Comenius University, Bratislava
Faculty of Mathematics, Physics and Informatics
Department of Applied Informatics

Modeling of color categorization using distinguishing criteria

Rigorózna práca

Mgr. Kristína Rebrová

Bratislava, 2011

Abstrakt

Ľudské myslenie a kognícia sú v úzkom vzťahu s našou schopnosťou kategorizovať. Keďže je fyziológia vizuálneho vnímania je univerzálna, môžeme skúmať a porovnávať kategorizáciu a pomenovávanie farieb v ľubovoľnom jazyku, kultúre či umelom systéme. Analýzy klasterizácie dát zo Svetovej farebnej štúdie (World Color Survey), ktorá obsahuje výsledky experimentov zo 110 jazykov neindustrializovaných kultúr bez písma potvrdili, že perceptuálne koreláty základných farebných termínov ľubovoľného jazyka spadajú do množiny univerzálnych základných farebných kategórií. V práci sme popísali náš model kategorizácie farieb založený na sémantike rozlišovacích kritérií. Zo simulácií dvojakého typu, s učiteľom, na báze WCS dát, a bez učiteľa, na báze jazykových hier, vyplynulo, že tento model je vhodný pre problematiku pomenovávania farieb a farebných kategórií a porovnateľný s podobnými, známymi modelmi.

Kľúčové slová: kategorizácia farieb, základné farebné termíny, World Color Survey, rozlišovacie kritéria

Abstract

The human thought and cognition are in close relationship with our ability to categorize. Since the physiology of color vision is universal, we can study and compare color categorization and naming in various languages, cultures, and artificial systems. Perceptual correlates of the basic color terms of any language fall into universal set of color categories, as confirmed by clustering analyzes of the data from the World Color Survey, consisting of color naming experiments from 110 unwritten languages from non-industrialized cultures. In this thesis we propose a model of color categorization based on the semantics of distinguishing criteria and its evaluation in experiments with the supervised learning of the data from WCS and unsupervised learning on the basis of language games. Concluding the results from the simulation we confirm that the distinguishing criteria are suitable for the modeling of color categorization and comparable with other, well established models.

Keywords: color categorization, basic color terms, World Color Survey, distinguishing criteria

Contents

Chapter 1

Introduction

The human thought and cognition are in close relationship with our ability to categorize. Categorization plays a dominant role in perception, reasoning, planning and action, speech, and many other cognitive abilities. Similar importance is attributed to the language ability, which can be considered a tangible feature discriminating the quality of human cognition from cognitions of any other species. Therefore the study of basic categorization mechanisms within language could be a way to understand the cognition and thought as such. Within cognitive science, cognitive linguistic school [20] is concerned with such notions.

The main topic of this thesis is color categorization and its computational models. The importance of color categorization rests in its universality. The visual perception can be considered the most important of our senses. The process of color discrimination must be therefore present in our thought, and consequently in language. Since the physiology of color perception is universal, the ability to discriminate colors is not culturally dependent. Thus we can study color categorization and naming in any language, culture, or a (biologically motivated) artificial system, search for similarities, general mechanisms, and differences. The research on color categorization and its modeling is closely connected to the emergence of modern paradigms explaining how categorization functions, namely the prototype theory.

In the following chapters of this thesis we provide an essential background for understanding the phenomena involved in color categorization, we describe the central theory about basic color terms, and finally we present our distinguishing criteria based model of color categorization. Firstly we introduce the basics of human color perception, i.e. the main theories of color vision, its physiology and psychology and a short overview on how the color can represented in a color space. The second theoretical chapter is dedicated to categorization, mainly to the prototype theory of categorization. It also provides a definition of the basic color categories and relevant background on color categorization as such.

In the fourth chapter we describe the main hypotheses regarding the basic color terms, their universality and evolution and the World Color Survey aimed to validate, invalidate or modify these hypotheses. The data from the WCS consist of the results from color naming experiment from 110 unwritten languages from non-industrialized cultures from all around the world and are freely available on the Internet. We also include a short overview on our version of the WCS experiment in the Slovak language we carried out in spring 2008. Lastly, we provide an overview of selected studies regarding the data from the WCS, aimed to verify the original hypotheses. We also describe the first visualization of these data, which inspired us to create our visualization methods. A section about these methods closes this chapter.

In the last chapter we describe our model of color categorization based on Rosch's prototype theory and the semantics of the distinguishing criteria. We provide two-fold evaluation of the model in experiments with supervised and unsupervised learning. In the first case we created simulated agents trained on the data from the WCS. In the second case, we tested the distinguishing criteria in multi-agent simulation of the perceptual color categorization and emergence of the shared lexicon of color terms using the semantics of language games [33]. The original studies with these simulations have shown that mild constraints on the perception and cultural (verbal) interaction among the agents result in categories that have a distribution similar to human color categories. However, the aim of this thesis is not to prove or disprove the original hypotheses of Berlin and Kay or any other hypotheses, but rather to probe the learning capability of the distinguishing criteria on the real data represented in a perceptually uniform space.

Chapter 2

Color perception and color coding

This chapter provides an introduction to basic principles of human color perception, basic theories of color vision, its physiology and psychology. In short, the visual apparatus perceives color by means of absorption of light by three different types of photoreceptor. The essential mechanism involved in many aspects of color processing is comparative principle. It can be found either on primitive level, where the response of each type of color-sensitive receptor is formed in comparison with reactions of other types of receptors, or on cognitive level, where the color of an object is distinguished based on comparison with its surroundings. Another principle present on every level of color perception is the principle of color opponency, generally known as opponent process theory, based on antagonistic relationship between pairs of primary colors, red-green, blue-yellow and black-white.

Color vision is the ability to detect and analyze changes in composition of the wavelength of light [37]. The definition of color in general is an undecided problem. On one hand, it is a property or a byproduct of special range of electromagnetic radiation. Color percepts then can be described as psychophysical properties of perceived object or material. On the other hand, color can be defined as a property of mind. It is possible to imagine certain color with our eyes closed. Therefore to describe color in general without binding on certain light and material conditions we describe it as a perceptual entity with three components – hue, value and saturation, where hue encompasses the quality of the percept. The other parameter, value (intensity, lightness or brightness) stands for the intensity of (white) light apparently coming from the colored object. When the light is at its fullest intensity, colors will become bright, at its least intensity, colors become dim. At last saturation, chroma or colorfulness tells us about the purity of color or more precisely about the amount of white light (or gray paint) mixed with the hue.

The visible spectrum (or simply light) is the portion of the electromagnetic spectrum that is visible to human eye. A typical human eye will respond to wavelengths from about 380 to 750 nm [32]. A very common mistake is an idea, that only the wavelength of light falling on perceived object determines the color perceived. On the other hand, the visual perception is in principle based on the comparison of attended objects with their background. Our perceptual system is not able to measure specific wavelengths, only compare them among objects perceived.

2.1 Theories of color vision

There are two generally accepted theories of color vision, both formulated in 19^{th} century, which were considered rival to each other. This section is based on [11] and [9].

The trichromatic theory (or Young–Helmholtz theory) suggests that there are three types of color receptors (red, green, blue) sending the values of their excitation the brain according to the color of the perceived light. Note that it was formed before physiological evidence for this phenomenon was found and described. The theory is based primarily on color mixing experiment and suggests that a combination of three channels is sufficient for creating any color. Although it truly corresponds with the three types of color sensitive receptors, this concept fails to explain the uniqueness of four color primaries (explained by the opponent process theory described below), and also why dichromats (people missing one type of color cones) are able to see white and yellow even though it is impossible to mix these colors with one channel missing.

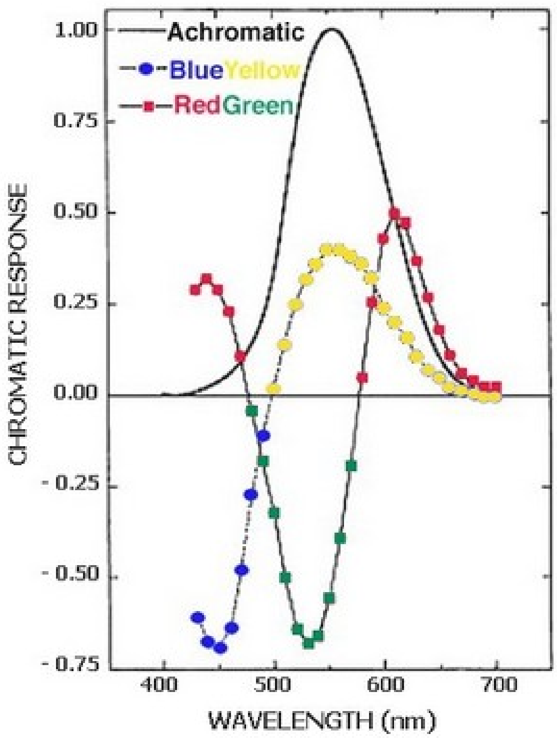
The the opponent process theory, important also for the further content of this thesis, also distinguishes three perceptual channels, but unlike trichromatic theory it defines six primary colors (also called Hering primaries after the author of this theory) into three antagonistic channels: red-green, blue-yellow and black-white (rather dark-light). It means that at any time either red or green is perceived, but never greenish-red (the same with blue and yellow). This principle easily explains the color-afterimage and other psychological phenomena (described in chapter 2.3), but in its original form it never challenged the trichromacy of initial stage of color processing expressed by the first theory.

The experimental background for this theory was provided by Hurvich and Jameson in 1957. Using hue cancellation method, they isolated psychophysical color opponent channels responsible for the antagonistic mechanisms in the perception of these unique colors. The figure below describes the chromatic response cancellation curves for the three Herring's color channels. Note that there is no single wavelength at which a pure red can be perceived. Hence the pure red is extra-spectral and can be perceived only when the yellow component is canceled by the blue component of specific wavelength.

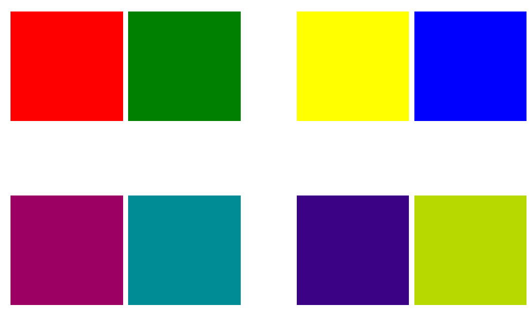
Some later studies considering color opponent processes summarized in [16] suggest that at physiological level, the pairs of mutually inhibiting color primaries are not exactly red-green and blue-yellow, but rather cherry-teal and chartreuse-violet or some simple rotation of these axes (similar hues with equal spacing in adequate color space). The uniqueness of original colors proposed is explained as resulting from the irregularity of shape of perceptual color space (described in section 2.4).

Figure 2.1 depicts opponent chromatic responses measured by Hurvich and Jameson, adapted from [11]. Figure on the right illustrates color hues which might be considered the classic opponent hues: red, green, blue and yellow and the new hues proposed by Jameson and D'Andrade: cherry, teal, violet, and chartreuse.

Concluding the proposals of these two basic theories of color vision we can say that they are not opponent, but complementary. Firstly, there is the trichromacy of receptors in retina; secondly there are opponent processes in the neural pathways visual cortex. On the other hand, the opponent process theory is valid also in three-type cones part, since yellow is sensed both by red and green cones, but not by blue cones (which's response is mutually inhibited with the "yellow signal"). More details are provided in the following section.



(a) Chromatic response cancellation curves for the three opponent channels



(b) Illustration of basic opponent hues

Figure 2.1: Opponent processes and hues

2.2 Physiology of color vision

The mechanism underlying the visual perception is the absorption of photons, the particles of light, carried out by special, light-sensitive receptive cells on the retina of human eye. There are two types of photoreceptors, rods and cones, named after their characteristic shape. Cones are those responsible for vision at daylight, acuity of vision and for color discrimination. This section is mainly based on [1] and [12].

Unlike many other mammals', the human eye owns three types of cones sensitive to different ranges of wavelength we call red (L), green (M) and blue (S). The red or L-cones (long-wavelength sensitive cones) react to light of wavelength between 500 nm and 700 nm with peak around 564–580 nm, green cones are tuned a little "shorter", to 450–630 nm with peak around 534–545 nm and blue cones to 400–500 nm culminating around 420–440 nm. Important is that all cones detect wavelength of incoming light only in comparison with at least one other type of cones.

Red and green cones are together with the color discrimination responsible for the fine discrimination between bright and dark. On the other hand the short-wavelength sensitive blue cones provide the perception of color contrast and are 10-times shorter in amount than the others. Due to a considerably big interception between the ranges of longer wavelength sensitive cones, the yellowish lights will stimulate them nearly equally, but blue cones will be hardly influenced. In addition, these two types of cones can be also considered red-green and green-red cones. The yellow color, perceived by these two types can be intuitively divided into two parts in the visible spectrum, which we can call the greenish and the reddish yellow. This relationship gives rise to an idea, that even if there were three types of cones, there is a possible explanation for four primaries, because yellow is perceived both by red and green cones opposite to blue, which is only in range of blue cones. The trivariance of color channels allows us to perceive also extra-spectral colors like cyan or magenta.

Color opponent mechanisms are present on the neural stage of color processing as well. After being perceived by photoreceptor in retina the visual information is sent via optic nerve to the thalamus to synapse at the lateral geniculate nucleus (LGN). In the LGN the red-green signal is processed by the parvocellular chromatic channel and the blue-yellow by the koniocellular channel both functioning on opponent principle [6]. The LGN transmits

its signals to the first visual area (V1) to process visual signals in cerebral cortex. Opponent processes continue also in V1 where double opponent cells are clustered within localized regions called blobs. Red-green cells compare the relative amounts of red-green in one part of a scene with the amount of red-green in an adjacent part of the scene, responding best to the local color contrast (red next to green). From the V1 blobs, color information is sent to cells in the second visual area, V2. Neurons in V2 then synapse it onto the cells in area V4, which provides input to the inferior temporal lobe ("IT" cortex), which integrates the color information with the information about the shape and form.

2.3 Phenomenology of color perception

The process of perceiving and distinguishing different colors is, similarly to visual perception mechanisms, based on the comparison of the object with its background. This process is generally based on comparing wavelength and intensity of perceived lights, but there are certain influential factors worth mentioning. The following part will briefly summarize most important of the mentioned properties and effects of cognitive psychological nature, which they cause. All of these effects are apparent changes (or inconsistencies) in the colors perceived[1]. This section is based on [9] and [13].

First of all, let us describe the physical properties influencing color perception. An important factor influencing the perception in general is size of the object perceived. In case of color, the smaller the object is, the stronger is the convergence of its apparent color – from dark hues to black (e.g. small blue object appearing as black) and from light ones (e.g. yellow) to white. The lightness of a color depends on various physical characteristics, mostly on the luminance of the material, the background and the properties of ambient light. Increase in the brightness of light causes an apparent shift of all colors of the visible spectrum, those characterized with wavelength below 500nm towards blue and over 500nm towards yellow (e.g. red seems yellower or more orange-like). This is called the Bezold-Brücke shift. On

[1]We will use basic color terms of English language, as this study claims in the next chapters, these terms can be universally translated to most of the western civilization languages. Please note that the following pages are not about linguistic categories, but rather about certain color percepts named with exact color terms for the simplicity of the explanation.

Figure 2.2: Illustration of simultaneous contrast (left) and assimilation effects (right)

the other hand in conditions with low illumination the retina becomes more sensitive to shorter wavelengths and less to longer wavelengths. This phenomenon is named Purkinje effect or shift after Czech anatomist Jan Evangelista Purkyně. For example a blue and a red flower, which appear of the same brightness in daylight, will begin to appear unequal in this respect as twilight deepens. The red color will appear darker more quickly than blue, which might appear even brighter than before.

Probably the most influential element of color discrimination principle is the background. Each and every color (similarly to shapes and orientation) is perceived by process of comparison of the focused object with its surroundings. In accordance with the Opponent process theory, the background influences the perception mostly when it is in an opposite-color relationship with the focused object's color. The background can both induce its complementary hue into an object (e.g. if the background is green, the object will appear redder) or reduce apparent saturation of similar hue (e.g. a very red background will induce green into an object), a highly saturated background will desaturate objects of the same hue and enhance saturation of objects with complementary hue. This principle functions with brightness and is generally called simultaneous contrast.

