Conceptual spaces and semantic similarities of colors

Gerhard Jäger
Universität Tübingen, Seminar für Sprachwissenschaft
Wilhelmstr. 19, 72074 Tübingen, Germany
gerhard.jaeger@uni-tuebingen.de

Abstract

The paper presents a statistical evaluation of the typological data about color naming systems across the languages of the world that have been obtained by the World Color Survey. In a first step, we discuss a singular value decomposition of the categorization data that led to a small set of easily interpretable features dominant in color categorization. These features were used for a dimensionality reduction of the categorization data.

Using the thus preprocessed categorization data, we proceed to show that the available typological data support the hypothesis by the cognitive scientist Peter Gärdenfors that the extension of color category are convex sets in the CIELab space in the languages of the world.

Keywords: color terms, conceptual spaces, semantic similarity, dimensionality reduction

1 Linguistic relativism and the semantics of color terms

The semantics of color terms, and the space of cross-linguistic variation regarding color semantics, has received a high amount of attention over the past decades. The domain of colors has a fairly simple and well-understood structure. Also, basic color terms are usually simple monomorphemic lexical entries. Still, relevant studies have unearthed interesting and non-trivial patterns of cross-linguistic variation. For these reasons, the semantics of color terms has been used as a paradigmatic case to address two fundamental issues: *linguistic nativism* and *linguistic relativity*.

The debate was initiated by the seminal study Berlin & Kay (1969). These authors investigated the color vocabulary of 98 typologically distinct languages (20 of these languages were studied in more detail). They found considerable variation in the size of basic color vocabularies (where *basic color terms* were defined as morphologically simple color terms with unrestricted applicability, excluding loan words), ranging from three to eleven. However, denotations of these terms were always drawn from a set of eleven universal categories (corresponding to the meanings of the English words *black, white, red, green, yellow, blue, brown, gray, pink, purple,* and *orange*). Furthermore they found that color term inventories follow *implicational universals* of the type: "If a language has a word for *yellow*, it also has a word for *red*."

These results seem to indicate that the range of possible cross-linguistic semantic variation is severely constrained, arguably by non-linguistic aspects of perception and cognition. This argument was reinforced by the study Heider (1972). It provided evidence that the ability

of test persons to recognize and remember colors does not depend on their native language's color vocabulary.

This sparked a controversial discussion lasting to the present day. While some authors have debated the existence of universal tendencies in color naming systems (e.g., Saunders & van Brakel 1997; Roberson et al. 2000), more recent results (such as Kay & Regier 2003; Lindsey & Brown 2006) have confirmed Berlin & Kay's basic conclusions while suggesting many modifications in detail. Also, some further experiments indicated that there is a correlation between the native languages of test persons and their ability to discriminate color stimuli (Witthoft et al. 2003). Gilbert et al. (2006) argue that this correlation exists but is restricted to the right visual field.

The present article studies the issue of the cross-linguistic universality of color naming systems under a somewhat different perspective. It will be argued that across languages, color categories are convex regions of a universal *conceptual space* (in the sense of Gärdenfors 2000) of low dimensionality. The statistical techniques utilized in this context are inspired by current work in computational distributional semantics (see for instance Erk 2012 for an overview), thus suggesting a potential synergy between computational and cross-linguistic semantics.

2 Semantic similarity and the geometry of meaning

It is a recurring thought in the history of natural language semantics that meanings are geometrically structured in some way. The *image schemas* developed in cognitive semantics (Langacker 1987; Lakoff 1987; Talmy 1988) are an early example. Peter Gärdenfors developed a full-blown program for semantic theory based on the idea that meanings are regions in some conceptual space which is endowed with a geometric structure (see for example especially Gärdenfors 2000; Gärdenfors 2014).

Quite independently of these developments in cognitive linguistics, research on semantics in quantitative natural language processing frequently employs geometric notions to represent semantic relations (see for instance Widdows 2004).

Utilizing geometric notions for semantic representations has several intuitively appealing aspects. The perhaps most attractive feature of this approach is the fact that the structure of a semantic space can be derived from the pairwise similarities between semantic objects. As similarity judgments can be empirically determined in various ways, such as via psychological experiments or corpus studies, the geometric approach offers an additional empirical foundation for semantic theory, complementing introspective judgments. Also, various meaning relations such as synonymy, hyperonymy, relevance etc. can be given a geometric interpretation. Furthermore, representing meanings geometrically allows to formulate semantic generalizations in geometrical terms not easily expressible in other frameworks. A case in point is Gärdenfors' (2000) *Criterion P*:

> "CRITERION P: A *natural property* is a convex region of a domain in a conceptual space."
>
> *Gärdenfors (2000), 71*

Together with Gärdenfors' assumption that simple adjectives denote natural properties, this leads to a semantic generalization that is not easily expressible in other semantic formalisms.

3 Human color perception

The semantics of color terms is well-suited to spell out a geometrical approach to meaning because there is little controversy that the the conceptual domain in question — colors as perceived by humans — does have a geometrical structure, but identifying this structure is non-trivial as it is not easily accessible to introspection.

Physiologically, human vision is based on photoreceptor cells in the retina sensitive to light of various wave lengths and intensities. There are two types of such cells: *cones* respond to bright light while *rods* work best in dim lights. Only cones are able to differentiate wave lengths, i.e. colors. There are three types of cones sensitive to different wave lengths. The three types are specialized roughly to monochromatic red, green and blue light respectively (see for instance Bowmaker & Dartnalli 1980 for details). Connected to this is the fact that human color vision is three-dimensional. All perceivable colors can be arranged in a three-dimensional structure in such a way that the Euclidean distance between two colors is inversely related to their perceived similarity. There are several three-dimensional color spaces in use, depending on the desired technical or psychometric application. In this study, I will use the CIELab color space, a three-dimensional representation of colors that has been designed with the goal of *perceptual uniformity*. This means that equal distances within the CIELab space approximately corresponds to equal perceptual dissimilarity. The CIELab *color solid* is roughly spherically shaped. White and black are located at the North Pole and the South Pole respectively. The rainbow color are arranged around the equator. Lighter colors, such as bright yellow or pink, are located in the Northern hemisphere and darker colors, such as brown or purple, in the Southern hemisphere. The various shades of gray are arranged along the earth's axis.

As the domain of colors has a well-defined geometrical structure, it is well-suited to study the predictions of the geometric theory of meaning. In this paper I will specificall be concerned with Gärdenfors' claim that simple adjectives denote convex regions.

4 The World Color Survey

In their above-mentioned path-breaking study Berlin & Kay (1969), the authors investigated the color naming systems of 98 typologically distinct languages. They argued that there are strong universal tendencies both regarding the extension and the prototypical examples for the meaning of the basic color terms in these languages.

As mentioned in the introduction, this work sparked a controversial discussion. To counter the methodological criticism that has been raised in this context, Kay and several co-workers started the *World Color Survey* project (WCS, see Cook et al. 2005 for details), a systematic large-scale collection of color categorization data from a sizeable amount of typologically distinct languages across the world.

The WCS researchers collected field research data for 110 unwritten languages, working with an average of 24 native speakers for each of these languages. During this investigation, the Munsell chips were used, a set of 330 chips of different colors that cover 322 colors of maximal saturation plus eight shades of gray. Figure 1 displays them in form of the Munsell chart.

