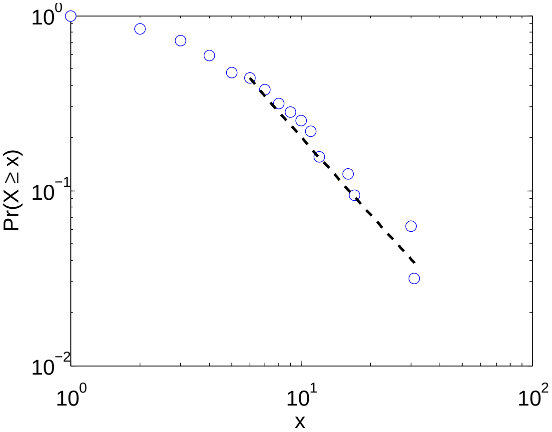
- for the entire range of type frequencies, the hypothesis can be rejected

The World Atlas of Language Structures

- however, Maslova is perhaps right in the assumption that languages are power-law distributed across WALS types
- worth to test it within features rather than across features
- 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

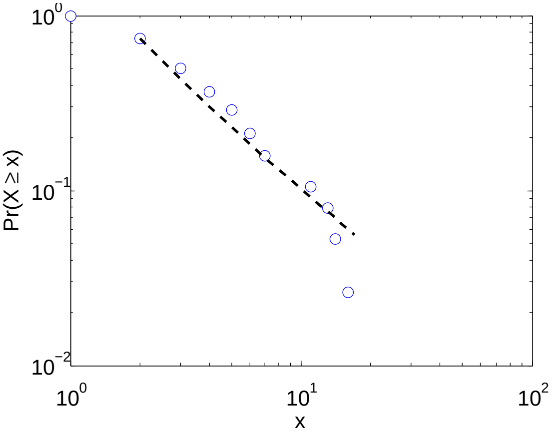
The World Atlas of Language Structures

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



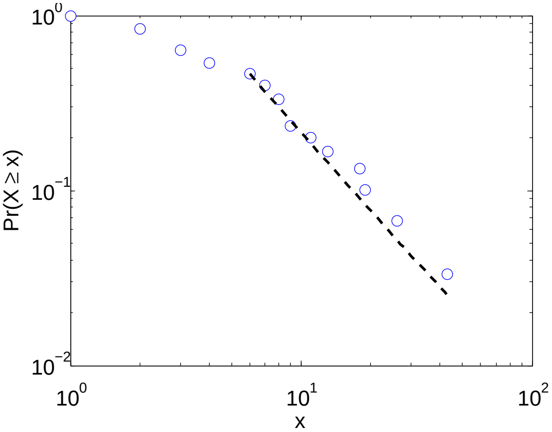
The World Atlas of Language Structures

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



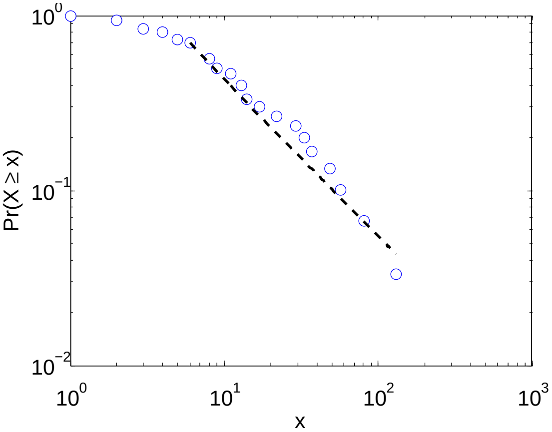
The World Atlas of Language Structures

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



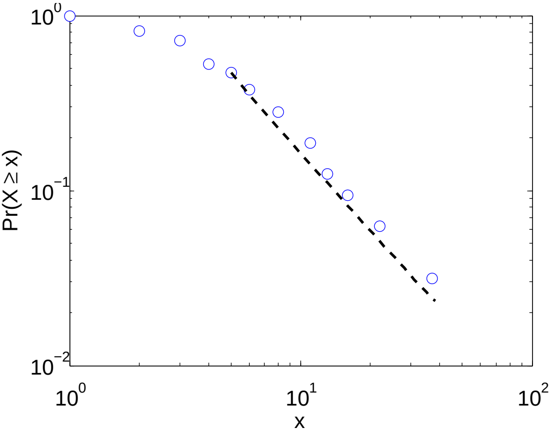
The World Atlas of Language Structures

- 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



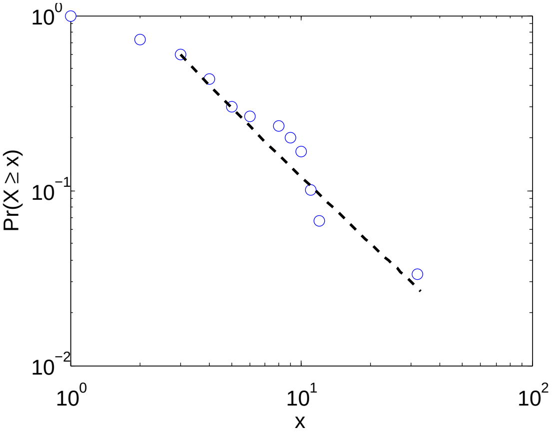
The World Atlas of Language Structures

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



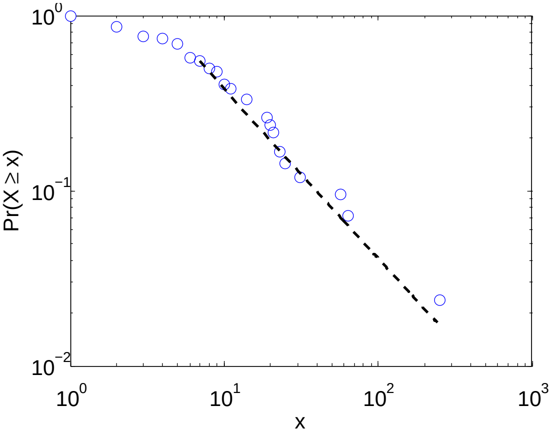
The World Atlas of Language Structures