Assimilation effects, opposite to the first type, cause an apparent change of the color of the background. These are the cases where the foreground instead of producing contrast causes the background to seemingly spread into it. This effect is complete at the point of spatial fusion when the stimuli are no longer viewed as discrete, but fuse into a single stimulus. Spreading, however, occurs at spatial frequencies below those at which fusion occurs. Thus, the stimuli are still observed as distinct from the background, but their colors begin to blend. In the example below white bars spread to make the blue look lighter and the black bars spread to make the same blue appear darker.

Another principle based on and accounting for the Opponent process theory is the successive brightness contrast. It occurs when the viewer has been under prolonged exposure to light of a particular color. This can produce something like a color afterimage effect resulting in inducement of complementary color. For example, viewing a red field would make a subsequently viewed yellow or white object appear greenish. On the other hand there can be a reduction of apparent saturation. For example, adaptation to a red field would then make a pink object appear whiter. This property can also induce or reduce brightness; for example, viewing a bright field would make a subsequently viewed object appear dimmer[2].

Next group of factors, which influence the human color vision are the properties of perceptual system itself. Together with other perceptual constancies as the constancy of shape, size or distance, color constancy is the perception of an object or its quality as constant under changing conditions. In this case it is the color of perceived object, which remains relatively constant under varying illumination conditions that is caused by the independent changes in responses of the three types of cone photoreceptors. This effect is also called chromatic adaptation. Color constancy experiments show that very large spectral changes in illumination cause only small changes in the appearance of objects. There is common consensus that the magnitude of color constancy corrections is very large, but also that the constancy is never perfect [23].

2.4 Color spaces

As mentioned before, color can be cognitively defined as perceptual entity with three qualities. A color space can consist of hue, value and saturation or any other reasonable set of quantitative parameters, which will stand for the axis in vector space. An important characteristic is the gamut – the certain complete subset of colors with a color space. The most common examples are color spaces based on popular working color models RGB or CMYK[3],

[2]For instance “rapid light adaptation”, the sensation of coming from strong sunlight to a dark room or other way round causes the colors perceived during first few seconds to look lighter and less saturated as there was some kind of white or gray aperture in front of one’s eyes.

[3]A color model unlike a color space has no associated mapping function to an absolute color space, in which perceptual difference between colors is directly related to distances between colors and where the interpretations of colors in the space are colorimetrically defined without reference to external factors. In other words it misses a globally understood system of color interpretation. For example there are several RGB-based color spaces like

used in CRT displays in first case or printing devices in the latter. Following lines will provide brief introduction to several color spaces designed to depict human color vision. This section is based on [16] and [19].

2.4.1 CIE color spaces

The CIE 1931 XYZ color space was one of the first mathematically defined color spaces, created by the International Commission on Illumination (CIE) in 1931. It is a linear transformation of older CIE RGB space, derived from a series of color perception experiments. It locates color within a tristimulus coordinate system (XYZ are the coordinates) derived from cone responses of retina. A great success is that it resembles perceptual appearance of colored light and the effects of the color mixture of lights by simple addition of vectors. On the other hand it does not correspond well to the perceptual color differences, i.e. the distance (the difference-similarity) measure, between different colors does not correspond directly to human perceptual judgment.

The next generation of color spaces created by CIE are those color-opponent based, derived from CIE 1931 XYZ, from which the most popular is the L*a*b* (or CIELAB) system. Strongly influenced by the Munsell color system (see next section), the intention of CIELAB is to create a space which can be computed via simple formulas from the XYZ space, but is more perceptually uniform [4] than XYZ. Inclining to the opponent process theory and more cognitively defined color spaces (Munsell, HSV or HSL[5]) the three coordinates of CIELAB represent lightness of color (L* = 0 yields black and L* = 100 indicates diffuse white), its position between red/magenta and green (a*, negative values indicate green while positive values indicate magenta) and its position between yellow and blue (b*, negative values indicate blue and positive values indicate yellow).

Adobe RGB and sRGB, created from the model using real colorimetric parameters.

[4]A change of the same amount in a color value should produce a change of about the same visual importance.

[5]HSV/L coordinates are hue, saturation, value/brightness or lightness/luminance, similar color models based on RGB. In both cases the color hues are arranged on a circle in a spectrum-like manner, while other parameters are in percents. Saturation decreases towards the center of circle, which is essentially gray and lightness/brightness decreases (or in the second case also increasing) to form a cylinder or a double-cone.

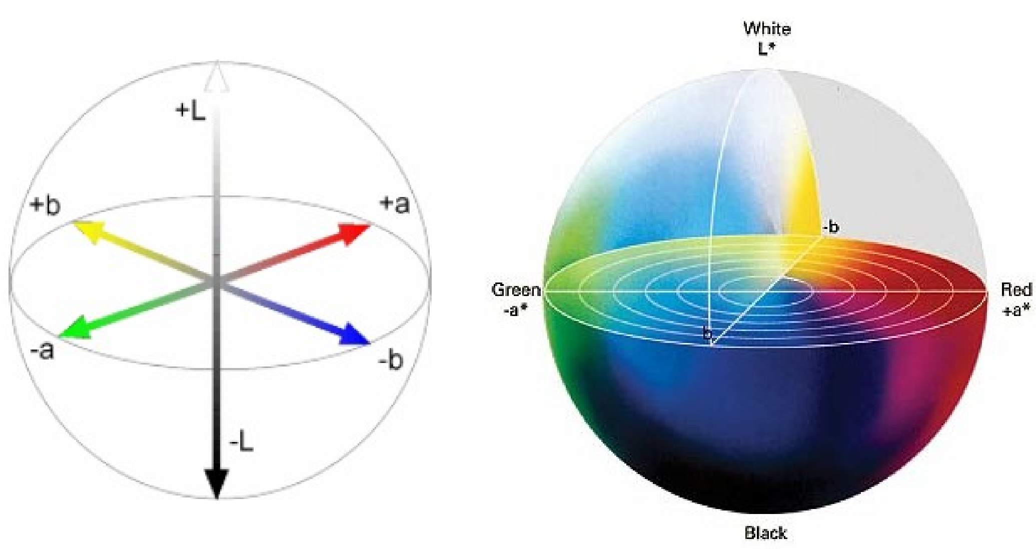


Figure 2.3: Illustration of L*a*b* Color Space

2.4.2 Munsell color system

The goal of perceptual color spaces is to reflect the perception of similarities between colors in the distance between each point of a color space. Munsell color system is one of the best known of these perceptual color spaces. Its coordinates are based on perceptual color properties – hue, value and chroma (approximately corresponding with saturation). The basic idea is that adjacent color samples in each dimension should have a constant perceptual difference, so that the color chips are located at equal perceptual intervals along each dimension. However, no method is given for comparing distances along these dimensions measured in different units.

In their theoretical summary [16] Jameson and D'Andrade suggest that scaling studies (scaling of color spaces to reach the hypothetical ideal color space) found out an overall pattern which clearly conforms to the Munsell type of organization. However these scaling studies do not describe an axis on which green is opposite to red, but rather red opposite to blue-green and green opposite to red-purple. The results also show the color perimeter divided into five not four equal sections, corresponding to red, yellow, green, blue, and purple. So pure red and pure green, assumed to be opposing colors, are not found to lie at opposite sides of an achromatic point in empirical scaling of perceptual space. This is another factor suggesting that a variant

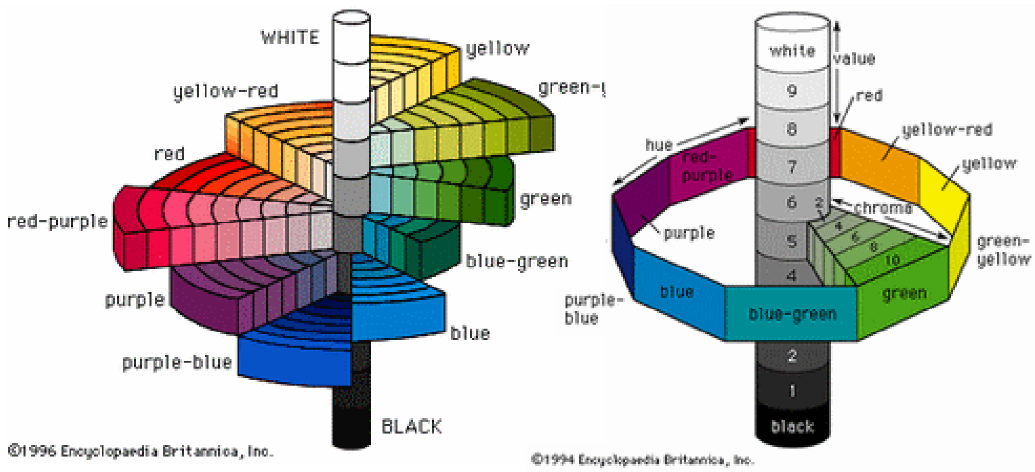


Figure 2.4: Illustration of L*a*b* Color Space

of opponent process theory, involving small changes of hues of the basic primaries is needed. Figure 2.4 depicts the Munsell color system and was taken from [25].

Chapter 3

Categorization

This chapter is dedicated to categorization, which we consider essential for the human cognition. The categorization as a mechanism is employed in perception, reasoning, planning and action, speech, and any other cognitive tasks. We will describe the *classical view* on categorization which dominated in philosophy and related sciences for more than two thousand years and is still very influential. However, with emergence of cognitive sciences, new theories arose to overpower the old theories, not only on categorization, but also on cognition as such. We will thoroughly describe the *prototype theory* of Eleanor Rosch, that challenged the rigid classical perspective. Since this view (proposed by Rosch and followed by Lakoff and many others) is the one we consider to be the most coherent and explanatory, and is also connected to the latter parts of this thesis, we will not include other influential theories that were constituted later. The two generally accepted examples of such theories are the *exemplar view* and the *theory view*. The exemplar view claims that there is no single representation of an entire concept; members of a category are formed by specific representations of its instances – exemplars. The theory view (also called the "knowledge approach") refers to concepts as to mental theories about the world; membership in a category is decided due to an individual's knowledge [24]. Following text is based mainly on [29] and [20].

3.1 From classical view to Eleanor Rosch

The classical view on categorization was established by philosophers of ancient Greece. From that time up to early 18^{th} century, categories were understood as some closed containers filled with things sharing the same (necessary and sufficient) properties. Respectively these properties defined the categories. Entities were either members of certain category or not (Boolean membership). There are other ideas of the ancient philosophy closely related to this notion, for example the idea that reason is a disembodied symbol manipulation, or that mind is separate from body, or that meaning is based on truth and reference (i.e. each word corresponds to one thing in the world).

The first one to notice the flaws in the classical view was Ludwig Wittgenstein. In his later work he pointed out that there are categories that have no clear-cut boundaries, constituted of members that do not necessarily share the same properties, and that members of the categories might be central and non-central. His most famous example was the word game. There are various types of games for example, children games, table games, or sports that share no properties at all, some are based simply on the chance, some have rules, some do not, there are even games that does not have a winner. The only thing they have in common is that they are grouped in the same category. For this principle Wittgenstein found a very nice parallel – the family resemblances. The members of a family resemble one another in various ways, but usually there is not a single feature they all share. This principle also accounts against the closed boundaries. There is always a possibility to gain a new family member (marriage, birth, etc.), without it, the world would not function. Another fine example is the category number. For a long time, since ancient Greeks, numbers have been only integers, but then rational numbers came along with the need to represent fractions, then real, complex, and transfinite numbers were discovered. This example also shows the centrality and non-centrality of category members. Every precise definition of number must include the integers, but not every definition must include transfinite numbers.

Another important step towards a new definition was the theory of fuzzy sets proposed by Zadeh. A fuzzy set is characteristic with its member function, which allows each member not just to belong or not belong to the category (0 or 1), but to be a part of it to some extend (a number between 0 and 1). This concept provides categories with fuzzy boundaries and accounts for the centrality character of categories, a central member for example may belong to the category at 100%, but non-central only at 20%.

The first one to propose a new general theory of categorization was Eleanor Rosch (in early 1970's). Her main proposal was that thought in general is organized in terms of prototypes and basic level structures. To demonstrate her statements she established new research paradigms in cognitive psychology and proved them on experimental basis.

3.2 The prototype theory of categorization

This theory was also called "the theory of prototypes and basic-level categories". Note that since Rosch is one of the first cognitive linguists she speaks mostly about categorization in language, with direct consequences on categorization in general. We are not aware of any reasonable objections against this generalization. Rosch [29] describes two principle of the formation of categories[1]:

Cognitive economy: categories are built in a way that maximizes information gain, but also preserves an adequate level of simplicity. The goal is to differentiate which information is important and which is not.

Perceived world structure: since the world is not unstructured total set of equiprobable co-occurring attributes (as expected in the classical view), there are attributes occurring together often and those that are never connected. The perception of such world is always driven towards a high correlation, so people will not think about (consider) features, which are not generally known to be interrelated or that have no logical connection.

The prototype theory, also called "the theory of prototypes and basic-level categories", proposes that categories and the process of categorization have following characteristics:

1. **Family resemblances** (Wittgenstein): members of a category may be related to one another without all members having any properties in common (i.e. a category does not have to have defining features).

[1]Under the formation of categories she means their formation in the culture, not the development of categories in children born into a culture. She also does not intent to constitute a model of how categories are processed (how categorizations are made) in the minds of adult speakers of a language.

2. **The basic level** (Rosch): the categories are not merely organized in a hierarchy from the most general to the most specific, but are also organized so that categories that are cognitively basic are in the middle of a general-to-specific hierarchy. These basic categories are simple and short words, most commonly used labels and most neutral terms for category members, first named and understood by children, and first to enter the lexicon of a language. Considering for example a hierarchy mammal – cat – Siamese. When talking about an animal sitting on the window it would sound strange saying: "The mammal sat on the window."

3. **Prototypicality** (Rosch): there are members of a category that are more "typical" than some others and there are methods for deciding this typicality. The more prototypical of a category a member is, the more attributes it has in common with other members of the category and the fewer with members of contrasting categories.

A very common notion is that according the prototype theory the membership function is a comparison of the category prototype with the considered samples. Rosch herself is against any tendencies trying to present prototypes as a processing model for categories, a theory of representation of categories (prototypes as representations of categories in mind), or the theory of learning categories. The prototypes or more precisely the judgments of degree of prototypicality are in these terms more of a property, than the base of categorization. However, some studies claim that in certain cases categories are built around prototypes – focal colors (see section 3.3).

The comparison with the prototype is also a useful mechanism for artificial intelligence based categorization frameworks. The distinguishing criteria we used for the color categorization model described in chapter 5 are working on the basis of such comparison: each distinguishing criterion stores a prototype and reacts to an input with the activity proportional to an exponentially decaying function of the squared distance between the input and the stored prototype. However, the distance function takes into account statistical characteristics of the sample set, such as variances of attributes and their mutual covariances (for details see further in the text or [36]). Unfortunately there are shortcomings of this approach, e.g. that categories (or categorization mechanisms) based on single prototypes are not able to encompass the composite categories, i.e. those that have multiple different prototypes.

3.3 Basic color categories

According to the color categorization definition in [37], there are two basic processes constituting color categorization. The lexical color categorization, characterized as the division of color sensations into classes corresponding to the significata of the color words of a particular language, and the perceptual color categorization, the division of the color sensations into classes by the perceptual processes of an (any) organism. There are various studies aimed to find a correlation between these two phenomena.

In 1969 Berlin and Kay [5] suggested that there exists a final set of basic color terms which describe all percepts in any language. This thesis will be discussed in more detail in the next chapter. Here we would like to concentrate on the definition of the basic color terms. A basic color term should satisfy following criteria:

1. it is monolexemic, i.e. its meaning is not predictable from the meanings of its parts (e.g. blue vs. greenish blue)
2. generality criterion
   - the color referred to is not contained within another color category (e.g. scarlet is contained within red)
   - it is not restricted to a narrow class of objects (e.g. blond is used only with hair, complexion, and furniture)
3. it is psychologically salient, common and generally known (e.g. yellow vs. saffron)

In other words, the basic color terms can be considered the smallest set of simple words with which the speaker can name any color. Basic color terms name basic color categories. This assumption is in accordance with the basic-level property from the prototype theory of categorization. Berlin and Kay claimed that each and every language has at least two maximum eleven basic color terms. The base for their proposal was a survey in which they studied not only simple color naming based categorization, but also the best examples of categories. They realized that the best examples for color terms used in each examined language cluster together, including the best examples from English language, so they assumed that color categories are built around this prototypes, and that additionally these prototypes are universal for all languages.