The main chart is a 8×40 grid, with eight rows for different levels of lightness, and 40 columns for different hues. Additionally there is a ten-level column of achromatic colors,

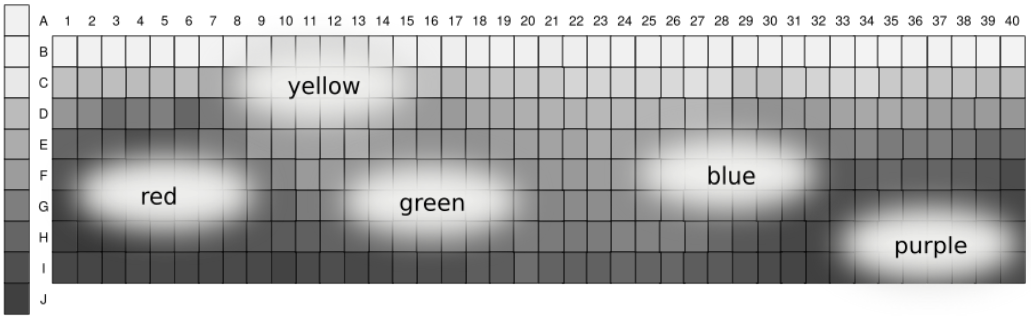

Figure 1: The Munsell chart

ranging from white via different shades of gray to black. The level of granularity is chosen such that the difference between two neighboring chips is minimally perceivable.

For the WCS, each test person was "asked (1) to name each of 330 Munsell chips, shown in a constant, random order, and (2), exposed to a palette of these chips and asked to to pick out the best example(s) ('foci') of the major terms elicited in the naming task" (quoted from the WCS homepage). The data from this survey are freely available from the WCS homepage http://www.icsi.berkeley.edu/wcs/data.html.

This invaluable source of empirical data has been used in a series of subsequent evaluations that largely confirm Berlin and Kay's hypothesis that there are universal tendencies in color naming systems (see for instance Kay & Maffi 1999; Kay & Regier 2003; Regier et al. 2005, Lindsey & Brown 2006).

In the present study I will look primarily at structural commonalities in how the test persons that participated in the WCS carve up the Munsell color space.

5 Cleaning up the data: dimensionality reduction

5.1 Motivation

For each informant, the outcome of the categorization task defines a partition of the Munsell space into disjoint sets — one for each color term from their idiolect.

An inspection of the raw data reveals a certain level of noise. This may be illustrated with the partitions of two speakers of a randomly chosen language (Central Tarahumara, an Uto-Aztecan language spoken in Mexico). They are visualized in Figure 2.

In the figure, shades of gray and the shape of the points in the center of each cell represent color terms of Central Tarahumara. We see striking similarities between the two speakers, but the identity is not complete. They have slightly different vocabularies, and the extensions of common terms are not identical. Furthermore, the boundaries of the extensions are unsharp and appear to be somewhat arbitrary at various places. Also, some data points seem to be due to plain mistakes. Similar observations apply to the data from most informants.

Since the raw data from the WCS are quite noisy, in a first step I applied heuristic methods to separate the linguistically/cognitively interesting variation in the data from noise. Techniques from distributional semantics proved to be suitable for this purpose.

Recall that the first task of the test persons in the World Color Survey was to name the 330 Munsell chips, using color terms from their native languages. The outcome of this task

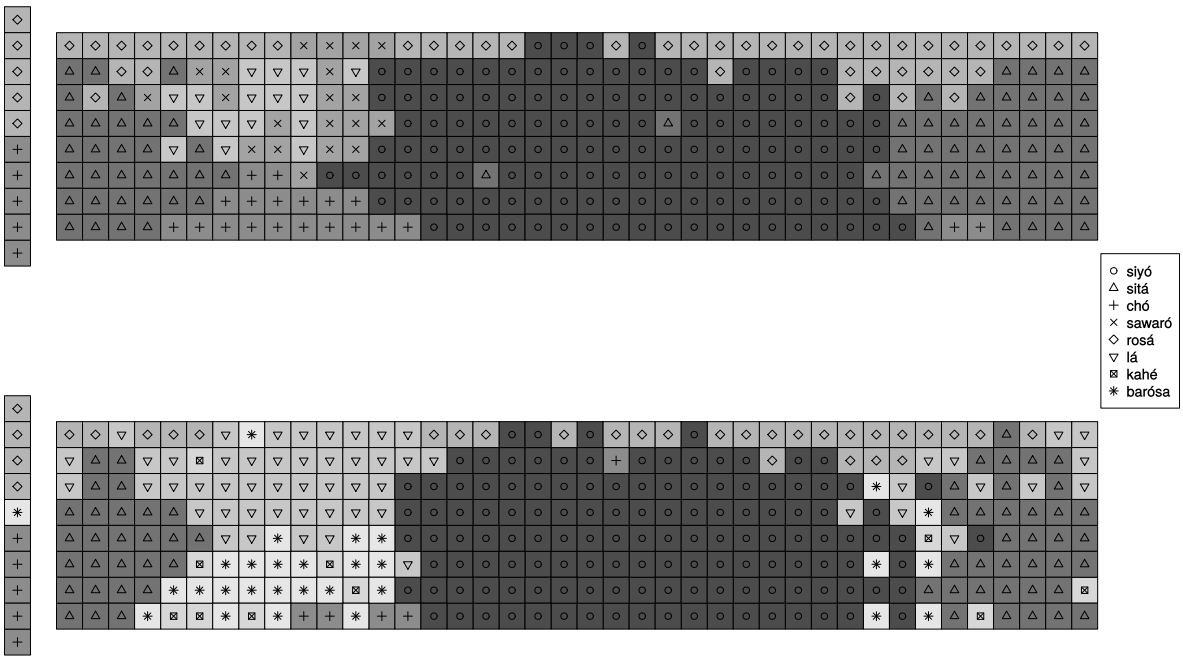

Figure 2: Partitions for two speakers of Central Tarahumara

can be represented as a *contingency table*. The rows of this matrix are word-language pairs, such as *red/English* or *noir/French*. (English and French are not covered in the WCS data, but they are used for illustrative purposes here.) There are 1,601 rows in total.

The columns of the contingency table are the 330 Munsell chips. Each cell contains the number of test persons that used the term corresponding to the row to name the Munsell chip corresponing to the color. The structure of this table is illustrated in Table 1. (The row names and the numbers are made up for the purpose of illustration.)

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

Table 1: Contingency table

This matrix is comparable to the *term-document matrix* (TDM) in distributional semantics (see for instance Widdows 2004: 145). A TDM is derived from a corpus which consists of a selection of documents. It is a contingency matrix with terms as rows, documents as columns and the frequency of occurrence of a term in a document as entries.

The WCS contingency matrix can be seen as a collection of vectors in a 330-dimensional space. The 330 Munsell chips define the dimensions of this abstract space, and the usage patterns of an individual term (of an individual languages) in the WCS gives a vector in this 330-dimensional space.

In this representation, each vector (usage pattern of a term) has 330 degrees of freedom. It is for instance possible to represent a checkerboard pattern on the Munsell chart as such vectors. However, we know that the color categorization patterns of human test persons are much more constrained — a checkerboard pattern is not a possible response of a WCS test person. Let us assume that there are actually only k degrees of freedom in human color categorization, with $k \ll 330$. We want to project the raw data vectors from the contingency table to a k-dimensional space while preserving as much information as possible.

5.2 Singular Value Decomposition in a nutshell

A commonly used method for this purpose (which also underlies the *Latent Semantic Analysis* technique in distributional semantics; see Landauer & Dumais 1997) is based on *Singular Value Decomposition* (SVD). Let me illustrate the underlying intuition with a toy example. Remember the fairy tale of Hansel and Gretel. When the father leads the children into the wood for the first time, Hansel leaves a trail of pebbles along the way, which leads them back to their home after the parents abandoned them. Suppose they walked on a straight line, and the pebbles are located as in the left panel of Figure 3.