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



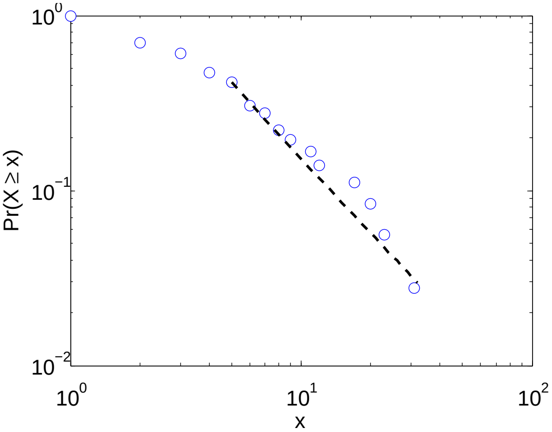
The World Atlas of Language Structures

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



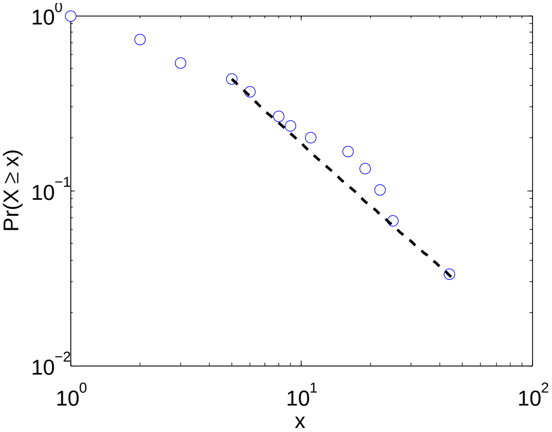
The World Atlas of Language Structures

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



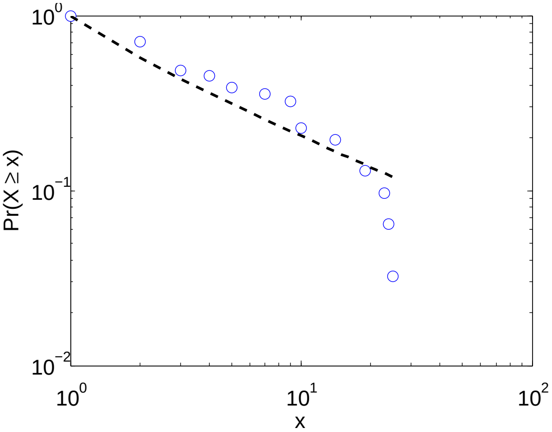
The World Atlas of Language Structures

- 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

- Power laws are characteristic of *critical states*
  - only small ice crystals in water above freezing point
  - one big ice crystal in water below freezing point
  - during transition from liquid to solid state:
    - ice crystals of many sizes
    - power-law distributed
- similar effect for all kinds of phase transitions in physics
- power laws are considered finger print of criticality

Why power laws?

Self-organized criticality

- some systems tend to return into a critical state due to their internal dynamics (see Bak et al. 1987)
- well-studied effect in computer simulations of cellular automata
- candidates for real-life examples are
  - earth quakes
  - forest fires
  - breakdowns of electricity networks