Chapter 4

The World Color Survey

In this chapter we introduce the main hypotheses about basic color terms proposed originally in 1969, and the World Color Survey aimed to asses these hypotheses against a broader empirical basis. Later on we describe an experiment reproducing the WCS experiment in Slovak language with a short overview on its conclusions with emphasis on shortcomings of technical details of the original method we have encountered while performing the experiment. Subsequently we provide an overview of selected studies regarding WCS data aimed to verify the original hypotheses. In the end of the section we introduce our original methods of visualization of the data from WCS, which will be used in the next chapter.

4.1 The main hypotheses of Berlin and Kay

In Basic Color Terms: Their Universality and Evolution [5] from 1969 Berlin and Kay proposed two general hypotheses regarding basic color terms and how they emerge in a language:

1. There is a restricted universal inventory of basic color categories

2. These terms emerge in any language in a constrained order, that can be interpreted as an evolutionary sequence.

The closed set of basic color categories consists of terms corresponding approximately to English black, white, red, green, yellow, blue, purple, pink, brown, orange, and gray. The evolution of basic color terms starts with the distinction between black and white or more precisely between dark and light and it is in correspondence with the Opponent process theory of color vision (described in 2.1). Later research of Eleanor Rosch showed that these first two terms represent rather categories covering warm (e.g. yellow, orange and red plus white) and cool colors (black, green, blue) called composite categories, characteristic with multiple best examples (e.g. warm colors are both red and yellow with different prototypes) [14].

Berlin and Kay proposed seven evolutionary stages characteristic by the color categories present in a language of that stage. First stage was the black-white or dark-light distinction, in the second stage, red emerged. Stage III contained a term for either green or yellow, stage IV both green and yellow; stage V added blue and stage VI brown. The last stage added the remaining of the eleven basic categories (purple, pink, orange, and gray). To this stage belonged also English and similar western languages.

Although the hypotheses of Berlin and Kay have been substantially confirmed by their research, it had certain drawbacks weakening their findings [17]. First of the important methodological objections against their empirical generalizations was, that the amount of twenty languages studied experimentally is not sufficiently numerous to justify universal conclusions. Moreover the number of speakers per language did not exceed three, which leaves us with roughly 40-60 speakers interviewed. Secondly, the data were obtained in Berkeley rather than in native communities, not all of the subjects were skilled speakers of the languages studied, and all of them spoke English as well as their native (inherited) language.

4.2 The course and methodology of the WCS

The World Color Survey started in 1976. Its major purpose was to validate, invalidate or modify the main findings of Berlin and Kay on a broader empirical basis. Data on the basic color term systems of 110 unwritten languages from all around the world were gathered with mean of 24 speakers per language. Once data gathering was completed (circa in 1980), data processing, quality control, and analysis were undertaken. The online data archive was published in 2003 [8]. The following text is based on [17] and [7].

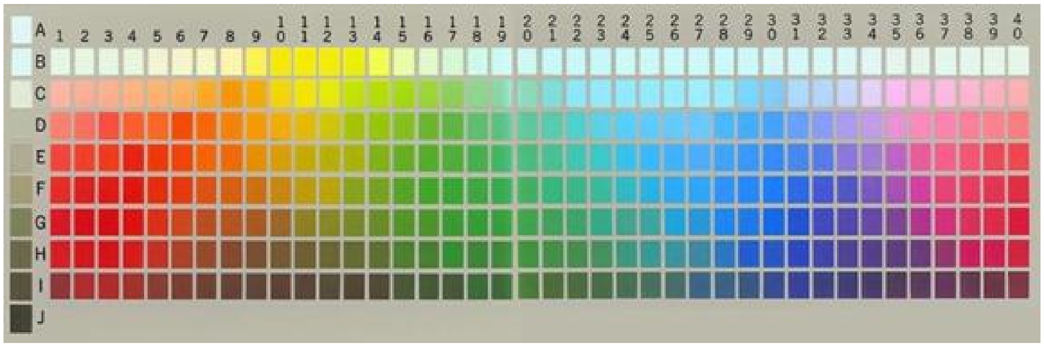


Figure 4.1: The WCS stimulus board

4.2.1 Stimulus materials

Stimulus materials used in both the WCS and Berlin and Kay's experiment consisted of 330 color samples from Munsell color system. There were individual chips in glass 35 mm slide and full stimulus board with all samples on it. The stimulus board, in form of a grid (from now on Munsell color grid) encompasses 320 samples, which represent forty equally spaced Munsell hues from R2.5 in column 1 to RP10 in column 40, at eight levels of lightness (Munsell value) in rows arranged from the lightest on top to the darkest on bottom. The color in each cell corresponds to the maximum available Munsell Chroma (saturation) for that hue-value combination. In addition a left-most column displays ten levels of lightness of neutral shade (black, gray, white). Figure 4.1 illustrates the stimulus board and was taken from [7] .

4.2.2 Experimental setup and methods

Unlike the original experiment of Berlin and Kay, the WCS method consisted of only two tasks. In the original version of the experiment by Berlin and Kay speakers were initially asked to elicit all the basic color names of their language. In the WCS experiment the task of collecting the set of used terms was carried out by the experimenter. The concept of basic color terms was explained to cooperating speakers as "the smallest set of simple words with which the speaker can name any color". Instructed to respond with short, simplest names, observers were shown the 330 samples, one by another in a fixed random order. This was the *naming task*.

In the latter task, called *foci task*, subjects were to choose so called focal color chips – the samples which they considered to be the best examples of color terms they used, separately for each term, on the full stimulus board (described above). Considerable factor influencing results of this task was that the speakers were allowed to choose as many focal samples as they wanted. Considering that they had a possibility to select all samples they previously labeled as forming one category (e.g. ten or twenty samples), the misunderstanding of the task might cause a full failure in detecting the best examples. Additionally there was no backward checking whether the samples selected as best examples of a category were previously classified as members of that particular category during the naming task.

4.3 Slovak color experiment

In spring 2008 we carried out an imitation of the World Color Survey experiment. We gathered data from 25 speakers, 12 men and 13 women, whose mother tongue was Slovak and who had no color vision deficiencies. Since the English language is one of the most popular and nowadays obligatory in compulsory education it would have been hard to get to monolingual speakers. Therefore we did not count this aspect as significant. Probably all of our subjects speak English, some of them German or French as well.

With slight modification of the method we gained data of two types, first consisting only of 11 basic categories proposed by Berlin and Kay and second including some other color terms considered by our subject as highly significant. Results of this experiment on one hand confirm the original hypothesis about eleven basic color terms, and on the other show that the evolution of basic color terms may continue.

4.3.1 Experimental setup and methods

We used methodology from the WCS experiment with several modifications. At first, the stimulus material, unlike the original study, consisted of color samples[1] printed on ordinary office paper with an all-purpose office laser printer. Although we used creation of the samples the original L*a*b* color values and a professional software, the quality must have been remarkably different from the original material.

[1]Rectangles app. 4x3cm glued to app. 5x4cm white thick paper frame.

The modification of the original method consisted of a two-answer questioning mechanism. Speakers were urged not only to use general terms, but also, when they used a word outside Berlin and Kay's 11 basic color terms (e.g. turquoise), they had to name also a counterpart for it from this closed set (e.g. turquoise-green or turquoise-blue). In this case we recorded both the first answer and the second specifying answer, so the data from this experiment finally consisted from two answer sets for each speaker.

Before we began the color naming task, we instructed the subjects to use the shortest and most basic names according to the definition in 3.3. However, unlike our expectations, the discussion about generality of almost every newly added color name was necessary. Subjects frequently considered the names of their favorite colors or terms they use often as generally known and used. Fortunately, after further explanation of the task, most of them agreed upon using more general terms.

An important factor influencing the results of this experiment is that speakers were examined not one by one, but in small groups. They were instructed that there are no "wrong answers" and that it is important, that they would react naturally, but the most likely influenced each other in some cases, especially when undecided about the color name of a particular sample.

4.3.2 Results and discussion

As one result of this experiment, we discovered, that besides the 11 basic color categories there were several other color terms that appeared quite often and were considered significant by the subjects. The most frequent of these were: tyrkysová, kaki, béžová, bordová, and okrová, which roughly correspond to English: turquoise, khaki, beige, burgundy (or maroon) and ocher. We will call them outer color categories.

What our results certainly confirm is that Slovak language contains firmly established 11 basic color terms as proposed by Berlin and Kay for English. To this also adds a fact that in foci task speakers had some difficulties with labeling the outer color categories, but not with selecting best examples for the basic ones. Only one outer category, turquoise, was strong (i.e. frequent) enough to win in a few color samples.

Figures 4.2 and 4.3 are outputs of the visualization method described in section 4.5. In short, the map on top of the figure displays the results of Slovak color naming and foci tasks in the basic (classic) mode. The small number in the left corner is the identification number of the category, which was elicited by most of the speakers. The larger numbers indicate that the color chip was selected as the best example of the category with the displayed serial number. The second map is so called fuzzy visualization, based on proportional mixtures of colors elicited categories derived from the best example colors, for more detail see 4.5. Figure 4.2 displays the results of the experiment with 11 basic color categories and figure 4.3 the results with outer categories.

The greatest drawback of this experiment is that it was not executed on exactly the same stimulus material as the original WCS experiment. During the foci task we encountered that the subjects complained that the color samples missed particularly representative hues of some colors categories, for example of red[2]. This might by caused by the print quality, or more precisely the conversion from Lab color space to the printer's color space, but it could also be a property of the original color values.

We also noted some disadvantages of the method in general. These findings can be useful for a general overview on the WCS experiment as such. First of all, 330 is a considerably high number of samples. Most of the subjects got bored and tired after only one third of the samples. A significant portion of them were of very similar color hues. The ten degrees of lightness for each hue and the proportionality according to the grid, not to perceptual features of the colors, might cause some samples to look nearly the same.

As we mentioned before, we see a great disadvantage in the option of choosing unlimited amount of focal colors. During the experiment we realized that some speakers, when not forced, pick a large number of samples just to "say something". The other drawback of this part of the method was that, when not corrected by the experimenter, subjects have chosen the samples previously labeled with one term, to be the best examples of another term, mostly in the cases of the outer categories. This fact accounts for the original thesis about 11 basic color categories. The inability to react properly to the outer samples and the misplacement of best examples of other categories into regions of outer categories confirms that they should not be considered parts of the basic color categories set.

The most important contribution of our experiment was that we learned more about the methodology of the WCS experiment in practice. Apart from

[2] A large number of the subject considered an ideal red a color hue, which can be in RGB color space coded as (255,0,0). In our printed version of WCS samples there was no such color.

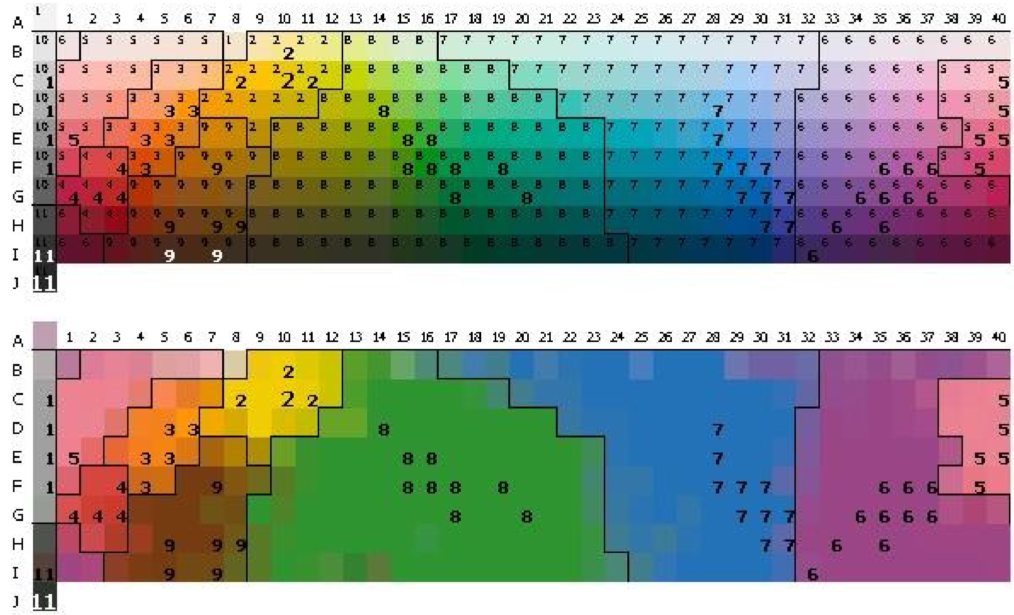


Figure 4.2: Visualization of the Slovak color experiment without outer categories

the difficulties with the amount of samples, we found out that the speakers must be aware of the task they do and be willing to use simple words with which they can successfully communicate about colors with all speakers of their language. Here we would like to emphasize that subject should be made aware of the task during the whole course of the experiment and to be urged to keep focused only on the task.

4.4 Analyses of the WCS data

There are several studies concerning the WCS data aimed to prove the original hypothesis of Berlin and Kay. They are based either on analysis of clustering of color terms, their distribution and salience, but also on artificial intelligence simulation or on the human color vision abstraction into the perceptual color spaces. The oldest way proposed to analyze the WCS languages was through a special conceptual framework, and on the basis of preliminary data summary and visualization.

The main aim of these studies was to confirm that the basic color terms from the WCS data resemble basic color terms of English using the results

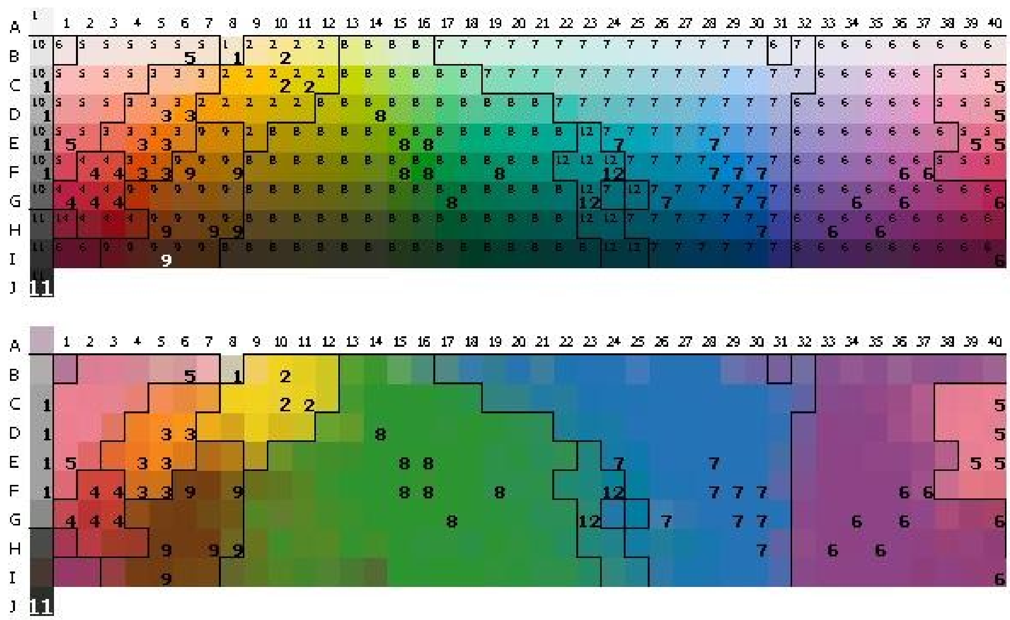


Figure 4.3: Visualization of the Slovak color experiment with outer categories

of the color naming experiments. From these data, items representing color terms used by individual, but also a whole group, can be extracted in form of points in color space. For each language a set of such points can be generated and compared to sets of points representing vocabularies of other languages, but also a universal set of color points, near which color points from all languages will fall.

4.4.1 Preliminary processing: the first visualization

Shortly after the WCS data were gathered, data processing, quality control, and analysis were undertaken for both data entry and data analysis. The preliminary data summary, presented in [7] and [17] included among the others the first visualization of the data, created using simple ASCII symbols.