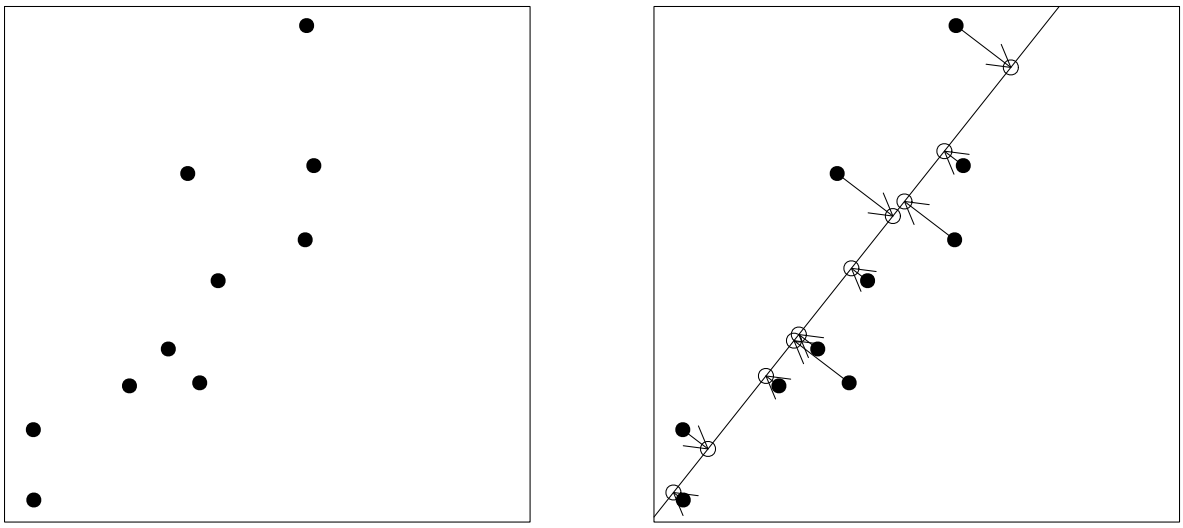

Figure 3: Dimensionality reduction from two dimensions to one

Hansel did not manage to leave the pebbles exactly where they walked; their location in two dimensions is perturbed by a certain amount of noise. To find back home, Hansel and Gretel are interested to reconstruct the one-dimensional linear manifold that generated the pattern — the path they walked in the morning — and to filter out the uninformative noise.

SVD achieves this by finding a lower-dimensional linear manifold — a straight line — with the property that its average distance to the observed points is minimized. This is shown in the right panel of Figure 3. The white circles represent the projections of the observed data points onto the inferred lower-dimensional manifold.

While the toy example reduces dimensionality from two to one, the WCS data are points in a 330-dimensional space that should be projected onto a k-dimensional linear manifold, for an unkown value of k.

In the remainder of this subsection I will spell out the underlying mathematics of SVD, as well as the postprocessing step applied, in somewhat more detail. This part can be skipped without loss of continuity.

There is a theorem of linear algebra which states that each matrix $m \times n$ M can be decomposed in the following way (see for instance Strang 2009, Section 6.7):

$$M = U\Sigma V^T \tag{1}$$

Here U is a $m \times m$ orthogonal matrix, i.e. a rotation of points in m-dimensional space. Likewise, V is an $n \times n$ orthogonal matrix, i.e. a rotation of n-dimensional space. Σ is a $m \times n$ diagonal matrix, i.e. a matrix where all entries except those on the main diagonal are zero.

The diagonal entries σ_i in Σ are the *singular values* of M. By convention, they are listed in descending order.

As explained above, the intuition underlying SVD is that there is an abstract space of *latent dimensions*, and the variation of the data (represented by M) along the various latent dimensions is pairwise independent. The matrix U maps data points in an m-dimensional space (i.e. columns of M) onto the latent space. Likewise, U maps n-dimensional vectors (rows of M) onto the latent space. The singular value σ_i captures the importance of the i-th latent dimension in distinguishing the data points in M.

When applied to the WCS contingency matrix, m is the number of term/language pairs (1,601) and n is the number of Munsell chips (330). The latent dimensions are, ideally, the criteria test persons use when categorizing Munsell chips. Additionally, the WCS data have degrees of freedem (i.e. latent dimensions) that reflect properties of the data collection process rather than cognitive features of color categorization. Let us assume, optimistically, that most variation in the WCS data reflect cognitively and linguistically meaningful information rather than noise due to data collection. Then there is a number k, with $k \ll 330$, such that the space of the k most important latent dimensions captures the interesting variation in the data, while the remaining $330 - k$ latent dimensions reflect noise. Figure 4 displays the singular values of the WCS contingency table.

There is no fool-proof criterion to determine the number k of "interesting dimensions". Figure 4 suggests that at most the first 20 latent dimension capture the relevant variation in the data. In the sequel, I will — somewhat arbitrarly — choose $k = 10$. None of the results of this study depends on this choice though.

Let Σ_k be the result of replacing σ_i in Σ by zero for each $i > k$. If we replace Σ in equation (1) by Σ_k, we get

$$M_k = U\Sigma_k V^T \tag{2}$$

M_k is a matrix with the same shape as M. However, its rank is reduced to k, i.e. both its rows and its columns are vectors within a k-dimensional linear manifold. In fact, M_k is the rank-k matrix which is as close to M as possible.

Each of the 10 latent dimensions identified via SVD can be mapped to a vector of the 330-dimensional space of Munsell chips. These vectors are the first 10 columns of the vector V of equation (1). In Figure 5 the 3rd and 5th of these vectors are visualized for illustration. The values of a 330-dimensional vector along a dimension is represented by the shade of gray

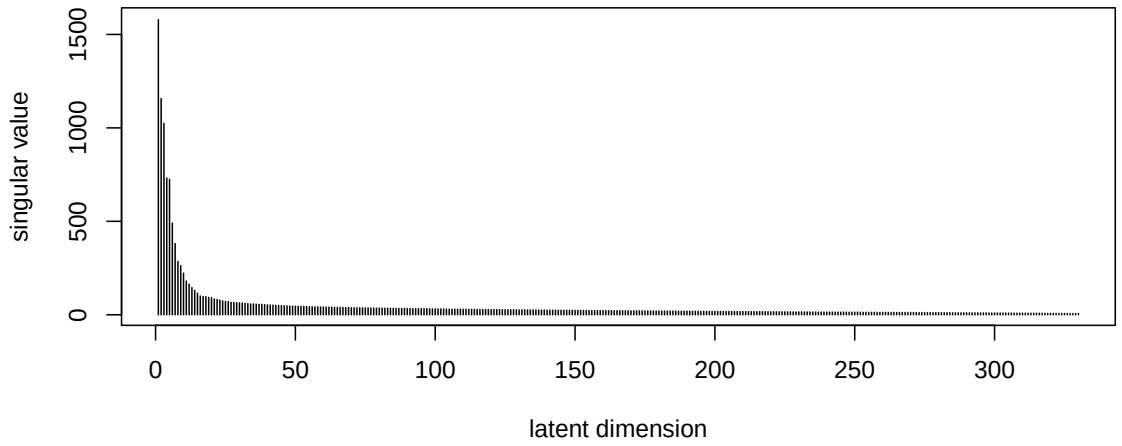

Figure 4: Singular values of the WCS contingency table

of the corresponding position in the Munsell chart. White corresponds to the minimal and black to the maximal value along any dimension.

As can be seen from these charts, the third latent dimension has its maximal values at red Munsell chips and its minimal values at white ones. So in short, this latent feature captures the contrast between white and red. Likewise, the 5th latent dimension captures the contrast between red and yellow. Similar — or more complex — patterns can also be discerned in the other eight latent dimensions.