  - landscape formation
  - avalanches
  - ...

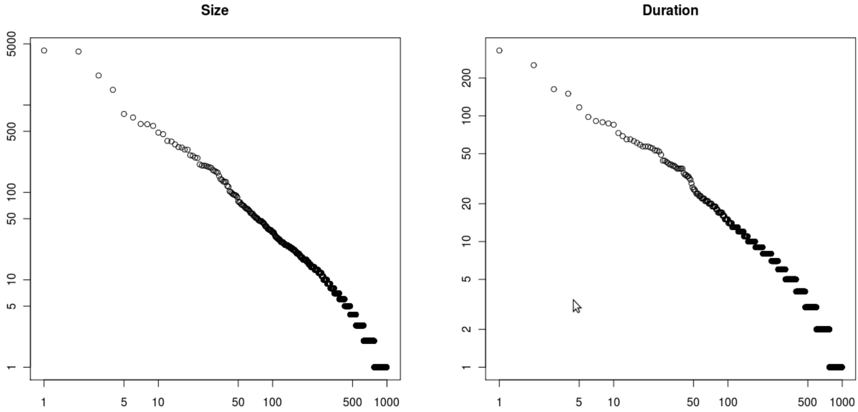
Why power laws?

The sandpile model

- cellular automaton; loosely inspired by real sand piles
- each cell has a certain value, its *slope*
- single grains are added at random, increasing the slope
- if the slope of a cell exceeds a critical value:
  - its slope is reduced by r
  - the slope of the four neighboring cells is increased by 1
- this may turn neighboring cells into the critical state, leading to further shifts
- see the computer simulation

The sandpile model

- both avalanche sizes and avalance durations are distributed according to a power law

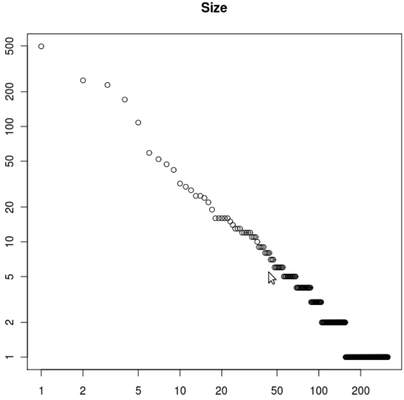


The forest fire model

- cellular automata model; inspired by behavior of wildfires
- each cell can be in either of three states: empty, tree, fire
- update rules:
    - fire $\rightarrow$ empty
    - tree and fire in neighoring cell $\rightarrow$ fire
    - with small probability: empty $\rightarrow$ tree
    - with even smaller probability: tree $\rightarrow$ fire
- simulation

The forest fire model

- size of contiguous clusters of trees or clusters of empty space are power law distributed