First, the "naming arrays" displayed the responses from the experiment from both naming and foci task, for each single speaker and for the whole language. These were aggregated results of the naming task across all speakers at various levels of inter-speaker agreement. The modal agreement array (100%) displays for each stimulus chip a symbol corresponding to the term

```
  40% Agreement Level, 10 of 25 speakers
            1         2         3         4
  01234567890123456789012345678901234567890
A +++++++++++++++++++++++++++++++++++++++++ A
B +++++++++++++++++++++++++++++++++++++++++ B
C +##############oooo++++++++++++++++++#### C
D +##############oooooooooooo+ooo++++###### D
E +#############ooooooooooooooooooo######## E
F o#############ooooooooooooooooooooo###### F
G o##########oooooooooooooooooooooo######## G
H o########oooooooooooooooooooooooooo###### H
I o####ooooooooooooooooooooooooooooo####### I
J ooooooooooooooooooooooooooooooooooooooooo J
     Coverage is 100% (330 of 330 chips)
```

```
  100% Agreement Level, 25 of 25 speakers
            1         2         3         4
  01234567890123456789012345678901234567890
A +++++++++++++++++++++++++++++++++++++++++ A
B ++  ++ ++      +++++++++++ +++++ +++ ++   B
C + #     ###                    + +      # C
D +# ## #####                         ##### D
E  ########                           ##### E
F  ###### #                          ###### F
G o#####                    o        ###### G
H  ###          o                      #### H
I o         o   o      o     o              I
J                                           J
      Coverage is 31% (102 of 330 chips)
```

Figure 4.4: Aggregate naming arrays for 25 Wobé speakers

most often applied to that chip, regardless of how often that was. The 30% Agreement array displays for each stimulus chip the symbol corresponding to the term most often applied to that chip only if that term was used for that chip by at least 30%, similarly for other levels of agreement (i.e. 70%, 40%, etc.).

The "term maps" are created for each term separately. In the map for a given term, each chip receives a typographical symbol (including blank) of visual 'density' showing the frequency with which speakers named a particular chip with displayed term. High-agreement symbols tend to occur in the interior of categories and lower agreement symbols at the edges.

Figures 4.4 and 4.5 (taken from [7]) display aggregate naming arrays and term maps for a three-term language Wobé, which will be used also for illustration of our visualization techniques in 4.5. In figure 4.5 symbols "+", "#" and "o" represent color terms. Note that at the 40% level of agreement all 330 chips were named, that is, at least ten speakers gave the modal response for each of the 330 chips, we can consider Wobé a high consensus language [7]. In figure 4.5 a perfect agreement of speakers is displayed by "@" (then gradually decreasing in order: "#","+","-", and ".").

A positive feature of this type of processing is that the terms are imaginable on the WCS color grid, but only for those who already can imagine the stimuli from the experiment. Additionally, there is no possibility to display more term maps in one image, or to display the other categories elicited for a single sample than the winning one (the most frequently used). The visualization method we propose in the next chapter can handle all of these issues, because it uses colored maps instead of characters, and those are able to display more information.

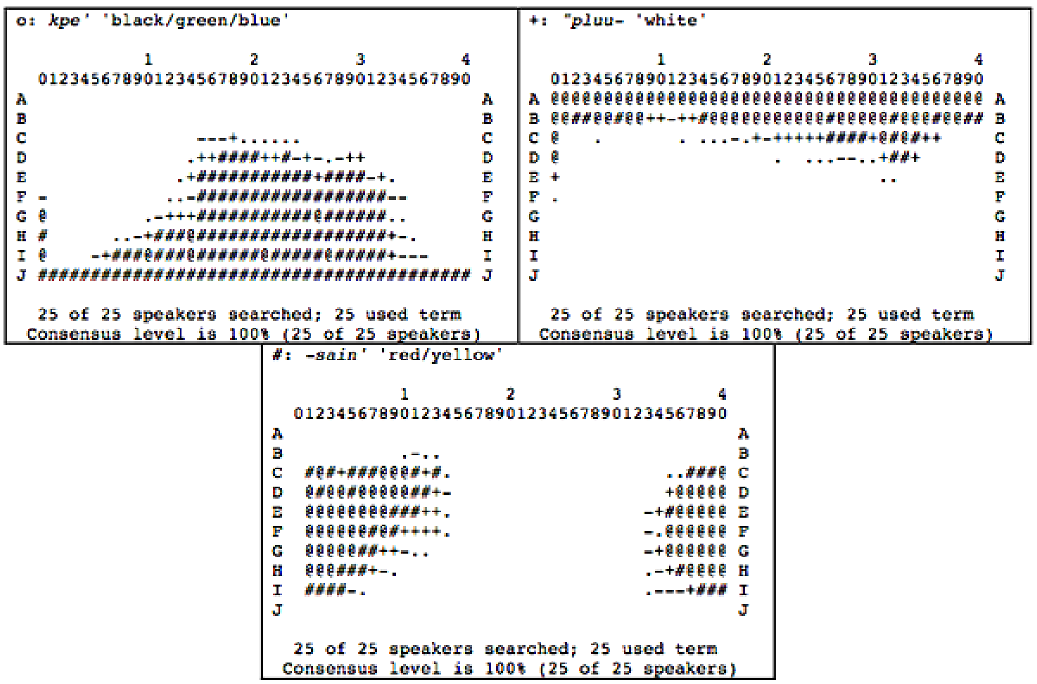


Figure 4.5: Term maps for the three terms of Wobé

4.4.2 Conceptual framework of evolutionary stages

The conceptual framework proposed by Kay and his colleagues [17] divides the WCS languages into groups based on the number of color terms they use. Note, that this significant number of terms is always derived from color naming task winners, i.e. most frequent names collected from all fields.[3]

This concept suggests that an evolutionary development of basic color terms should not be seen as a single process, but as two partially independent processes: the division of composite categories into the six fundamentals (black, white, red, green, yellow, and blue) and later the combination of fundamental categories into derived categories (the other 5 basic color terms). The first progress finally yielding six primaries is categorized into five stages, corresponding with systems containing two to six composite or fundamental categories. Beginning with two basic composite categories, there follow two partially independent processes: dissolution of the white/warm channel and dissolution of the black/cool channel.

Through the two-process mechanism Kay and colleagues explain the fact that a large set of languages developed separate terms for white, red and yellow (or similar distribution of colors covered by warm category), but did not developed separated terms for green and blue, or even the whole cool category. However, this phenomenon still accounts against the originally proposed opponent process theory basis, stating that the next step of the progression will be the opponent color (e.g., when a language already has black, white, red and green, the next stage will be yellow and blue so no primary will be left without its opposite).

In summary, the stage I. consists of two categories – warm and cool, in stage II there happens the separation of warm channel to white and red-yellow in all cases. Differentiation starts in stage III where either decomposition of cool channel or of the red-yellow category occurs. The latter decomposition is influenced by the previous distribution. There are three possible pairs of colors remaining together in this stage, green-blue, black-blue and yellow-green. In the last stage all six primaries are separate.

[3] Here we mean a set of terms, from which each term appeared at least once as a most frequent answer in color naming task. Please note that this winner-take-all-like method, despite its simplicity and power to generalize, has its real drawbacks, mostly in case of colors, that are not typical members of a category. It is possible that from 25 speakers, in 13 cases the winner is some category a, and in 12 category b, so a is a winner, but a weak one.

4.4.3 Statistical evaluation of the WCS data

Reflecting the objections against previous intuitive analyses done by "a hand" several statistical analyses were provided to proof the original hypotheses of Berlin and Kay. In most important study [18] Kay and Regier analyzed whether categories of WCS languages represented in L*a*b* color space[4] were more clustered across languages than would be expected by chance. To represent each color term in each language they first constructed a term centroid[5] for each term by each speaker and then made an average for the whole language. Consequently they derived the degree of clustering of color terms across languages using the measure of their dispersion. For each color term its geometrically nearest term in each language was found and their distance was added to the dispersion sum (see equation 5.4). Finally the dispersion of color terms in WCS languages was compared with the dispersion of randomized data set. Similarly they compared the dispersion of WCS data with the date from the original study of Berlin and Kay, and the hypothetical data. In both cases the dispersion was significantly smaller than within the hypothetical data.

An important aspect of the debate on universality of color categories is the question of the mechanism of category formation. The question is, whether the categories are being formed around prototypes, which are universal, or if they are only constituted by their boundaries and the process in which they emerge is random. The latter view was posited by [28], where authors describe their unsuccessful attempt to replicate Rosch's results [14], in which the prototypes of color categories appeared to by cognitively privileged (easier to remember, identify, etc.). As a reaction on this skepticism another clustering analysis of WCS data [27] showed that the best examples of color terms (collected as color samples in the second – foci task of the experiment) tend to cluster together even more than centroids of categories and that on the basis of these foci the boundaries of categories can be predicted.

In pursue of universal set of color categories in WCS data, Lindsey and Brown [21] used k-means cluster and concordance analyses. The results showed that when divided into 2 up to 10 clusters, the average color-naming patterns of the clusters all gloss easily to single or composite English patterns, and also that the structures of these k-means clusters unfold in a hierarchi-

[4] L*a*b* is perceptually uniform so it well resembles human color cognition.

[5] A centroid is a geometrical center of a shape produced by all color chips named with the particular term

cal way similarly to evolutionary scheme described above. Processing of the WCS data also showed that 8 was the optimal number of WCS chromatic categories: red, green, yellow-or-orange, blue, purple, brown, pink and grue (green-blue)[6], what is also roughly consistent with the framework proposed above. In the second part the analysis of concordance in color naming within WCS languages revealed statistically significantly high concordance across languages in small regions in color space that agreed well with five of six primary focal colors of English. Recently [22], Lindsey and Brown addressed also the problematic of evolution of color terms describing a small amount (3-6) of universal motifs of color category systems appearing in WCS languages. Interesting point is that several motifs can be present in one language and used differently by particular speakers.

4.4.4 Universality of color categories

There are several explanations for the universal tendencies in color naming and categorization. One is that color categories are universal because they are salient in the environment and our perceptual system is then predestined to attend to them [38]. Similarly it can be an evolutionary tuning for the properties of the daylight [31].

Interesting explanation was provided by Jameson and D'Andrade [16], who propose that a possible explanation for color naming universalities is that the developmental order of color names is due to the irregular shape of the color space. They illustrated the irregularity of the perceptual color space (described in 2.4.2) using the Munsell color system, depicted on figure 4.6, as the best example. In this color space there are areas, where hue interacts with saturation and lightness producing large bumps. These are located at focal yellow and focal red. The entire blue-green area is depressed (of low chroma), as is the area below focal yellow. These areas are more salient and therefore form best examples of color categories. They also assume that a property of names assigned to the color space at any stage is that they have to be most informative about color.

In case of two color terms, the most informative categories will be dark/cool versus light/warm. In correspondence with framework described in 4.4.2, the region of color space that is most distant from the regions specified by these

[6] Unfortunately this term refers not only to a composite green and blue category, but also stands for a color which is a mixture of green and blue. The latter case is severe, but still present in some WCS languages.

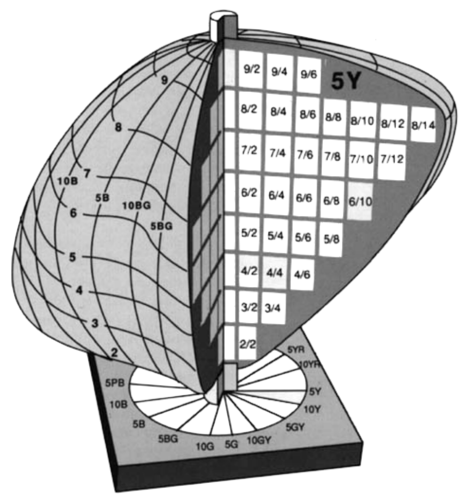


Figure 4.6: Munsell Color System

two terms is red. Further, after three terms specified it becomes more difficult to determine which is the next most distant region, because the differences in distances are smaller and depend in part on how the focal areas are determined. Expected is either yellow or blue to be the next, followed by green, purple, pink, orange, brown, and gray. Figure 4.6 depicts a diagrammatic representation of the Munsell color solid with one quarter removed. The numbers displayed in boxes represent individual color samples – with various hue, value and chroma. The circle on the bottom displays various hues, the value (lightness) increases from bottom (black) to top white and the chroma (saturation) increases from center to the side. Figure adapted from [15].

The last important factor regarding the universality is the social interaction. In their computational model of the formation of color categories, Steels and Belpaeme [35] and Belpaeme and Bleys [3, 4] showed that in an artificial system that uses perceptual color space, color categories qualitatively similar to those of WCS can emerge. This model, simulations with it and their results will be described in more detail in section 5.4.

4.5 World Color Survey Visualization

In this section we present visualization techniques we developed for better understanding of the WCS data and the results of simulations in the next chapter. Inspired by the old visualization based on typographical symbols described in section 4.4.1 the ambition of our conception is to overcome the limitations of the old one by employing more dimensions to display more information and to facilitate the imagination of the original data stimulus material. We project each from the WCS languages onto the Munsell color grid in three various complementary ways called *classic*, *reliability* and *fuzzy* visualization. We implemented our visualization methods in a separate software utility for visualization of the World Color Survey data as a part of master thesis *Basic Color Categories* [26], which is predecessor of this thesis.

Here we will at first define some specific terms that will be used in the whole section, describe the three types of visualization, and the program for generating these visualizations in form of JPEG images. We will present an illustration of a confrontation study of selected outputs with the conceptual framework from section 4.4.2 and finally discuss the possible usage of this visualization, its advantages and drawbacks.

4.5.1 Terminology

First, it is important to note that we will interchange freely the terms color term and color category, despite the fact that they do not mean the same thing. Generally, the color term is the name of the color category. Unlike color terms, color categories can be linked with many terms and consequently, universal. We are aware of the distinction between them, but it makes in no semantic difference in this case.

A color category will be most of the time described only by its serial (identification) number, because the lexical name in the case of the WCS languages plays no role for us. For simplicity we will shorten the serial number of the category into *category number* or just *category*. Additionally the word color chip refers to one color sample from the WCS experiment placed in one field of the Munsell color grid.

A *winning category* for a certain color sample stands for the color term, which was elicited for a certain color sample most of the times. Similarly to winner-take-all algorithm in this case the winner may be one of several

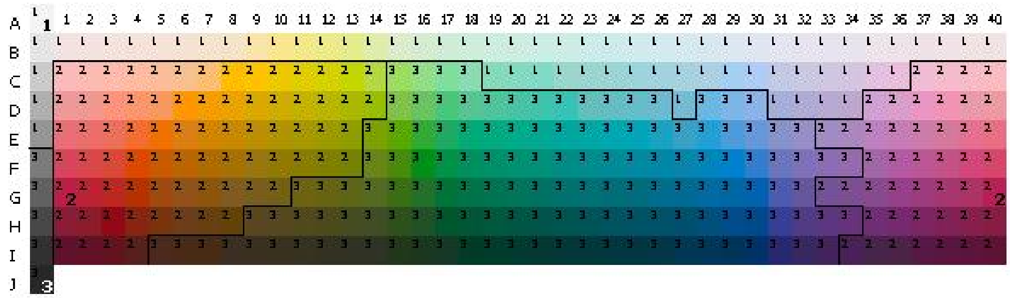


Figure 4.7: The visualization map for Wobé

frequently used terms with advance of small percentage, so it might not be exactly a dominating color term of that color sample. Still, this method has already been used in the first visualization of the WCS data described in 4.4.1 and has no equivalently simple and strong counterpart. A *focal chip* is the color sample selected as the best example of a certain category. Respectively a *focal category* for one chip stands for the serial number of the particular category of which the chip was selected as focal.

By *reliability* we mean the percentage of speakers that contributed to the winning category, i.e. how many speakers of all elicited the winning color term for a certain chip. This can be also called inter-speaker agreement. Average reliability is the average of reliability values from all 330 color samples. The Munsell color grid, as mention before, stands for the full stimulus board described in 4.2.1.

4.5.2 Classic visualization

This type of visualization, depicted on figure 4.7, is the simplest one. It projects the winning category numbers on the Munsell color grid with original colors used in the color naming experiment. The black lines around color chips represent the borderlines between categories. A small number in the left-top corner of each field represents the winning category number. The larger number in the right-bottom corner of some chips stands for the focal category, i.e. labels the best example chip with the number of category for which it was considered the best example. The size of the font of this number represents the agreement of speakers. We include this focal category numbers only in fields with at least 10% agreement of speakers.