The *Varimax* algorithm (Kaiser 1958) is a statistical technique to facilitate the interpretability of latent features. It amounts to a rotation of the latent (in our cases: 10-dimensional) subspace in such a way that the correlation of the latent with the observable dimensions is maximized. This procedure was applied to the 10-dimensional space resulting from SVD.

5.3 Results

The resulting 10 relevant dimensions, or *features*, are visualized in Figure 6. For each of these features, the high values are concentrated within a contiguous region of the Munsell chart. In most cases, these regions even correspond to the extension of English basic color terms: *green, red, white, black, yellow, blue, purple, pink* and *brown*. Only the 10th latent dimension has to be described with a non-basic term, *light blue*.

These findings indicate that the WCS test persons categorized Munsell chips mostly according to their degree of greenness, redness, whiteness etc. So the WCS data provide support to Berlin and Kay's claim that there are universal color categories that underlie the structure of the color vocabulary in typologically diverse languages.[1]

[1]The WCS data do not fully support Berlin and Kay's more specific proposals about universals of color vocabularies. These issues are discussed in detail in Jäger (2012), using a methods similar to the ones described here.

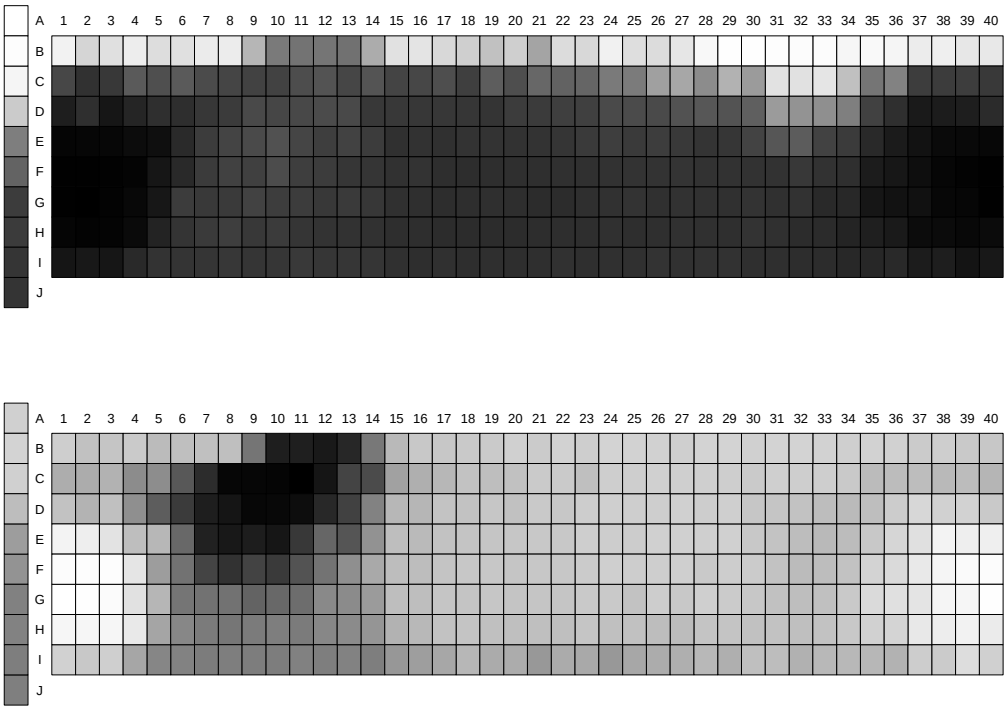

Figure 5: Third and fifth latent dimension

The effect of mapping the raw data vectors to the 10 most important latent dimensions is illustrated in Figure 7. The first panel visualizes the row corresponding to the term *aleluya* from the Bolivian language Chiquitano. This term was only used by a single test person. That person used it for some, but not all, deep red Munsell chips. The vector that results from applying dimensionality reduction is shown in the second panel. It has high values for all red Munsell chips.

Another example is displayed in the third and fourth panel, representing the term *madsmas* of the Bantoid language Gunu (spoken in Cameroon). This term was used by two test person. One of them only used it to name two light gray Munsell chips, while the other one also applied it to white and several whitish, very light blue and very light pink chips. After dimensionality reduction (bottom panel), the corresponding vector has high values for white, all light shades of gray and very light shades of blue and pink.

Dimensionality reduction can also be applied to the extension of a term as it is used by a single test person. The set of chips which were categorized by the same term by a given test person can be represented as a 330-dimensional binary vector. Let V_k be the matrix consisting of the first k columns of V. V_k can be conceived as a transformation that maps 330-dimensional vectors (over Munsell chips) to k-dimensional vectors in the latent subspace. The matrix V_k^T represents a transformation that maps vectors from the k-dimensional latent space to an k-dimensional linear sub-manifold of the 330-dimensional Munsell space. Combining these two operations to the matrix $V_k V_k^T$ is an operation within the 330-dimensional Munsell space that supresses all but the first k latent dimensions. (The rows of M_k are the result of applying

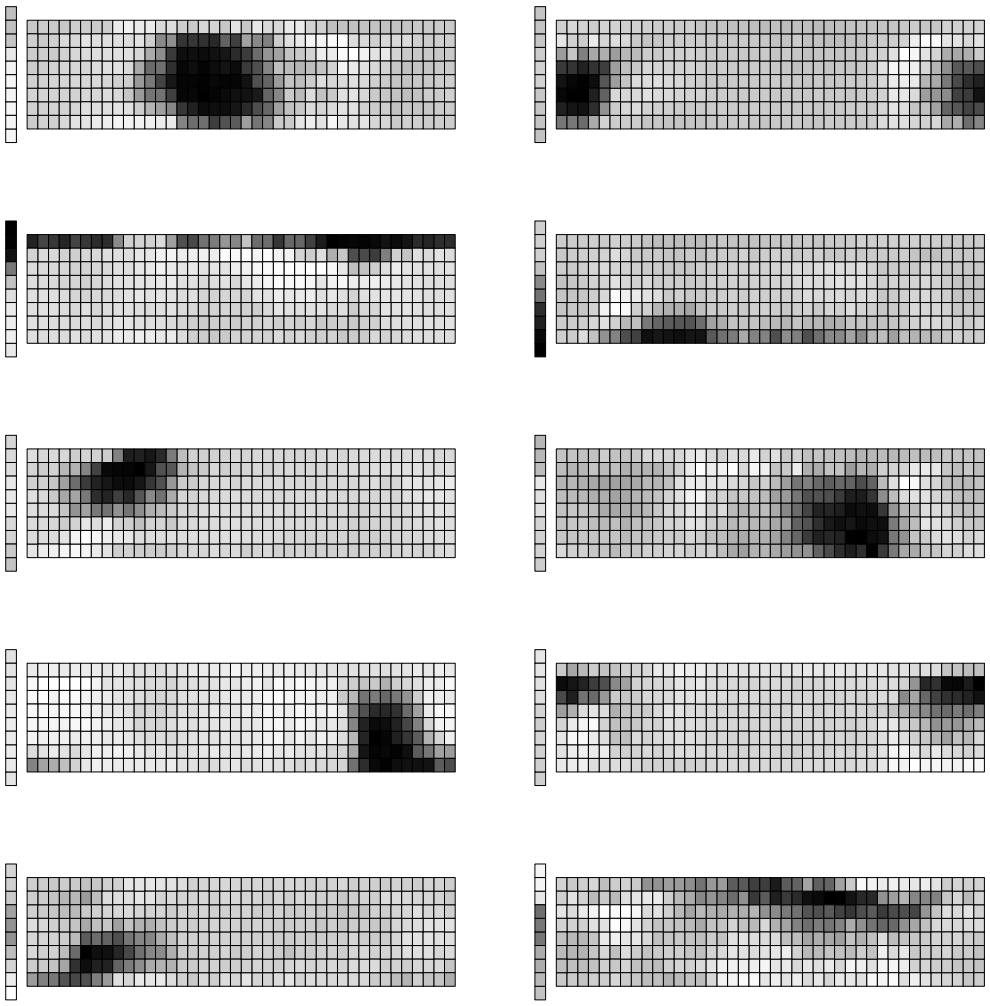

Figure 6: Latent dimensions (Singular Value Decomposition + Varimax)

this operations to the corresponding rows or M.)