Simulating the evolution of color terms

Communication game

- game between a sender and a receiver
- two-dimensional conceptual space ($n \times n$ cells, periodic boundaries)
- small number of signals
- one round:
  - nature picks out a point in the concpetual space at random and shows it to the sender
  - the sender sends a signal to the receiver
  - the receiver has to guess which point the sender was referring to
  - both receive the same payoff:

$$\text{payoff} \sim \exp(-||p_s - p_r||^2)$$

  - if the distance between the sender's point and the receiver's guess is small, the payoff is high, and vice versa

Simulating the evolution of color terms

Evolutionary dynamics

- each player has a memory for point-signal associations
- after each round, the association between the signal and the point which were used in this round are strengthened proportional to the payoff of this round
- amounts to an evolutionary dynamics of associations:
  - succesful associations have a high fitness and are selected
  - unsuccesful associations have a low fitness and die out
- simulation

Simulating the evolution of color terms

Long-run dynamics

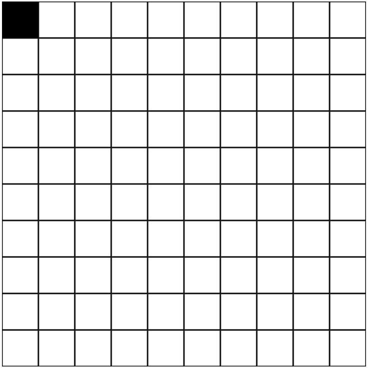
- players quickly evolve towards a local fitness maximum (*neutrally stable states*)
- induces a partition of the conceptual space into convex categories (each corresponding to one signal)
- most of the time evolution ends in one of the four global maxima (*evolutionarily stable states*)
- once a stable state has been reached, evolution comes to a standstill

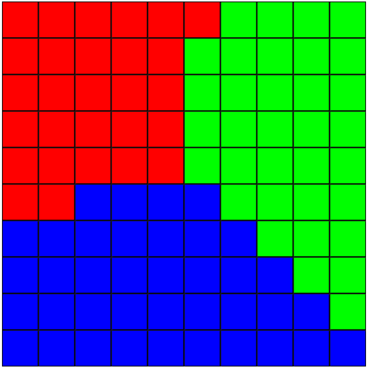
Spatial evolution

- 100 agents
- arranged on a 10×10 grid
- periodic boundaries
- in each round
  - a pair of neighbors is selected at random
  - they talk to each other and update their point-signal associations accordingly
- this is repeated thousands of times

Spatial evolution

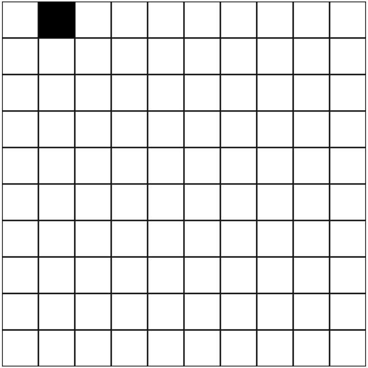
- population does not reach a stable homogenous state
- "languages" of neighbors are similar, but not identical

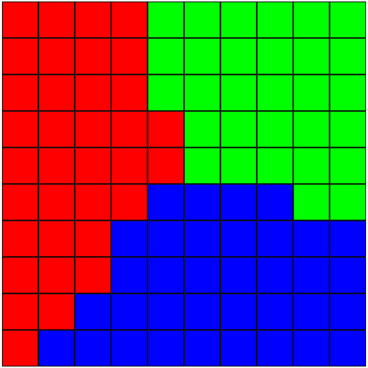




Spatial evolution

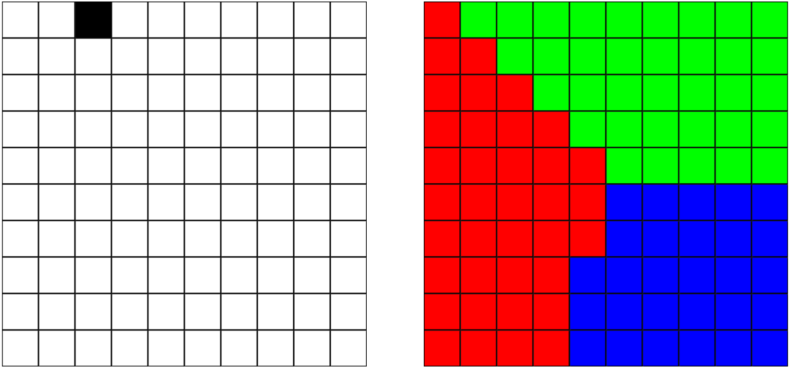
- population does not reach a stable homogenous state
- "languages" of neighbors are similar, but not identical