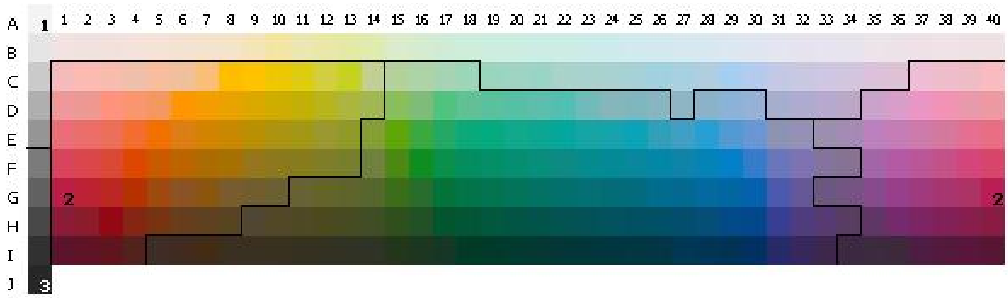


Figure 4.8: The reliability visualization for Wobé

4.5.3 Reliability visualization

In this type of visualization the color of the chips of color grid depends on how many speakers contributed to the winning category of that chip. The weaker agreement, the lower is the saturation of a chip. The new color of each sample is counted as the proportional mixture of its original color and gray color from the first column in the same row, to preserve the lightness of the sample. For example when only 50% of speakers agreed upon calling a sample s with the winning term t, the color of this chip will be half the original color and half gray. One drawback of this visualization principle is that there is no possibility to desaturate the shades of gray, so the leftmost column does not show the reliability of its category.

Two examples, figures 4.8 and 4.9 display Wobé as a language with one the highest average inter-speaker agreement from all WCS languages (89%) and the language with the lowest average inter-speaker agreement, Tifal (42%). Note that in this case we omitted the numbers of winning categories, just for simplicity; we kept the focal choices and borderlines. In the case without category numbers, but focal numbers present, we still include category numbers in separated chips, i.e. those which are not surrounded by any samples of the same category.

4.5.4 Fuzzy visualization

The most interesting and innovative is the fuzzy visualization. In this case the color chips of the grid no longer represent the original samples, but visually encode the results of color naming task and in certain sense also the

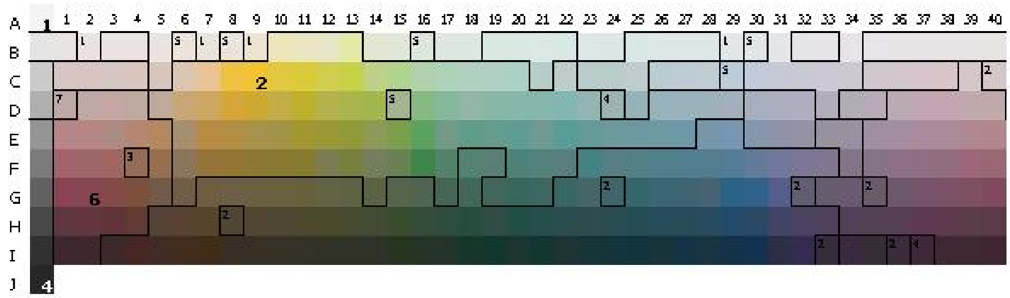


Figure 4.9: The reliability visualization for Tifal

foci task. Color for each chip is computed as a weighted mean of elicited categories represented by the average focal color, i.e. the mean of colors of all focal chips.

$$c = \sum_{i=1}^{N} r_i f_i, \tag{4.1}$$

where c is the resulting color[7], N is the number of color terms in language, r_i represents the portion of speakers responding with the category i and f_i stands for the averaged focal color counted for the category i globally from all samples labeled as best examples of it, in a similar proportional way.

For continuity we present the fuzzy visualization for Wobé (figure 4.10). Note an interesting phenomenon – the focal colors are consistent, but consist only of 3 primaries, even if the winning term area for these categories, especially number 3 (dark/cool color) covers also blue, green and other hues. This accounts for the stage theory of color term systems. The second illustration, figure 4.11, is Chavacano, a language from Philippines with relatively high average inter-speaker agreement (73%) represents (according to Berlin and Kay) a fully developed language with 11 winning categories. Note that distribution of these categories is very similar to those from Slovak language (see 4.3).

An interesting property of this visualization is that it displays distribution of all color categories on the grid and with fuzzy borders, since chips, which are on the edges of categories or with low reliability gain a more neutral

[7]This is the abstraction independent of color space in which the color is encoded. The parameter c can either represent a triple of RGB or L*a*b values, or other coordinates.

color[8]. On the other hand, the colors that are placed over the grid are highly related to the foci task responses, so they are prone to the errors and imperfections of the experimental data. If data contained no focal responses, prototypes of color categories were counted from the most salient chips (with high inter-speaker agreement).

It is important to note that the colors on the grid are not equivalent neither to the real best examples of the categories from the foci task (since we used an average from all foci chips for a category) nor to the real percepts of the speakers interviewed. The strength of this visualization method is the simplicity and intuitiveness of display. Even if the colors on the maps do not resemble the real percepts, they show the real distribution and salience of color terms and their best examples[9].

A notable drawback of this visualization is that it does not respect the nature of composite categories. For example the category grue will not appear green or blue as the speakers would perceive or, more precisely, categorize in finer distinction, since they have named multiple foci for any composite categories. In a typical case the focal responses for grue are located in the "middle" of blue and in the "middle" of green samples (approximately around F16, G16 and F28, G28 on the grid). However, the visualization will display grue as a proportional mixture of foci, so it will produce something like teal or turquoise. On the other hand if the prototypical grue was really a green-blue color it will account for the hypothesis about the different hues of Hering primaries mentioned in 2.1. What the fuzzy visualization can show, is which of these composite sub-categories of basic terms (not yet emerged into two separate terms) is stronger. Here again, to emphasize the useful properties of this visualization, we will bring to front the intuitiveness. The colored areas of the map are intuitively comprehensible and comparable with each other and also with the real percepts of the observer, so it indirectly suggests the real number and quality of basic color categories in a language.

[8]More precisely a mixture of colors of categories on which borderline the chip is. For example a chip between red and white will be light pink.

[9]Note that in hypothetical case, if a certain color category had only one focal chip and the inter-speaker agreement was 100%, in the fuzzy visualization it would be displayed with the same color value (the same color hue) as in the Munsell color grid.

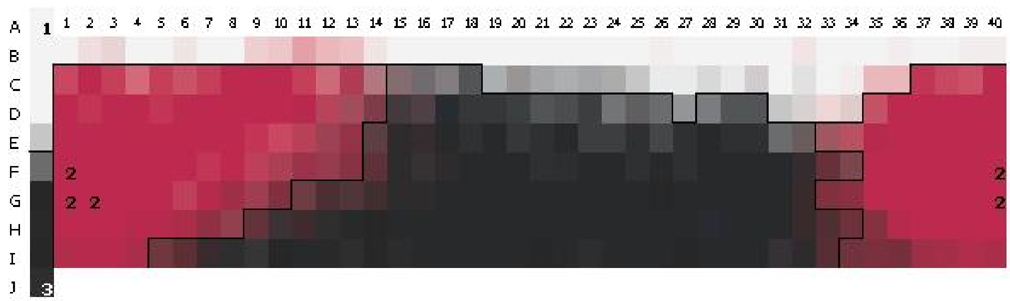


Figure 4.10: The fuzzy visualization for Wobé

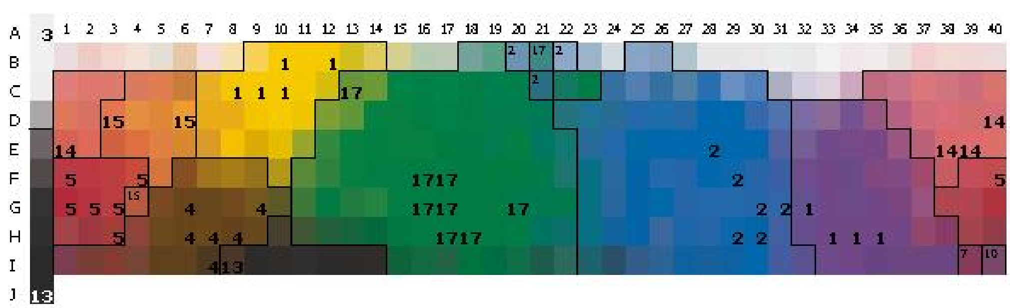


Figure 4.11: The fuzzy visualization for Chavacano

Chapter 5

Distinguishing criteria based model of color categorization

In this chapter we present a model of color categorization based on Rosch's prototype theory and on the semantics of distinguishing criteria. We tested this model in two situations; first we created simulated speakers of WCS languages as an example of supervised learning. In the next step we tested the distinguishing criteria in unsupervised learning task using multi-agent simulation mentioned in 4.4.4. The aim of this work is not to prove or disprove the original hypothesis of Berlin and Kay, but rather to probe the learning capability of the distinguishing criteria on the real data represented in a perceptually uniform space.

5.1 Distinguishing criteria

Distinguishing criteria were conceptually proposed by Šefránek [30] and implemented by Takáč [36]. Each distinguishing criterion (DC) functions as a locally tuned detector reacting to some part of its input space, and represents one category. It is able to distinguish whether or, more precisely, to what degree is the presented input a member of a category it represents. A core of each DC is its prototype – its best example, which is computed from the inputs it receives during learning (so it learns only from positive examples). The mechanism of evaluation of inputs after learning is called the *activation function* of the criterion and it computes the degree of membership using the distance of the presented input and the stored prototype.

Distinguishing criteria have these important properties:

1. **Learnability**: they can be incrementally and continually constructed from an incoming sequence of examples. Note that unlike classic artificial neural networks, the distinguishing criteria are able to function even with a smallest amount of presented examples.

2. **Identification**: each criterion can express for each given input its degree of membership in a category represented by given criterion. In other words, the distinguishing criterion can express if the given input is an instance of the concept, which it represents. The value of activity of the DC for an input is expressed by a real number from $[0, 1]$. The closer it is to 1, the more the input belongs to the category, if it is 1, the input is identical with the prototype of the category.

3. **Auto-associativity**: even if the input is noisy or incomplete, the DC returns the best example (prototype) of the category and completes the missing aspects of the input when it is needed for further evaluation.

The second property of the distinguishing criteria exists due to their character of locally tuned detectors, which can be intuitively represented by conceptual spaces [10]. A conceptual space is a something like a geometric space, but with dimensions corresponding to the attributes of represented objects. These attributes can be of various natures (metric, psychological, etc.) and are organized in domains. A particular object is represented as a point (a vector of coordinates) in a subspace of one or several domains. In our case of color categories we will have no difficulties with the accordance of the domains of compared objects, because inputs will be always represented as points in L*a*b* color space, which is only three-dimensional and has Euclidean metric.

It is important that the natural categories are represented by convex regions in the space. If two points represent objects that are good examples of a category, then any point between them must also be a good example of that category. Thus the best examples should be situated in the geometric centers of these regions. Figure 5.1 illustrates the receptive field (with the threshold of 0.1) of a 2-dimensional locally tuned detector, and was taken from [36].

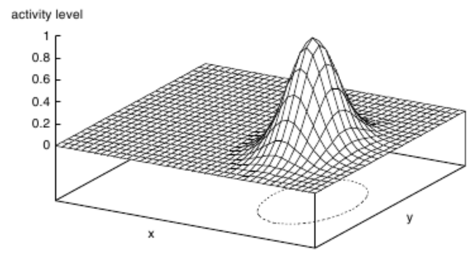


Figure 5.1: Illustration of a locally tuned detector

The implementation of the DC we use computes the prototype as an average of all inputs. Each color category c is represented by a locally tuned detector. The degree of membership of an input $\vec{x}$ presented to the DC in form of one point in L*a*b* color space (a three-dimensional vector) will be computed as an exponentially decaying function of the distance from the prototype $\vec{p}$:

$$r_{\vec{p}}(\vec{x}) = \exp\left(-k.d(\vec{p}, \vec{x})\right), \tag{5.1}$$

where k is some positive constant, and d is the metric used.

We used the covariance based distinguishing criteria, which track down the distribution of inputs in the covariance matrix sigma and compute the activation for an input using the Mahalanobis distance:

$$d^2_{\Sigma^{-1}}(\vec{p}, \vec{x}) = (\vec{x} - \vec{p})^T \Sigma^{-1} (\vec{x} - \vec{p}), \tag{5.2}$$

where $\vec{p}$ and $\vec{x}$ are column vectors and Σ^-1 is the inverse of the covariance matrix Σ of the sample set used for training. For more detailed description see [36].

5.2 Prototype based model of color categorization

Our model of color categorization draws on Rosch's prototype theory. The color categories are formed around their prototypes. They are embodied through distinguishing criteria operating on three-dimensional L*a*b* color space, which is known to well resemble human color perception system (see 2.4). In case of supervised learning with data from WCS, we introduce additional factor influencing learning and categorizing colors - the cultural factor.

Each distinguishing criterion representing one term in particular language is coupled with reliability parameter, which then plays role in evaluation of the activation function. In this chapter we provide two examples of usage of our model in color categorization tasks, supervised and unsupervised, setting and results of simulations and the discussion.

5.3 Supervised learning with the WCS data

We implemented the simulation in Java using the distinguishing criteria framework developed by Takáč [36]. For our purposes we processed the data from [8] to a suitable form. For each of the 110 WCS languages we created a simulated idealized speaker, who learned to categorize and name colors from all speakers of that language. We can imagine this process as if each speaker from particular language came to our simulated speaker and pointed to each of the 330 samples and named the color. The simulated speaker is then "examined" from all or some of the color samples and his repertoire of color categories compared to repertoire summarized from all the speakers.

In the process of learning our simulated speaker (agent) is firstly endowed with "blank" distinguishing criteria for all terms named in one language. During the training each DC receives color values of chips (color samples) named with the represented term. So each DC learns only from positive examples of the certain color category. The size of the whole dataset is approximately 330 x 25 = 8250 categorical responses, which, from the perspective of our agent, have a form: #: {L, a, b}, where # represents the serial number of the category assigned to a color sample by one speaker and L, a, b are values of the color coordinates of that sample.

Since the nature of DC allows them to produce quite good judgments only after few presentations from training set, and since the data in many cases include noisy values [1], the agents also keeps track about the frequency of use for each term. The higher this number is, the more significant is the term for the language. We will label this "frequency" factor $f_C(s)$ (frequency

[1] Probably not all experiments from the WCS were successful, possibly due to poor understanding of instructions. This phenomenon is supported by the fact, that in some languages the number of terms named and terms that were agreed upon by majority of speakers differ significantly. Largest difference is in language Mampruli, 79 named to 8 winning terms.

of C for sample s[2]) and use it in evaluation of the *winning categories* for test samples.

5.3.1 Methods of evaluation

We provide two main methods of evaluation inspired by the preliminary and general methods of processing of the WCS data (see 4.4). In this section we will use terminology from 4.5.

Winning categories

First we evaluated the *winning categories*. For each of the 330 samples we computed the winning category from the existing data, it is the serial number of the term, which was elicited for the particular sample by the largest number of speakers, as in the winner-take-all mechanism.

Evaluating the responses of the agent we compared the activity of the distinguishing criteria multiplied by the reliability factor $f_C(s)$. The response of each color category (term) C to color sample s, which has color value $\vec{x}$ then is:

$$C(\vec{x}) = r_{\vec{p}}(\vec{x}).f_C(s), \tag{5.3}$$

As in the case of data evaluation, we computed the winning categories from the simulated agent in a similar winner-take-all fashion. Finally we compared the winners of the agent and of all speakers of the language, chip by chip, and in case of match we added one, otherwise zero. The output for each language was an average of these matches and non-matches for all samples.

Vectors of activity

Reflecting the objection against generalization nature of winning categories evaluation we introduced a more thorough way of comparison, namely using vectors of activity. An activity vector can represent the distribution of answers of subjects from the experiment as well the distribution of activations of the trained distinguishing criteria.