To illustrate this, let us consider again the data from the first test person from the Mexican language Central Tarahumara (cf. the upper panel of Figure 2). The upper left panel of Figure 8 displays the extension of the term *siyó*. It covers almost all shades of green and blue, except two single chips in the center of green and blue respectively. The upper right panel shows the extension of the term *sitá*, which covers all shades of red plus the two green and blue outliers. The lower two panels visualize the result of applying dimensionality reduction to these two extensions (or rather, their vector representation). The dimensionality-reduced version of *siyó* now covers all shades of blue and green, while *sitá* is confined to all shades of red.

After dimensionality reduction, the values of the extension of a term for a single speaker is not binary anymore but contains values. It can be conceived as a fuzzy set with smooth boundaries.

For a given speaker, each Munsell chip c can be assigned to the term t for which c has the highest membership value after dimensionality reduction. This leads to a partition of

the Munsell space into disjoint categories. The net effect of this transformation is that the boundaries between categories are crisp but outliers in the raw data are removed. Figure 9 displays the raw data for the first test person from Central Tarahumara (upper panel) and the partition that results from dimensionality reduction (lower panel). Figure 10 displays the raw and the denoised partition for test person 11 from the Nilo-Saharan language Didinga (spoken in South Sudan). The comparison between the upper and the lower chart in both figures illustrates how dimensionality reduction smoothes the boundaries between categories and removes outliers.

6 Convexity in the CIELab space

6.1 Motivation

The visualizations that were discussed so far suggest the generalization that after dimensionality reduction, category extensions are usually contiguous regions in the 2d Munsell space. This impression becomes even more striking if we study the extensions of categories in a geometrical representation of the color space with a psychologically meaningful distance metric. The CIELab space has this property. It is a 3d space with the dimension L* (for lightness), a* (the green-red axis) and b* (the yellow-blue axis). The set of perceivable colors forms a three-dimensional solid with approximately spherical shape. As described above, white is at the North Pole, black at the South Pole, the rainbow colors form the equator, and the gray axis cuts through the center of the sphere. The CIELab space has been standardized by the *Commission Internationale d'Eclairage* such that Euclidean distances between pairs of colors are monotonically related to their perceived dissimilarity.

The 320 chromatic Munsell colors cover the surface of the color solid, while the ten achromatic chips are located at the vertical axis. The color solid is actually not completely spherical but irregularly shaped.

Visually inspecting CIELab representations of the (dimensionality-reduced) partitions led to the hypothesis that the boundaries between categories are in most cases linear, i.e. two-dimensional planes. This is in line with the main claim of Gärdenfors' (2000) book *Conceptual Spaces*. Gärdenfors suggests that meanings can always be represented geometrically, and that *natural categories* must be convex regions in such a conceptual space. The three-dimensional color space is one of his key examples.

In this section it will be tested to what degree this prediction is borne out for the dimensionality-reduced partitions obtained from the WCS. For this purpose, a computational method was devised that modifies a partition (by re-categorizing individual Munsell chips) in such a way that each category corresponds to a convex sub-region of the CIELab space and these sub-regions do not overlap. The *degree of convexity* of a partition is the proportion of Munsell chips that need to be re-categorized for convexification.

6.2 Finding convex approximations using Support Vector Machines

The convexification algorithm is described in more detail in this section. This material is not essential for the main thread of this article.

The algorithm can be described as follows. Suppose a partition $p_1, \cdots, p_k$ of the Munsell colors into k categories is given.

1. For each pair of distinct categories p_i, p_j (with $1 \leq i, j \leq k$), find a linear separator in the CIELab space that optimally separates p_i from p_j. This means that the set of Munsell chips is partitioned into two linearly separable sets $\tilde{p}_{i/j}$ and $\tilde{p}_{j/i}$, that are linearly separable, such that the number of items in $p_i \cap \tilde{p}_{j/i}$ and in $p_j \cap \tilde{p}_{i/j}$ is minimized.

2. For each category p_i, define
$$\tilde{p}_i \doteq \bigcap_{j \neq i} \tilde{p}_{i/j}$$

As every $\tilde{p}_{i/j}$ is a half-space and thus convex, and the property of convexity is preserved under set intersection, each $\tilde{p}_i$ is a convex set.

To perform the linear separation in the first step, I used a soft-margin Support Vector Machine (SVM). An SVM (Vapnik & Chervonenkis 1974) is an algorithm that finds a linear separator between two sets of labeled vectors in an n-dimensional space. An SVM is soft-margin if it tolerates misclassifications in the training data. As SVMs are designed to optimize generalization performance rather than misclassification of training data, it is not guaranteed that the linear separators that are found in step 1 are really optimal in the described sense. Therefore the numerical results to be reported below provide only a lower bound for the degree of success of Gärdenfors' prediction.

The output of this algorithm is a re-classification of the Munsell chips into convex sets (that need not be exhaustive). Figure 11 illustrates the convex approximation of a partition with the data of test person 12 from the Nilo-Saharan language Murle (spoken in South Sudan and Ethiopia). The upper panel shows the categorization performed by this test person after dimensionality reduction. The lower panel gives the convex approximation of this partition using the procedure described above. It should be kept in mind that these charts are two-dimensional projections of the three-dimensional CIELab sapce. The region covering purple and pink, for instance, is not convex in the projection. Nevertheless it does represent a convex region in CIELab space.

It can be seen that there are minor differences between the two partitions concerning the boundaries between adjacent categories. The five white squares in the lower chart mark Munsell chips that could not uniquely be assigned to any convex category. In total the two partitions agree for 306 out of 330 Munsell chips.

The *degree of convexity* "conv" of a partition is defined as the proportion of Munsell chips that are not re-classified in this process. If $p(c)$ and $\tilde{p}(c)$ are the class indices of chip c before and after re-classification, and if $\tilde{p}(c) = 0$ if $c \notin \bigcup_{1 \leq i \leq n} \tilde{p}_i$, we can define formally:

$$\text{conv} \doteq |\{c | p(c) = \tilde{p}(c)\}| / 330$$

For the example from Figure 11, this value is $306/330 \approx 92.6\%$.

6.3 Results

The mean degree of convexity of the partitions that were obtained SVD and dimensionality reduction is 93.8%, and the median is 94.5% (see the first boxplot in Figure 12).

One might wonder how important the dimensionality reduction step is in obtaining this result. To address this question, convex approximation via SVM training was also applied to the raw partitions. Here the degree of convexity is only 77.9% (see the second boxplot in Figure 12).

Since the difference between these values is considerable, one might suspect that the high degree of convexity for the cleaned-up data is actually an artifact of the dimensionality reduction algorithm and not a genuine property of the data. This is not very plausible, however, because the input for dimensionality reduction were exclusively categorization data from the WCS, while the degree of convexity depends on information about the CIELab space. Nevertheless, to test this hypothesis, a random permutation of the category labels was applied to each original partition. To the permuted data, the same analysis (SVD, dimensionality reduction, computation of the degree of convexity) was performed. The mean degree of convexity for these data is as low as 35.1% (see the third boxplot in Figure 12). The fact that this value is so low indicates that the high average degree of convexity is a genuine property of natural color category systems and not a side effect of dimensionality reduction.