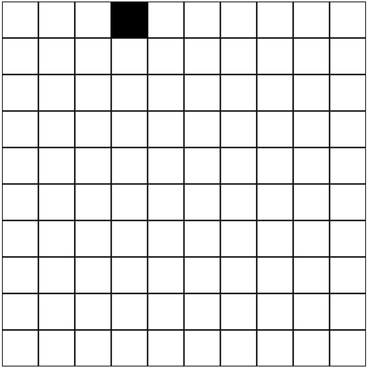
Spatial evolution

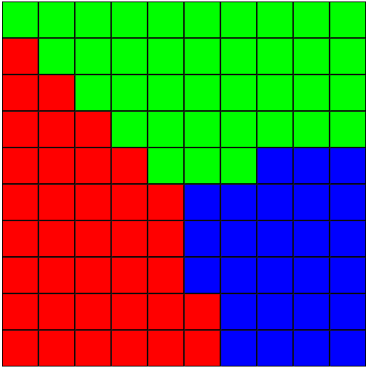
- population does not reach a stable homogenous state
- "languages" of neighbors are similar, but not identical



Spatial evolution

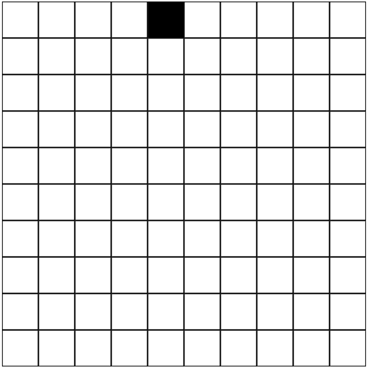
- population does not reach a stable homogenous state
- “languages” of neighbors are similar, but not identical

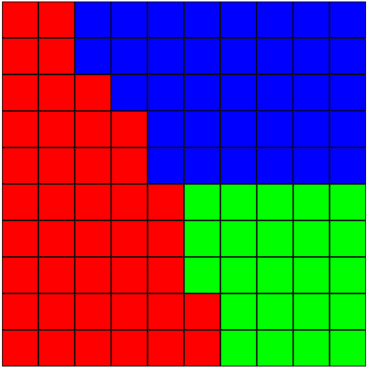




Spatial evolution

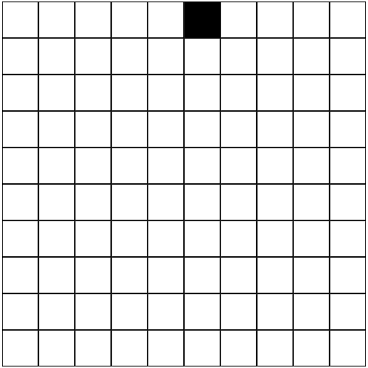
- population does not reach a stable homogenous state
- “languages” of neighbors are similar, but not identical

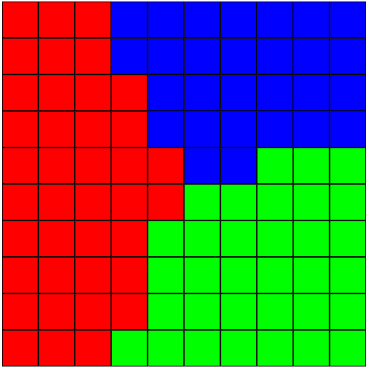




Spatial evolution

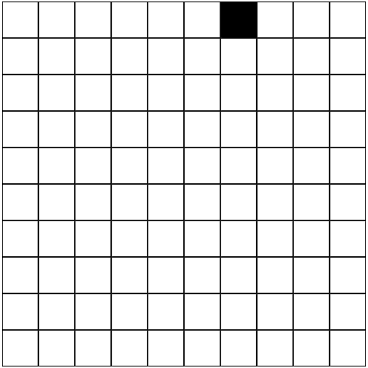
- population does not reach a stable homogenous state
- "languages" of neighbors are similar, but not identical