For each of the 330 samples we first created a vector of numbers of speakers that used each term from the language's vocabulary, and another vector

[2] s is a number from $[1, 330]$ and represents the serial number of sample presented

of responses of color categories of the agent. We normed both vectors using the Euclidean norm[3], and then compared them as two unit vectors in n-dimensional hyper sphere (where n represents the number of elicited categories). In two-dimensional case it would be two unit vectors on the unit circle.

For the comparison of the two vectors we used the scalar product (dot product), which directly corresponds to the cosine of the angle between these two vectors (since we have unit vectors, their size will not count). If the scalar product is 1, they are identical; if it is -1, they are opposite to each other. For the summary comparison with other languages we recorded the average value of scalar products of normed activity vectors for all samples.

Dispersion

The last method for evaluation we took from [18]. Namely we counted the dispersion, in pursue of evaluation of clusterization, among WCS data and our simulated agents from all languages. If the representations of color terms in WCS languages tend to cluster together, then the areas, to which these representations from all languages fall can be considered the universal basic color categories (for detailed explanation see previous chapter, especially section 4.4.3).

For each color term c from each language we computed its geometric center as one point in CIEL*a*b* space. These centroids were first computed from all responses of each speaker from a particular language. Then centroids for all terms elicited in that language were computed from the centroids of its speakers. We compared each color term centroid c in each language l with all other categories in every other language l^*, and found the centroid c^* with the shortest Euclidean distance to c. We computed the dispersion D as a sum of these shortest distances between color centroids (equation from [18]):

$$D = \sum_{l,l^* \in WCS} \sum_{c \in l} \min_{c^* \in l^*} distance(c, c^*) \tag{5.4}$$

Subsequently we used the same mechanism for evaluating the dispersion of color categories of all our simulated agents (each agent “speaking” a different

[3]Each component is divided by the size of the vector, which is the square root of the sum of squares of each component

Table 5.1: Evaluation of winning categories

k	mean	maximum	minimum
–	89.854%	97.879%	54.545%
2	89.344%	97.576%	54.242%
3	89.518%	97.576%	54.545%
5	89.532%	97.273%	54.545%

WCS language). With use of established fact from [18], that data from WCS cluster together do a degree greater than chance we expected that the dispersion among the lower number of "individuals" (simulated agents) will be smaller than for all subjects from the WCS.

5.3.2 Results and discussion

For sufficient extent of testing we used in both methods of evaluation in addition to the whole dataset also the k-folded cross-validation with different values of k, namely $k = [2, 3, 5]$. We divided the dataset by chips into k sets and ran k simulations always using one set for testing and the others for training. Since divided the datasets by particular chips, not by speakers, the simulated agent was always trained on some samples and then presented with new color hues during testing. When computing the overall results we first evaluated simulations for each language alone and then computed means throughout all languages for different values of k.

The results (averages from all WCS languages) are displayed in tables 5.1 and 5.2. High percents of agreement (around 90%) in winning categories and cosines of the activity vectors very close 1 (around 0.96), point to the success of the simulation. Especially the small difference in results from different values k in cross-validation show that the model is in this case stable, successful and robust. Lastly, the dispersion was, as we expected, significantly smaller than dispersion of the WCS data (around 2-3 times). Following figures depicts color centroids for WCS data (fig. 5.2) and simulation results (fig. 5.3) plotted on the WCS grid (stretched to enable a 3D view). Our results proof that the distinguishing criteria are able to successfully generalize the input data.

Table 5.2: Evaluation of activity vectors

k	mean	maximum	minimum
–	0.96700	0.98965	0.78902
2	0.96389	0.98761	0.78451
3	0.96525	0.98877	0.78620
5	0.96522	0.98911	0.78517

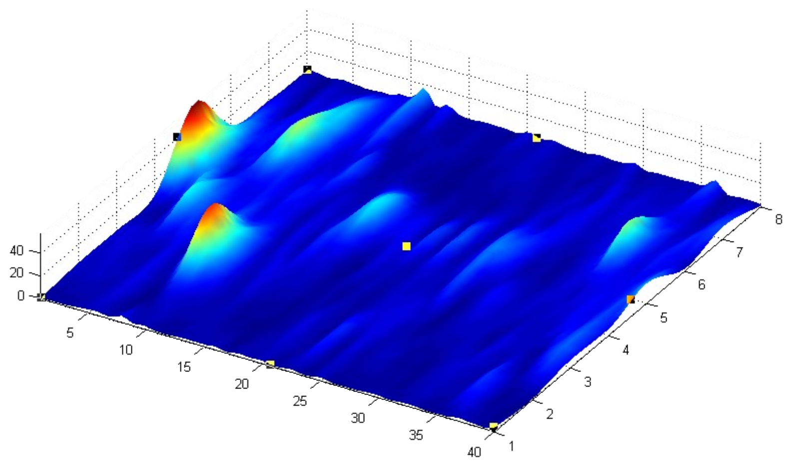


Figure 5.2: A histogram of color centroids computed from WCS data

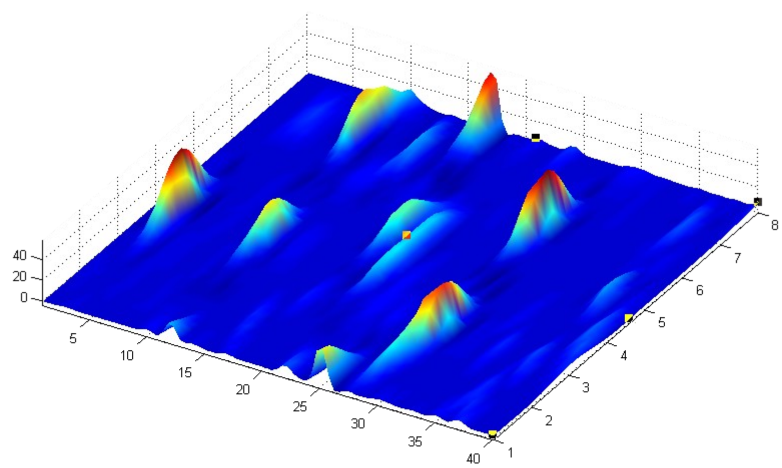


Figure 5.3: A histogram of color centroids computed from simulated agents

5.3.3 Results visually

To provide better overview on the results from the simulations we dedicated this section to visual depiction of selected languages and their simulated speakers. We used our visualization methods from 4.5, specifically the fuzzy visualization, which displays not only borders around categories (evaluated on the basis of winning categories) and the best examples, but also projects the color naming results in form of weighted color mixtures. Since the results were quite similar both in cases with and without cross-validation, we used results from simulations where agents were trained and tested on the whole data set.

The fuzzy visualizations from the data were created according to algorithm described in 4.5. In the case of visualizing the simulation results the prototypes were taken directly from the distinguishing criteria, unlike the case of visualization created from the data (the upper map), where the best examples were taken from the second task of the experiment (see 4.2.2). The color of each chip is a sum of the color values of prototypes of all categories (points in L*a*b* space), each multiplied by the normed value of activation of the category for the given chip[4]. The black borders represent boundaries between winning categories for the chips. The images contain two maps, the upper one created from the data and the second one from the simulation results. Note the little black numbers in the first image, which represent the serial number of categories and the chips labeled with them the best examples elicited by the subjects of the experiment (evaluated in winner-take-all-like fashion).

The examples we selected are depicted on figures 5.4, 5.5, 5.6, and 5.7. Figure 5.4 depicts our already well known example, Wobé. As we already mentioned, Wobé is a high consensus language (high percentage of inter-speaker agreement) with only three basic color terms and six elicited terms. The simulation data from this language produce a slightly different image, with green, orange and something grayish, in comparison to visualization of the real data, where only red, white and black can be found. This color shift emerges due to the learning mechanism of the distinguishing criteria. Since the prototype is computed from all examples presented to a DC, the wide category containing not only black and brown, but also bluish and greenish

[4] Similarly to fuzzy visualization from data in 4.5, the color of a given chip is a mixture proportional to the activations of distinguishing criteria, if the activation is high, it will contribute significantly to the resulting color.

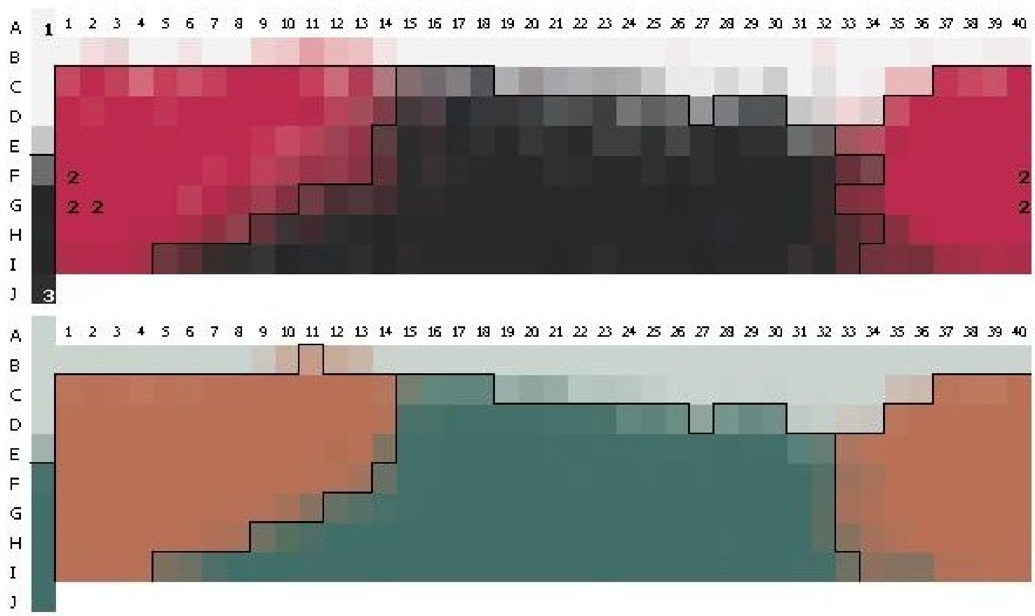


Figure 5.4: The fuzzy visualization of data and simulation results for Wobé

hues will gain a blend of these colors, which is not similar to best examples of the category chosen by the subjects in the experiment. However, the 97.88% match of the winning terms and cosine of activity vectors of 0.99 confirms, that the categorization ability of the DC remains intact. In the next example, figure 5.5 we present a low consensus language, Maring from Papua New Guinea with inter-speaker agreement around 49%, 33 elicited and 7 winning terms. Note an interesting phenomenon produced by the simulation: there is a significantly lower amount of chips located inside some categories, but belonging to other categories. From this point of view we can expect, that our model does not only fit the original data, but also "denoises" them to some extend.

The latter two examples are presented in relation to the hypotheses of Berlin and Kay from previous chapter. The first one, Ticuna (Peru) on figure 5.6 is a language, in which green and blue hues are named together with one term. Since the prototypes are both from green and blue hues, this term names a composite category (we call this color grue). In this case the winning DC acquires a blend of green and blue as a prototype, but is not able to model the composedness of the category. This language is also a case

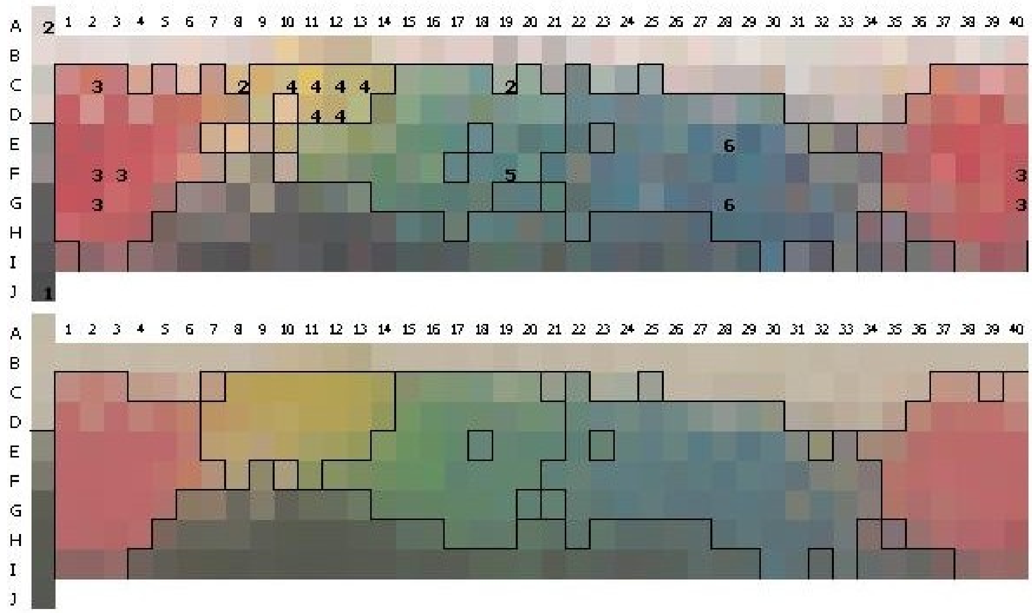


Figure 5.5: The fuzzy visualization of data and simulation results for Maring

in agreement with the evolutionary framework by Kay and colleagues [17] (see 4.4.2). Nevertheless, there are cases that do not fall into any schemes in this framework. One of such languages is Chayahuita (also from Peru) on figure 5.7, which also has an overall term that covers what we call green and blue, but has best examples only in blue. This is the case where label grue does not fit for the category. Chayahuita violates the proposed framework also in other aspects, for instance from the visualization we can clearly see that the speakers of this language already distinguish pink from red, which should not happen before the division of green and blue.

5.3.4 Slovak language simulation

The last experiment with supervised learning were the simulations with the data from the Slovak experiment 4.3. Since we collected two sets of data, with and without the six *outer* color categories[5], we created two different simulations.

[5]The colors which were successfully recognized among significant amount of speakers other than 11 basic categories proposed by Berlin and Kay, more in 4.3

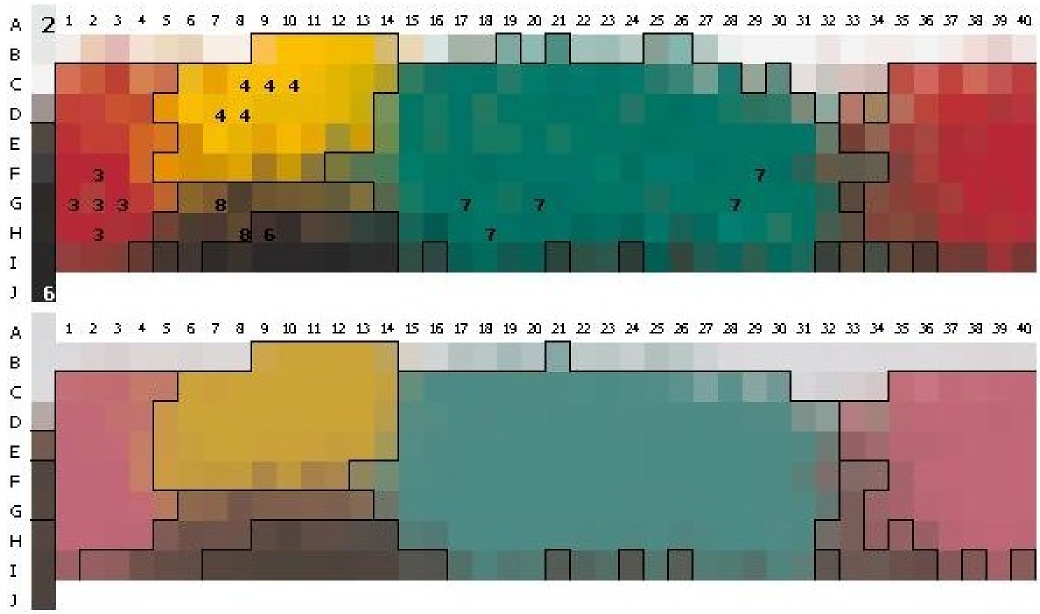


Figure 5.6: The fuzzy visualization of data and simulation results for Ticuna

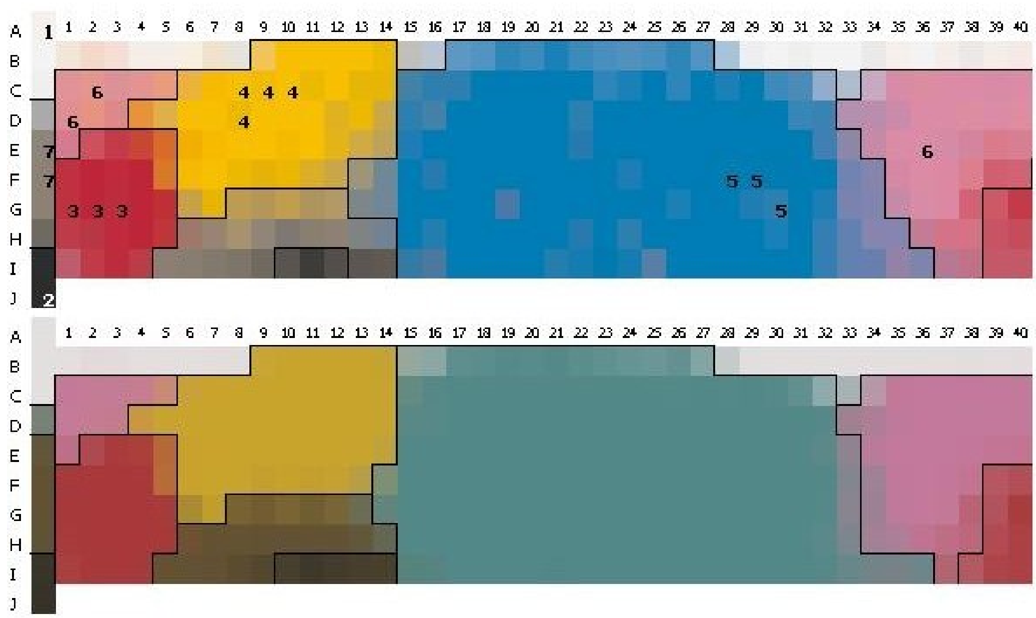


Figure 5.7: The fuzzy visualization of data and simulation results for Chayahuita