The choice of $k = 10$ as the number of relevant latent dimensions was somewhat arbitrary. Therefore it is important to test to what degree the results from this section depend on this choice.

For this reason, the same analysis was performed with the original data for all values of k between 5 and 20. The dependency of the mean degree of convexity on k is displayed in Figure 13. It can be seen that the degree of convexity is not very sensitive to the choice of k. The highest value is achieved for $k = 6$ with mean degree of convexity of 94.3%. For higher values of k convexity slightly drops off, but at a slow rate. For $k = 20$, mean degree of convexity is still at 92.8%. Most importantly, the mean degree of convexity only differs very slightly (by less than 0.5%) between adjacent values of k. We can thus conclude that the choice of k does not seriously affect the qualitative conclusions of this study.

These results strongly indicate that color terms across typologically diverse languages denote convex regions within the CIELab space. This finding strongly supports Gärdenfors' hypothesis that simple adjectives denote natural properties, which are, in turn, convex regions within some conceptual space.

7 Acknowledgments

I would like to thank the editors and reviewers of this volume for valuable feedback. This research has been supported by the ERC Advanced Grant 324246 EVOLAEMP (*Language Evolution: The Empirical Turn*) and the DFG-KFG 2237 *Words, Bones, Genes, Tools*, which is gratefully acknowledged.

A Methods

Investigation where confined to the data from those informants for which the WCS contains a definite category term for each of the 330 Munsell chips. In total, 13,490 category extensions/vectors from 1,771 speakers from 102 languages (out of a total of 21,992 categories for 2,616 speakers from 110 languages) have been used.

The convex approximations where computed by using the e1071 implementation (Dimitriadou et al. 2005) of an SVM with a penalty term of 100 to minimize the number of misclassified training data rather than to maximize the margin.

References

Berlin, Brent, and Paul Kay. 1969. *Basic color terms: their universality and evolution*. University of California Press: Berkeley and Los Angeles.

Bowmaker, James K., and Herbert J. A. Dartnall. 1980. 'Visual pigments of rods and cones in a human retina'. *The Journal of Physiology* 298, 501–511.

Cook, Richard, Paul Kay, and Terry Regier. 2005. 'The world color survey database: History and use'. In: Cohen, Henri, and Claire Lefebvre (eds.), *Handbook of Categorisation in the Cognitive Sciences*. Amsterdam: Elsevier, 223–242.

Dimitriadou, Evgenia, Kurt Hornik, Friedrich Leisch, David Meyer, and Andreas Weingessel. 2005. *e1071: Misc Functions of the Department of Statistics (e1071)*. Tech. rep. Vienna: Technical University Vienna.

Erk, Katrin. 2012. 'Vector space models of word meaning and phrase meaning: A survey'. *Language and Linguistics Compass* 6, 635–653.

Gärdenfors, Peter. 2000. *Conceptual Spaces*. MIT Press: Cambridge, MA.

Gärdenfors, Peter. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press: Cambridge, MA.

Gilbert, Aubrey L., Terry Regier, Paul Kay, and Richard B. Ivry. 2006. 'Whorf hypothesis is supported in the right visual field but not the left'. *Proceedings of the National Academy of Sciences of the United States of America* 103, 489–494.

Heider, Eleanor Rosch. 1972. 'Universals in color naming and memory'. *Journal of Experimental Psychology* 93, 10–20.

Jäger, Gerhard. 2012. 'Using statistics for cross-linguistic semantics: a quantitative investigation of the typology of color naming systems'. *Journal of Semantics* 29, 521–544.

Kaiser, Henry F. 1958. 'The varimax criterion for analytic rotation in factor analysis'. *Psychometrika* 23, 187–200.

Kay, Paul, and Luisa Maffi. 1999. 'Color appearance and the emergence and evolution of basic color lexicons'. *American Anthropologist*, 743–760.

Kay, Paul, and Terry Regier. 2003. 'Resolving the question of color naming universals'. *Proceedings of the National Academy of Sciences* 100, 9085–9089.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press: Chicago.

Landauer, Thomas K., and Susan T. Dumais. 1997. 'A Solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge'. *Psychological Review* 104, 211–240.

Langacker, Ronald Wayne. 1987. *Foundations of Cognitive Grammar*. Vol. 1. Stanford University Press: Stanford.

Lindsey, Delwin T., and Angela M. Brown. 2006. 'Universality of color names'. *Proceedings of the National Academy of Sciences* 103, 16608–16613.

Regier, Terry, Paul Kay, and Richard S. Cook. 2005. 'Focal colors are universal after all'. *Proceedings of the National Academy of Sciences* 102, 8386–8391.

Roberson, Debie, Ian R. L. Davies, and Jules Davidoff. 2000. 'Color categories are not universal: Replications and new evidence from a Stone-age culture'. *Journal of Experimental Psychology: General* 129, 369–398.

Saunders, Barbara A. C., and Jaap van Brakel. 1997. 'Are there nontrivial constraints on colour categorization?' *Behavioral and Brain Sciences* 20, 167–228.

Strang, Gilbert. 2009. *Introduction to Linear Algebra*. Wellesley-Cambridge Press: Wellesley.

Talmy, Leonard. 1988. 'Force dynamics in language and cognition'. *Cognitive Science* 12, 49–100.

Vapnik, Vladimir, and Alexey Chervonenkis. 1974. *Theory of pattern recognition [in Russian].* Nauka: Moscow.

Widdows, Dominic. 2004. *Geometry and meaning.* CSLI Publications: Stanford.

Witthoft, Nathan, Jonathan Winawer, Lisa Wu, Michael Frank, Alex Wade, and Lera Boroditsky. 2003. 'Effects of language on color discriminability'. In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.*

Biographical note

Gerhard Jäger obtained his PhD at Humboldt-University in Berlin in 1996. He became professor of semantics in syntax at Bielefeld University in 2004. Since 2009 he is professor of general linguistics in Tübingen.

He has worked on a variety of topics in semantics and mathematical linguistics, including dynamic semantics, categorial grammar, and Optimality Theory. In recent years his research has mostly focused on evolutionary and game theoretic approaches to the dynamics of language.