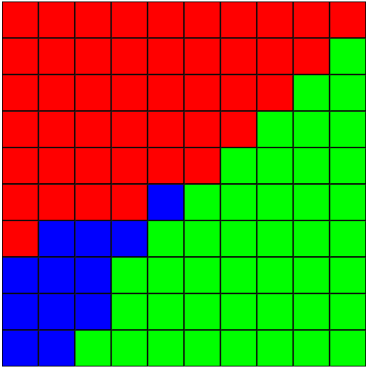




Spatial evolution

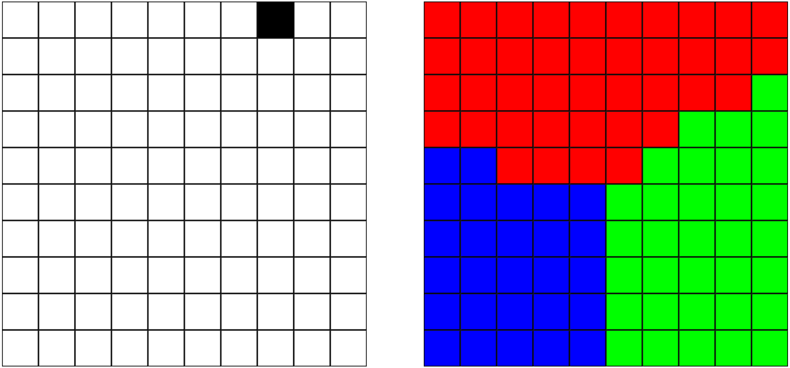
- population does not reach a stable homogenous state
- "languages" of neighbors are similar, but not identical





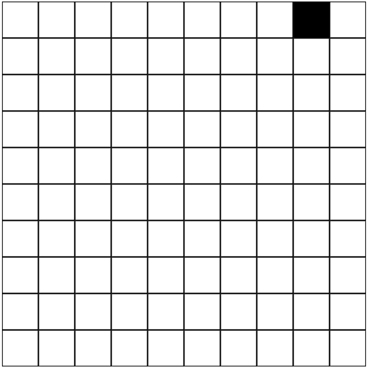
Spatial evolution

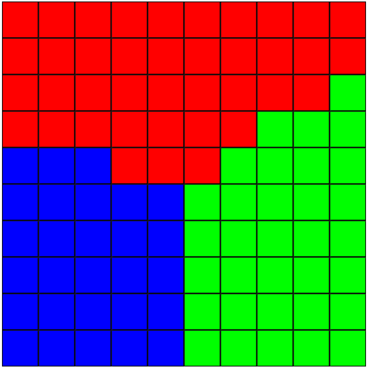
- population does not reach a stable homogenous state
- “languages” of neighbors are similar, but not identical



Spatial evolution

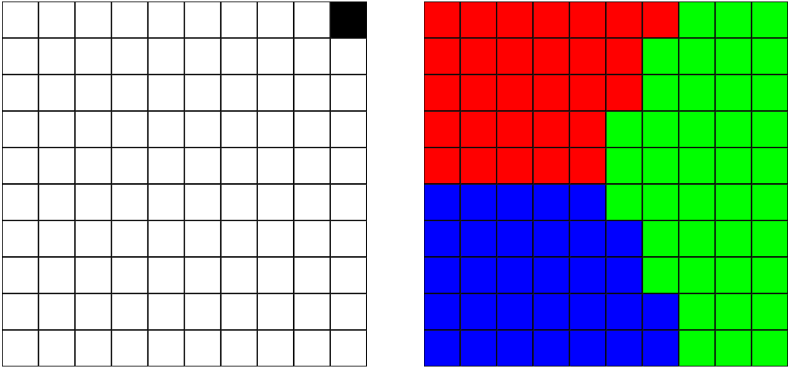
- population does not reach a stable homogenous state
- “languages” of neighbors are similar, but not identical