The design was the same as in simulations with the WCS data, also introducing the reliability factor. The results were quite similar to the WCS results. The winning categories matched at 96.06% for data containing only 11 categories and at 90.00%, when the outer categories were included. Cosines of activity vectors were more similar, 0.985 with 11 and 0.974 with all 14 categories. The relatively big difference in winning categories results can be explained, but also explain the nature of the data. Since the outer categories were not used by all subjects in the experiment[6], their presence in the data introduces a noise similar to noise in the WCS data (the elicited terms that were not winning for any samples). Since our model generates the while repertoire of "blank" categories according to all elicited terms, we can conclude that the inter-speaker agreement and the number of elicited terms influence the performance of our model. The results of these simulations are displayed on figure 5.8 (11 categories) and figure 5.9 (outer categories). These figures were generated in the same way as figures in 5.3.3.

[6]Only one of the outer categories can be considered to be added to 11 basic color terms. It is the turquoise located between green and blue. The other two gained in the winner-take-all competition only one or zero chips. This turquoise dominance can be explained as a culturally valid contribution to basic color terms, but also can be caused by the properties of the printed samples, which have hues only similar, but not identical to those from WCS survey

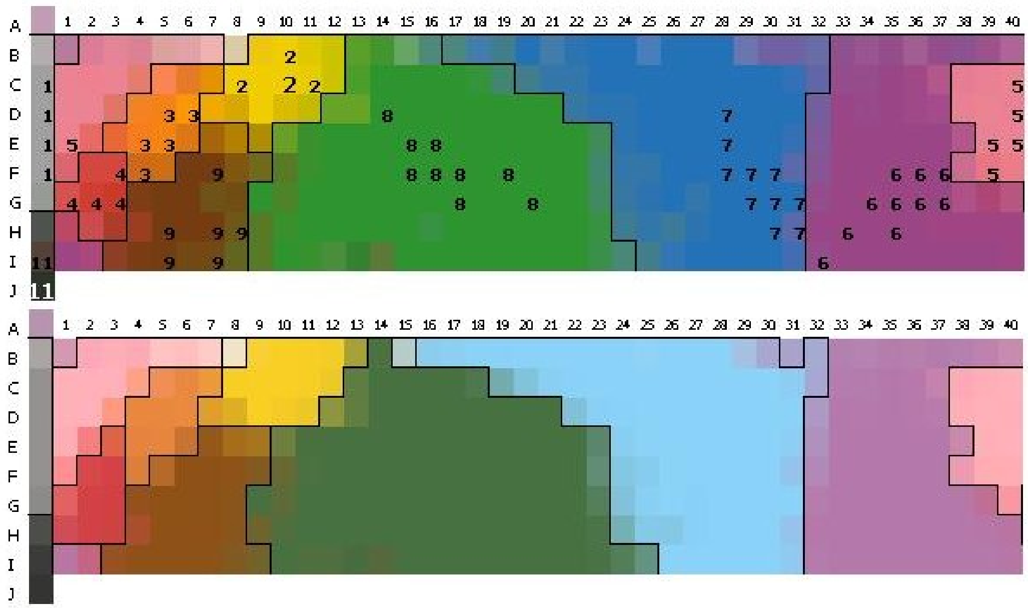


Figure 5.8: Fuzzy visualization of results for Slovak (11 terms)

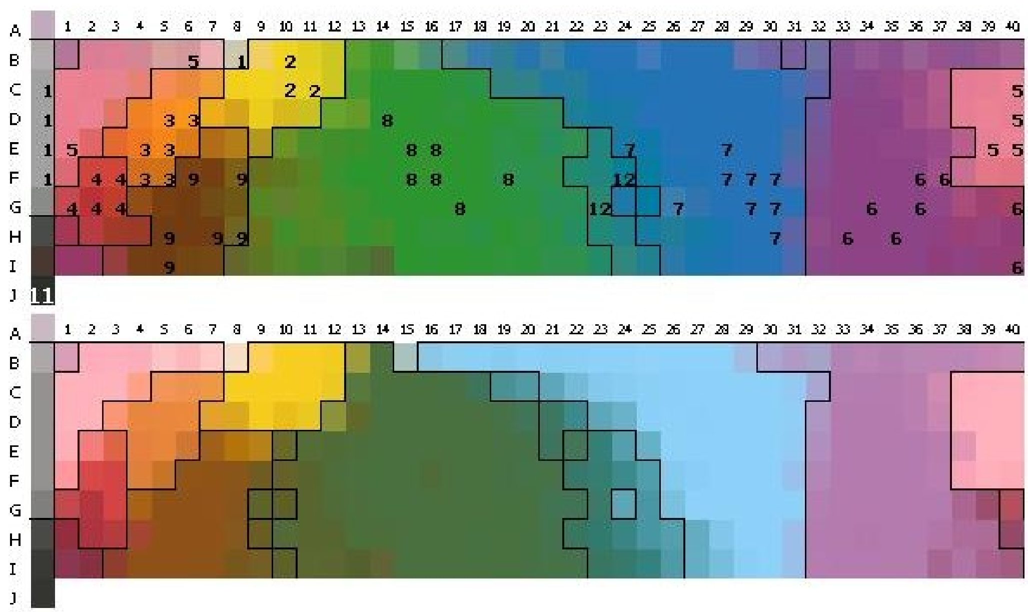


Figure 5.9: Fuzzy visualization of results for Slovak (all terms)

5.4 Simulations with unsupervised learning

In this section we verify the abilities of our distinguishing criteria based model in simulations with unsupervised learning, namely with the semantics of language games developed by Steels [33] and implemented by Belpaeme [2], resulting into influential research in field of color categorization [35]. As we mentioned in 4.4.4 Belpaeme and Bleys [3, 4] showed, using the semantics of language games, that a group of simulated agents able to discriminate colors and communicate will develop a shared lexicon of color names representing basic color categories qualitatively similar to basic color categories from the WCS. We will first briefly describe the language games, then the specific simulations and results and lastly our simulations with the games and their results. We will compare the original mechanism for representing the color categories, the *adaptive networks* with our model based on the distinguishing criteria.

5.4.1 Language games

The language games used to model the emergence of color categories in simulated agents were originally developed by Steels [33] and implemented in the famous Talking head experiment and other examples (for conclusive overview see [34]). In our case of color naming there are two games, the discrimination game and the guessing game, of which only the latter is actually a true language game. The purpose of the discrimination game is to gain the ability to discriminate colors and form the repertoire of the internal representations of different color categories sufficient enough to distinguish one color from other, notably different colors[7]. The guessing game, which implies having the ability to discriminate, is the mechanism responsible for consolidation of color vocabulary among agents.

The discrimination game

The discrimination game is played by each agent alone. In the course of the game an agent first views a small set of color values, which must be sufficiently distant from each other in the color space used within the agent's perceptual

[7]There must be some minimal distance between their point representations in perceptual color space, so they would present perceptually differentiable color hues

system. This set is called the context and is randomly chosen from a larger set of color values (training data). The agent randomly chooses one sample from the context, the topic, and tries to discriminate it from the context. First, the agent confronts the topic and the remainder of the context with its internal repertoire of color categories labeled with serial numbers. The winning category, chosen similarly to our simulation in a winner-take-all-like mechanism is chosen for each sample.

The topic is successfully discriminated from the context only when the serial number of its winning category is different than other categories identified for the remainder of the context. If the game fails, in most cases the agent adds a new color category to its repertoire. In the course of the simulation an agent starts as a tabula rasa, the only way it acquires new meanings (categories) is through the discrimination game. This is why we call this method of learning unsupervised. The discrimination game is schematically depicted on figure 5.10.

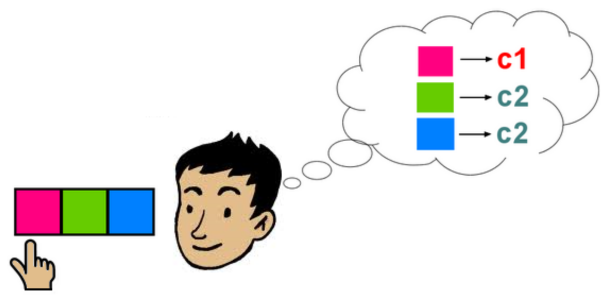


Figure 5.10: Illustration of a successful discrimination game

The guessing game

This game is more complicated than the first one and involves two agents. It implies that agents first play discrimination games until they are successful enough in perceptual categorization to begin the interaction. In the beginning of each guessing game the two players are randomly chosen from the population and assigned roles of the *speaker* and the *hearer*. The first move belongs to the agent in the role of speaker, which chooses the topic and tries to discriminate it from the context. If the topic is identified successfully, the speaker searches for a term with the strongest association to the winning category and utters it. If there is none a random new term is generated, assigned to the category and uttered. The hearer receives the word and looks it up in its vocabulary. If the term is present, the hearer assigns categories to all samples in the context. If there is a sample, that can be uniquely identified by the category represented by the processed word, the hearer "points" to it. If the hearer points to the same category the speaker "has had in mind", the guessing game is successful and the speaker's vocabulary reinforced.

The guessing game can fail in many aspects. Both the speaker and the hearer can fail in the discrimination game. If this failure happens to the speaker, the game is terminated immediately and only the speaker learns the new category, but his vocabulary and the mapping to the categories remain intact. Similarly, no learning of words occurs if the hearer fails to discriminate. If the color sample chosen by the hearer is different from the speaker's topic, the hearer's lexicon (the association between term and color category) is shifted towards the speaker's lexicon. The guessing game is schematically depicted on figure 5.11.

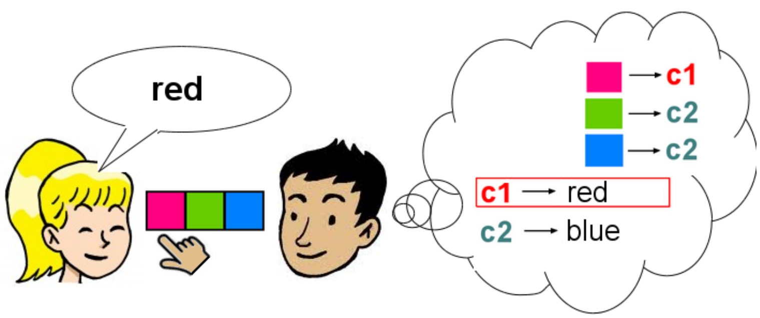


Figure 5.11: Illustration of a successful guessing game

5.4.2 Color naming models based on language games

The structure of an agent in these models is very similar to our model. Each agent is endowed with perception in CIEL*a*b* color space (which is perceptually uniform), has an internal repertoire of color categories, which is private to the agent, the lexicon of words and mapping between colors and words, and the ability to play the discrimination and the guessing game. The mapping between categories and words can be of many-to-many type to allow synonymy and homonymy during the learning.

In [2] and [35] color categories are represented using the so called *adaptive networks*. An adaptive network is a classic feed-forward neural network built from RBF (radial basis function) neurons. Interestingly, these neurons do not adapt, they are only initialized to a particular prototype (color value). The RBF neurons have a fixed width[8] of the regional extent around the prototype determined by a normalized Gaussian function (the activation function of the network). The only parameters adapted during the learning are the weights from the RBF neurons to the summing unit. The activation of one adaptive network (equation 5.7) is computed as a sum of activations of the RBF units (equation 5.5) multiplied by their weights (from $[0, 1]$).

$$z_j(\vec{x}) = e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i - m_{ji}}{\sigma_i}\right)^2}, \tag{5.5}$$

where z_j is the activation of RBF unit z for the input $\vec{x}$ and m is the prototype of this unit and σ is the width of the Gaussian.

$$y_k(\vec{x}) = \sum w_j z_j(\vec{x}), \tag{5.6}$$

where y_k is the activation of the adaptive network k for the input $\vec{x}$ and $z_j s$ are the activations of the RBF neurons in this adaptive network.

In the second version of these simulations, Belpaeme and Bleys [3, 4] use much simpler representations. One color category is embodied by a single point in color space (a prototype) and the activation computed from the distance of the input from the prototype.

In the experiments with these culturally based models were executed with different data sets. The first simulations with adaptive networks [2, 35] used a dataset of color values from Munsell color system (which is proprietary), and unfortunately was not freely available. On the other hand, data from latter two studies [3, 4] are freely available online and were used also in our simulations. They contain two different sets, with ten thousands color values each, extracted from digital photographs of either urban scenes or nature. Populations of agents typically contained 10 agents that played thousands of games (according to the task and environment constrains). Regarding the learning, the first design included both cultural and genetic evolution of color categories. The latter cases were aimed on cultural evolution and comparison with color categories in the real world. We describe the learning mechanism in more detail in the next section.

In conclusion, Steels and Belpaeme [35] have shown that the coupling of category formation with verbal communication leads to the coordination of

[8] This width, σ, is usually set to 10.

perceptually grounded categories, even if there is no statistical structure in the data (cf. section 4.4.4). In addition Belpaeme and Bleys [3] conclude that the cultural acquisition of color categories together with mild constraints on the perception and categorical representation result in categories that have a distribution similar to human color categories.

5.4.3 Language games simulations with adaptive networks and distinguishing criteria

In this section we present a comparison of the performance of the distinguishing criteria with the adaptive networks (AN). To gain most conclusive results, we first implemented the model from [2] and [35], and then adapted it for the distinguishing criteria. First we implemented the discrimination game with both representations and compared the results. In the next step we implemented also the guessing game, unfortunately without notable success (see below).

Implementing the model with distinguishing criteria we simply used the same mechanism for discrimination game and replaced the adaptive networks with the distinguishing criteria. However, we had to modify the learning mechanism, to fit the properties of the distinguishing criteria. In the original model, learning depends on the outcome of the game. When the discrimination game is successful, the weights of the AN that identified the topic are updated according to the following equation:

$$\Delta w_i = \alpha z_i(x), \tag{5.7}$$

where Δw_i is the contribution to the weight, α is the learning rate (usually set to 1.0), and $z_i(x)$ is the activation of the AN for the input x.

If the game fails, there are two cases of learning. If the cumulative discriminative success of an agent is above a predefined threshold (usually 95%), only a new locally reactive unit is added to the adaptive net representing the best matching (winning) category for the topic. Otherwise the learning is much more progressive: if the discriminative success is below the threshold, a new AN is added to the agent's repertoire with the color value of the topic as the prototype.

The same principles apply for the DC. If the game fails, a new blank DC is added and the topic is presented to it. In this moment the criterion's prototype is the color value of the topic. Unlike the AN, the DC consist of only one locally tuned detector, so the learning must be different in the "above threshold scenario". Since the occurrence of such cases is limited to the end of the game or none (see results in the following paragraph), we decided to apply the same mechanism for this settling phase as in case success. If the game is successful, the winning DC is updated with the value of the topic (for more details on learning mechanism of the DC see Appendix C in [36]).