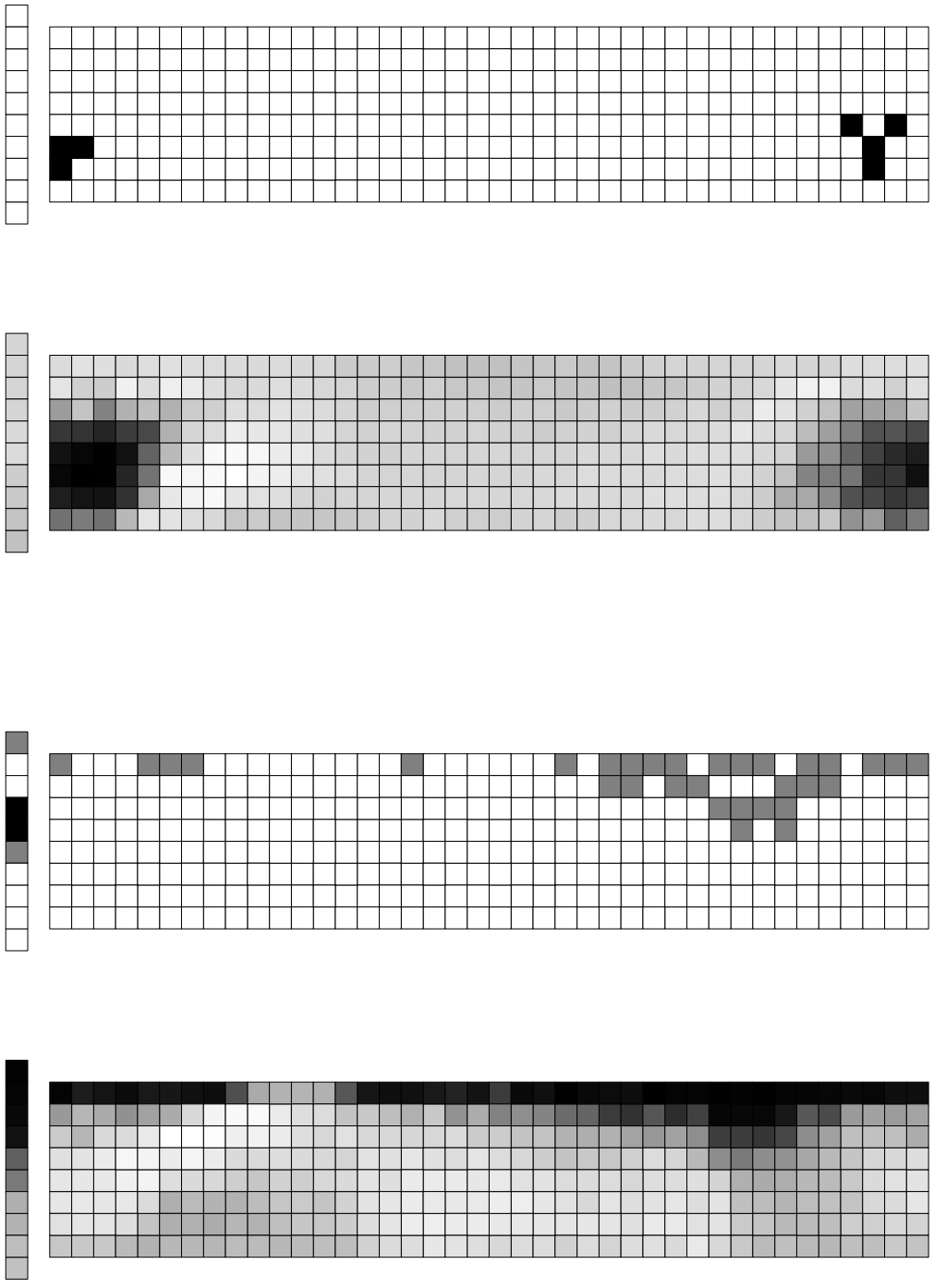

Figure 7: The effect of dimensionality reduction

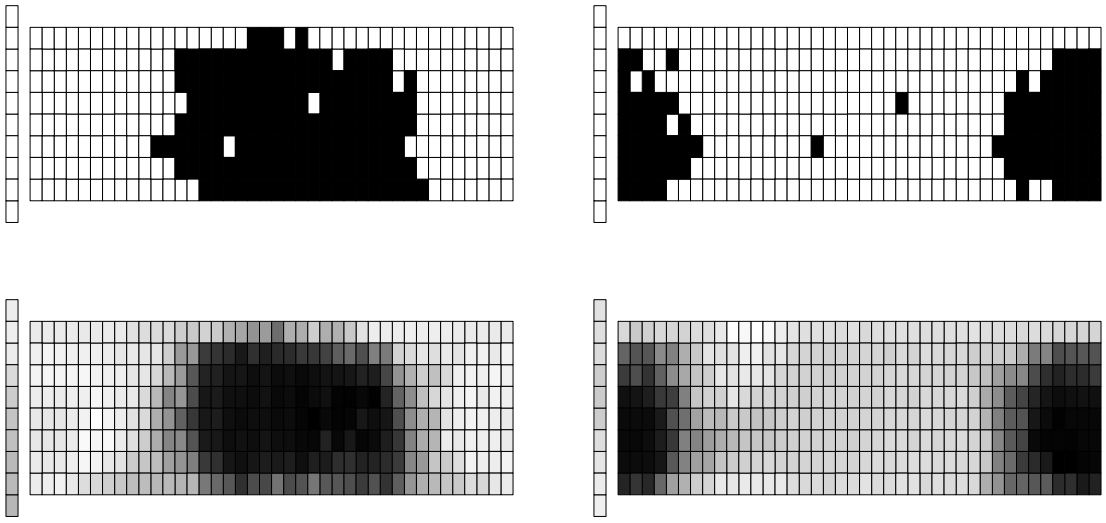

Figure 8: Dimensionality reduction applied to data from a single test person

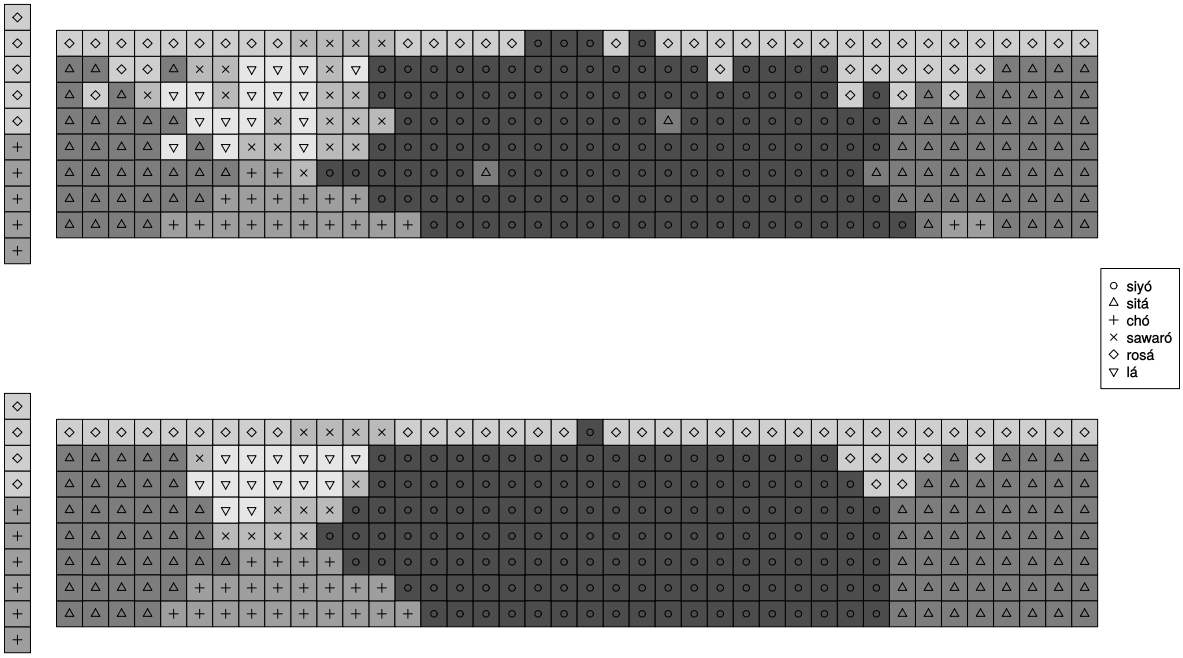

Figure 9: Raw and smoothed partition: Central Tarahumara

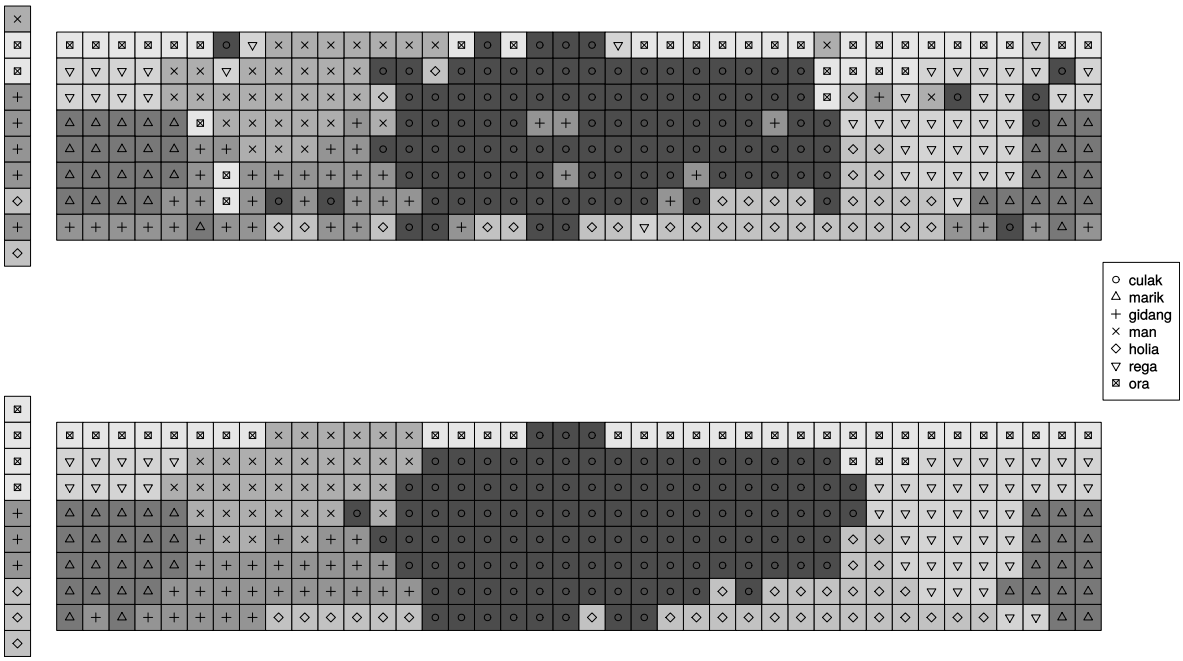

Figure 10: Raw and smoothed partition: Didinga