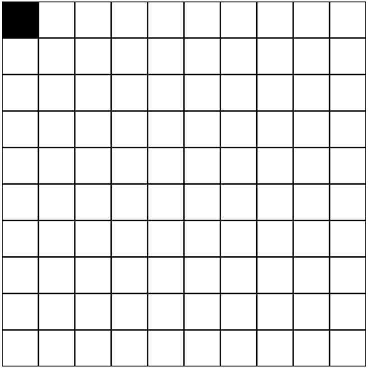
Spatial evolution

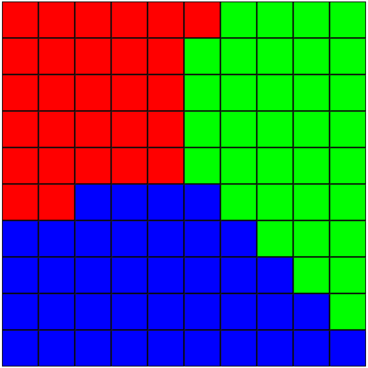
- population does not reach a stable homogenous state
- “languages” of neighbors are similar, but not identical



Spatial evolution

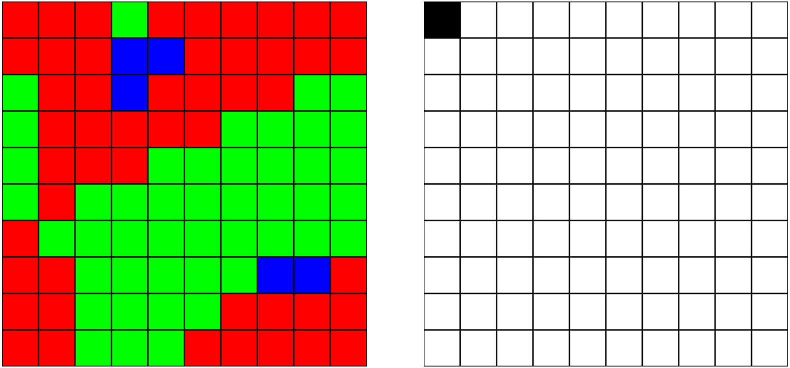
- population does not reach a stable homogenous state
- “languages” of neighbors are similar, but not identical





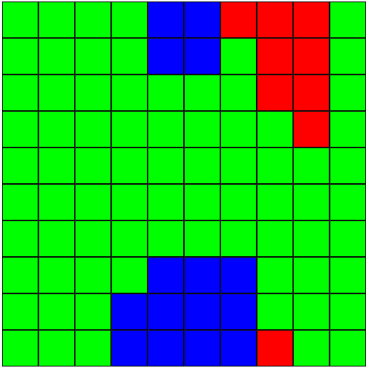
Spatial evolution

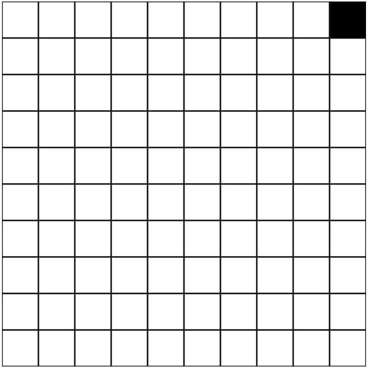
- no regions of completely identical languages
- however, clear “isoglosses” for single “concepts”



Spatial evolution

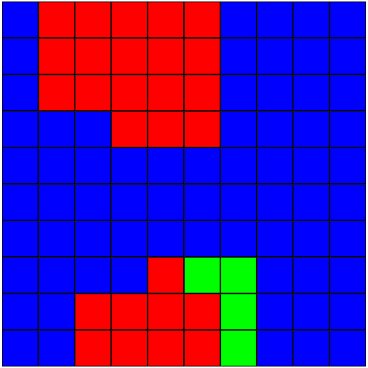
- no regions of completely identical languages
- however, clear “isoglosses” for single “concepts”