To stop an uncontrolled growth of the color category repertoire a deleting mechanism must be implemented. In case of the AN, it is implemented as a continual small decay of the weights. If all weights in an AN are smaller than a predefined threshold, the AN is removed. To produce similar effect with the DC we evaluated their usage, more precisely, the number of games in which they were not used, and according to this value and an arbitrary threshold we deleted the unused criteria.

For evaluation of the discrimination games we adopted measures from [2, 35], specifically the average discrimination success of the population of agents. The success of one agent is expressed as a moving average of results of discrimination games it played, with size of the window typically of 20. In other words, it is a percentage of successful cases to failures.

Simulations with discrimination games

We executed 100 simulations with sole discrimination games for both adaptive networks and distinguishing criteria implementations. In each simulation a populations of 10 agents played 1000 games (as in [2]). In each game one of the agents was chosen randomly and played the discrimination game alone. We used the dataset with 10000 color values from nature photographs[9]. The size of the context was set to 3. Figure 5.12 show the average discriminative success for an example simulation with AN, in black, and one with DC, in red. The horizontal axis displays simulation numbers and the vertical the average discriminating success.

The average discrimination success was for the AN simulations 95,095% and for the DC simulations 92,539%. These results are quite close to results from the original studies. Taken into consideration the difference in sizes of the datasets used in our and original study, the slightly worse results (not reaching 100%) are acceptable. We conclude that the categorization ability

[9] In this aspect our simulations differ with the original study, where values from Munsell color system were used.

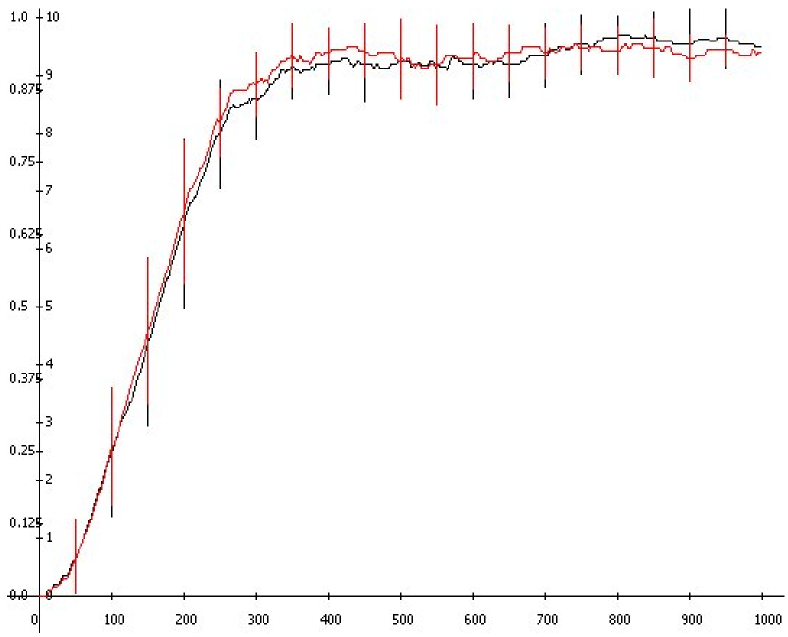


Figure 5.12: The discriminative success in simulations with AN (black) and DC (red)

of the distinguishing criteria is comparable to the ability of the adaptive networks, and that they are suitable for tasks requiring unsupervised learning.

Simulations with guessing games

Regarding the guessing games, we implemented the model according to specification in [2] and ran simulations with 10 agents, 5000 games and the nature data. Unfortunately even the AN simulation results were quite different from the original results. They differed not only in the final amounts of internal color categories in agents, in average amount of words in the shared lexicons, and communicative success of the agents, but also in the course of the simulations. Unlike we expected, the discriminative success did not rise gradually with the course of the simulation, but culminated around 85%. It seems that it was caused by some kind of overlearning.

These discrepancies might have occurred due to incompleteness of our implementation. Since the guessing game is more complex than the discrimination game, there might be some specific parameters not included in the description of the model, for instance a stopping criterion for learning of color categories, or some new parameters in the guessing game, etc. The performance of the distinguishing criteria was in this case similar to the performance of the adaptive networks. However, the failure to replicate the simulation correctly prevents us from drawing any significant conclusions from this part of our research.

Results of discrimination games visually

In this section we present visual examples of the categories learned by the agents in the discrimination game. The motivation is, similarly to section 5.3.3, to illustrate simulated perceptual categories of the agents and their similarity to color categories (labeled by basic color terms) in languages of the WCS. To gain visualization comparable to visualization of the WCS data we first created two groups of artificial agents trained in discrimination games, as described in the previous section. After training, we presented the agents with color samples from the WCS and recorded the winning categories for each sample. Then we created the fuzzy visualization of the results (see 4.5).

Figures 5.13 and 5.14 display selected agents of both the AN and the DC type. Note that the color categories are slightly different from the human color categories. Especially some of the basic opponent hues (red, green, blue, yellow, black, and white, see 2.1) are missing, what is not in accordance with the evolutionary framework described in 4.4.2. This is probably caused by the composition of hues the agents encounter. The dominance of these primaries in human color categorization might be as well caused by the environment and properties of the daylight as we mentioned in 4.4.4.

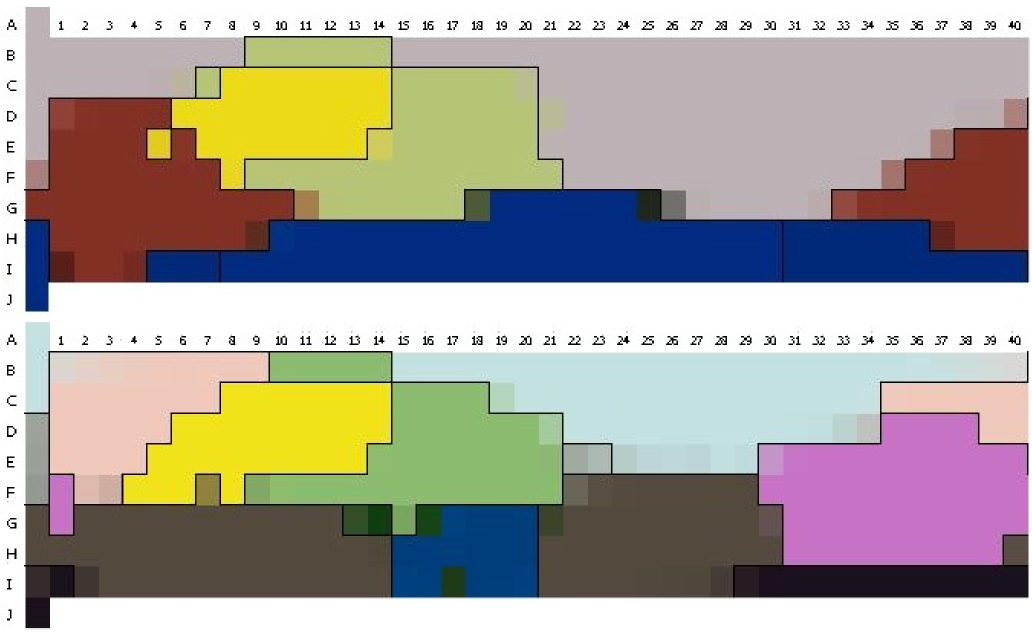


Figure 5.13: Example of two AN agents categorizing WCS samples

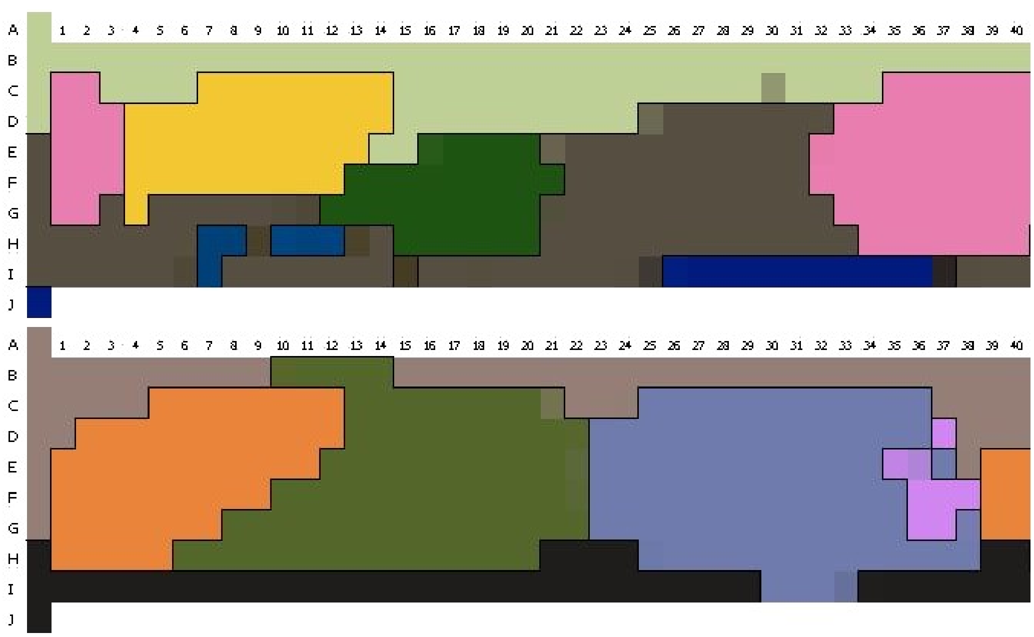


Figure 5.14: Example of two DC agents categorizing WCS samples

Chapter 6

Conclusion

In the first part of this thesis we provided an essential theoretical background for understanding the phenomena involved in color categorization – color perception and categorization in general. We presented a definition of basic color terms, which are labels for the basic color categories. They constitute a small set of simple words with which all speakers of a language can consistently name any color. Consequently we described the main hypotheses about them, which state that each language has at least 2 and at most 11 basic color terms and that the perceptual categories connected with these terms are universal and correspond to 11 basic color terms of English (black, white, red, green, yellow, blue, purple, pink, brown, orange, and gray). The first six of these terms are so-called Hering primaries that are involved in opponent processes underlying the color perception and are expected to emerge first in any language.

Various clustering analyzes of the data from the World Color Survey confirmed that the perceptual correlates of the color terms from the WCS data resemble basic color terms of English. Similarly, the results of our simple color experiment confirmed that Slovak language contains firmly established 11 basic color categories that can be directly translated to English terms. The most important contribution of our experiment was that we learned more about the methodology of the WCS experiment in practice. The crucial factor in this experiment is that the subjects have to be fully aware of the task to use only basic terms throughout the whole experiment. The insufficiency of instructions might have been a cause of discrepancies in some of the WCS data. Regarding our visualization methods we would like to emphasize the innovative fuzzy visualization, which uses proportional color mixtures to display not only the winning color terms, but also the whole distribution of the responses of subjects.

The core of this thesis is our model of color categorization based on Rosch's prototype theory and on the semantics of distinguishing criteria. We tested this model in two steps. Firstly we experimented with the WCS data. For each of the WCS languages we created an idealized simulated speaker, who learned how to categorize and name colors from all responses of all subjects from the experiment. We evaluated these simulations using the pre-processed data from WCS. We compared the responses of the simulated speakers and real speakers on the basis of winning categories and normed vectors of distributions of categorical responses. For sufficient extent of testing we used also the k-folded cross-validation. Results of the simulations confirm that that the distinguishing criteria are able to successfully generalize the input data.

In the next step we tested the distinguishing criteria in unsupervised learning task using multi-agent simulation of color categorization. We compared the performance of the distinguishing criteria with the adaptive networks used in the original simulation with the result, that the categorization ability of the distinguishing criteria is comparable to the ability of the adaptive networks, and that they are suitable for tasks requiring unsupervised learning. In general, we can conclude that the distinguishing criteria are suitable for the modeling of color categorization.

Bibliography

[1] *Ľudské Telo*. Gemini, 1991.

[2] T. Belpaeme. The origins of colour categories. 2002.

[3] T. Belpaeme. Explaining Universal Color Categories Through a Constrained Acquisition Process. *Adaptive Behavior*, 13(4):293–310, December 2005.

[4] T. Belpaeme and J. Bleys. Colourful language and colour categories. 2005.

[5] B. Berlin and P. Kay. *Basic color terms: Their universality and evolution*. University of California Press, 1969.

[6] B.R. Conway. Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *Journal of Neuroscience*, 21(8):2768, 2001.

[7] R.S. Cook, P. Kay, and T. Regier. The World Color Survey database. *Handbook of categorization in cognitive science*, pages 223–241, 2005.

[8] R.S. Cook, P. Kay, and Regier T. Wcs data archives, 2010.

[9] M.D. Fairchild. *Color appearance models*. Wiley, 2005.

[10] P. Gärdenfors. *Conceptual spaces: The geometry of thought*. The MIT Press, 2004.

[11] P. Gouras. Color perception in webvision, the organization of the retina and visual system, john moran eye center, university of utah, 2010.

[12] P. Gouras. Color vision in webvision, the organization of the retina and visual system, john moran eye center, university of utah, 2010.

[13] M. Green. Basic color & design sbfaq, 2010.

[14] E. Heider[Rosch]. Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1):10–20, 1972.

[15] ID-Studiolab. Id-studiolab: onnos colour show : onnos colour pics, 2010.

[16] K. Jameson and R.G. D'Andrade. It's not really red, green, yellow, blue: An inquiry into perceptual color space. *Color Categories in Thought and Language*, pages 295–319, 1997.

[17] P. Kay, B. Berlin, L. Maffi, and W. Merrifield. Color Naming Across Languages. *Color categories in thought and language*, pages 21–58, 1997.

[18] P. Kay and T. Regier. Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15):9085, 2003.

[19] R.G. Kuehni and G. Woolfe. Color space and its divisions: color order from antiquity to the present. *Journal of Electronic Imaging*, 13:907, 2004.

[20] G. Lakoff. *Women, fire, and dangerous things*. The University of Chicago Press, 1987.

[21] D.T. Lindsey and A.M. Brown. Universality of color names. *Proceedings of the National Academy of Sciences of the United States of America*, 103(44):16608–13, October 2006.

[22] D.T. Lindsey and A.M. Brown. World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences*, 106(47):19785, 2009.

[23] J.J. McCann. Local/global mechanisms for color constancy. *Die Farbe*, 34:275–283, 1987.

[24] G.L. Murphy. *The big book of concepts*. The MIT Press, 2004.

[25] Encyclopædia Britannica Online. Munsell colour system, 2010.

[26] K. Rebrová. Basic color categories. Master's thesis, Univerzita Komenského v Bratislave, 2009.

[27] T. Regier, P. Kay, and R.S. Cook. Universal foci and varying boundaries in linguistic color categories. In *Proceedings of the XXVII annual conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.* Citeseer, 2005.

[28] D. Roberson, I. Davies, and J. Davidoff. Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology General*, 129(3):369–398, 2000.

[29] E. Rosch. Principles of categorization. *MIT Press, Cambridge, MA*, 1999.

[30] J. Šefránek. Kogníicia bez mentálnych procesov. *Kognitívne vedy. Bratislava: Kaligram*, pages 200–256, 2002.

[31] R.N. Shepard. The perceptual organization of colors: An adaptation to regularities of the terrestrial world?. 1992.

[32] C. Starr, C.A. Evers, and L. Starr. *Biology: Concepts and Applications Without Physiology*. Brooks/Cole Pub Co, 2007.

[33] L. Steels. Perceptually grounded meaning creation. In *Proceedings of the International Conference on Multi-Agent Systems*, pages 338–344, 1996.

[34] L Steels. Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7(7):308–312, July 2003.

[35] L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and brain sciences*, 28(04):469–489, 2005.

[36] M. Takáč. Autonomous construction of ecologically and socially relevant semantics. *Cognitive Systems Research*, 9(4):293–311, 2008.

[37] R.A. Wilson, F.C. Keil, and Massachusetts Institute of Technology. *The MIT encyclopedia of the cognitive sciences*. MIT Press, 1999.

[38] S.N. Yendrikhovskij. Computing color categories from statistics of natural images: Color Imaging Science. *The Journal of imaging science and technology*, 45(5):409–417, 2001.