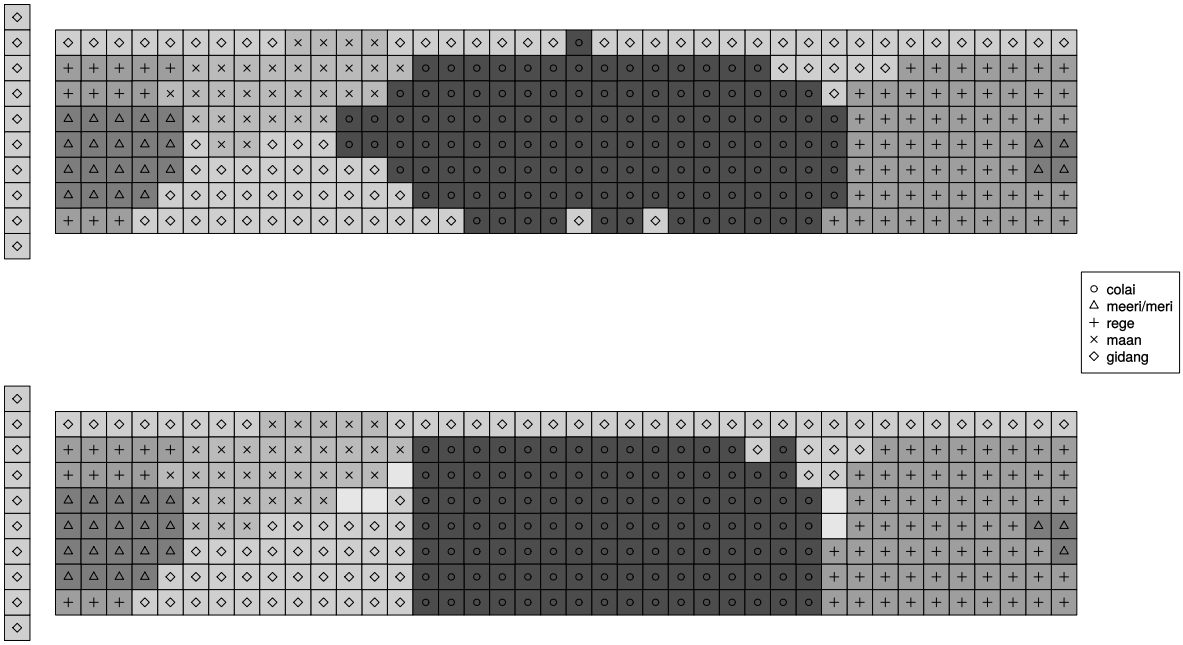

Figure 11: Convex approximation of a partion: an example from Murle

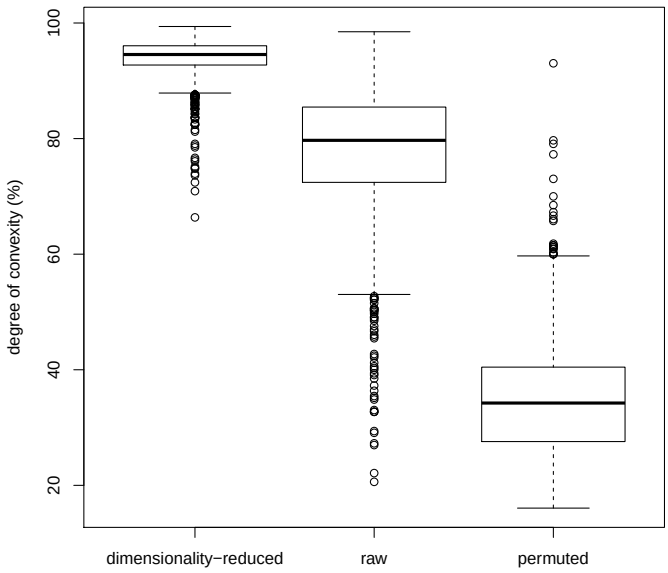

Figure 12: Degrees of convexity (in %) of 1. cleaned-up partitions, 2. raw partitions, and 3. randomized partitions

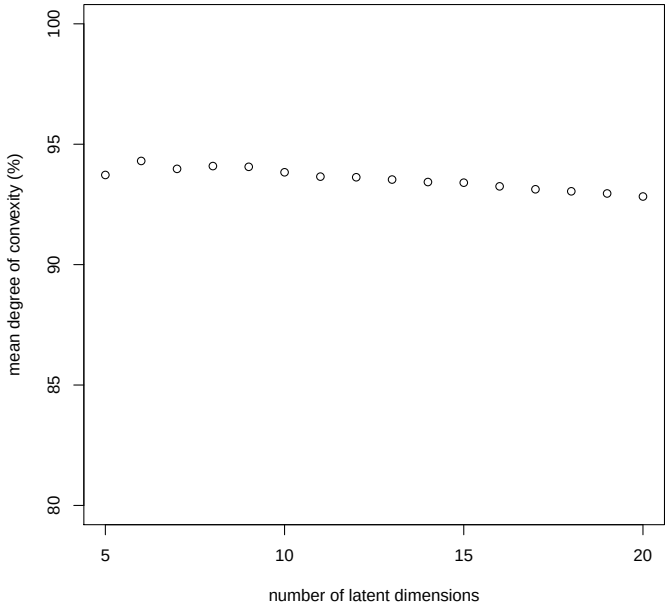

Figure 13: Mean degree of convexity as a function of k