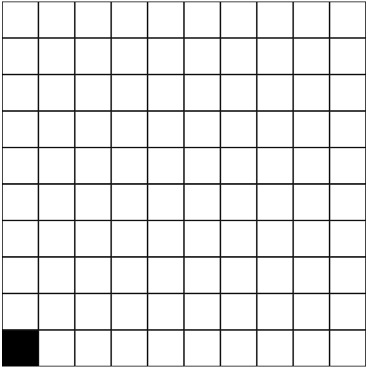




Spatial evolution

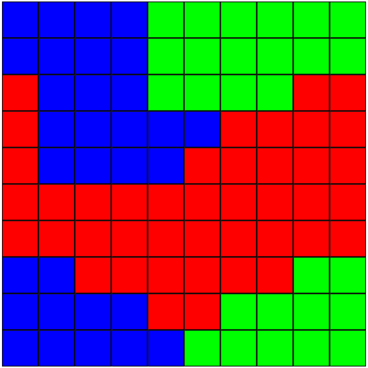
- no regions of completely identical languages
- however, clear “isoglosses” for single “concepts”

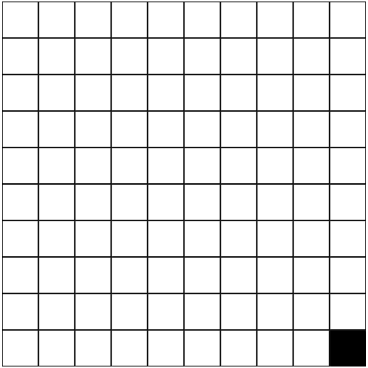




Spatial evolution

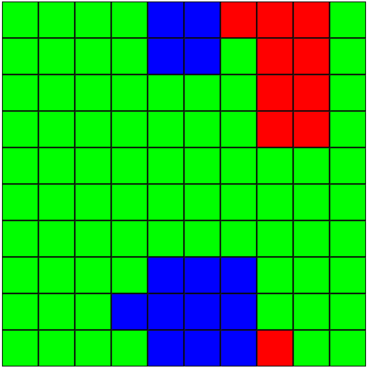
- no regions of completely identical languages
- however, clear “isoglosses” for single “concepts”

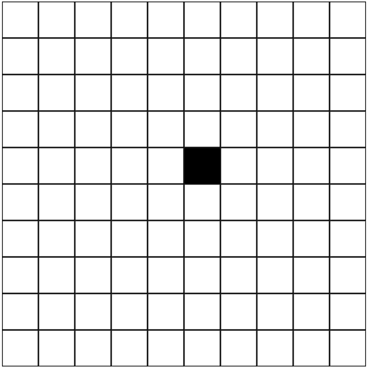




Spatial evolution

- no regions of completely identical languages
- however, clear "isoglosses" for single "concepts"

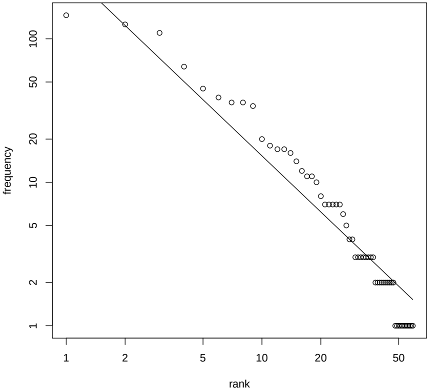




Spatial simulation produces power law

- Simulation data produce similar data structure like World Color Survey
- each artificial agent is treated as test person
- points in conceptual space $\approx$ Munsell chips
- signal with strongest association to that point $\approx$ categorization judgment
- same method of data evaluation:
  - Principal Component Analysis
  - dimensionality reduction
  - automatic classification of speakers into categorization types

Spatial simulation produces power law



$$frequency \sim rank^{-1.3}$$

Conclusion

- power laws are very common in cross linguistic variation
- indications that they are also characteristic of language change processes
- they are typical of self-organized criticality
- simulation of language evolution in a spatially structured population produces power law behavior (plus other characteristics that are observed in natural language)
- still inconclusive, but encouraging