Usage-related variation in the referential range of *blue* in marketing context

Alena Anishchanka, Dirk Speelman, Dirk Geeraerts

University of Leuven

The paper explores language-internal variation in the referential meaning of the lexical form *blue*. Taking a usage-based cognitive approach, we analyze the referential range of *blue* in several marketing contexts from a semasiological and an onomasiological perspective. The study develops an interdisciplinary method that combines frequency analysis with mapping of the referent distributions in the three-dimensional *CIELab* color space. It is argued that the observed referential variation in *blue* is influenced by usage-related factors such as availability of the referents, diversity of color naming strategies and onomasiological competition between lexical forms in the individual product categories.

1.   Introduction

Understanding the relation between the variability in the lexicalization of color concepts across languages and the universal aspects of human color perception has defined the core of the multidisciplinary research in color semantics over its century-long history (for an overview see Berlin & Kay 1999 [1969]; Dedrick 1998; MacLaury 1997). The ongoing debate over the principles of color categorization has revealed the multivariate nature of the linguistic construal of color concepts and generated a range of cognitive models that have been successfully applied to other sensory and non-sensory domains (see, for instance, Dubois 2007; Majid & Levinson 2011; Plümacher & Holz 2007). At the same time, due to the divergent theoretical frameworks and methods in color research some aspects of color conceptualization remain outside the scope of attention. One such underexplored area involves the question about how constant the referential range of color terms across different usage contexts is.

This paper takes a linguistic approach to referential variation in color vocabulary exploring it as a case of language-internal variation. Using the lexical form blue as an example, we compare its frequency distribution and referential range in several contexts represented in marketing discourse. Our main question is whether blue consistently refers to the same color shades when used to speak about colors of cars, clothing or house paints. The goal of the analysis is to uncover the linguistic conceptual and pragmatic mechanisms that are involved in adapting the structure of a general color concept like BLUE to specific usage situations.

The analysis builds on the referentially and lectally-enriched model of semantics proposed in Geeraerts *et al.* (1994, see also Geeraerts 2006). Following this cognitive usage-based approach to linguistic meaning, the referential structure of blue is explored from a semasiological and an onomasiological perspective, focusing on situation-related factors such as the distribution of the different types of blue referents across product contexts and the linguistic competition between lexical forms. The case study relies on referentially-enriched data including color names and their referents represented by color samples retrieved from online marketing materials. Methodologically, it combines frequency analysis, as traditionally applied in text-based linguistic studies of color vocabulary, with referent mapping techniques, developed in experimental color categorization research.

The paper is divided into five sections. Section 2 situates the study among the existing cognitive and linguistic approaches to universality and variation in color semantics. Section 3 introduces the data and the three-dimensional color mapping techniques used in the analysis. Section 4 presents the results of a case study exploring variation in the referential structure of blue in marketing discourse and Section 5 summarizes the conclusions.

2.   Analyzing variation in color lexicalization patterns

The mismatch between the seeming arbitrariness in color lexicalization across cultural and linguistic contexts and the assumed universality of human color perception mechanisms remains one of the central issues in the multidisciplinary color semantics research. Two

traditions — which can be loosely labeled as 'cognitive' and 'linguistic' — have been particularly active in understanding universality and variation in color conceptualization and its linguistic expression. While the relation between the perceptual, cognitive and linguistic mechanisms of color semantics is central to both traditions, a marked difference in their theoretical assumptions and methods often leads to different understanding of the relevant dimensions of variation observed in the linguistic construal of color concepts. The following two sections discuss the views on variation adopted in color categorization and in linguistic research of color vocabulary. In Section 2.3, we discuss the possibility of integrating the two views using the framework of referentially and lectally-enriched language analysis.

2.1　　Cognitive perspective on variation in color categorization

The cognitive paradigm, developed by cognitive psychology and anthropology, focuses on the principles of color categorization and the nature of color categories. This type of research is pursued from two opposing perspectives. At the one end, universality-oriented theories emphasize the pan-human color categorization principles; the major type of evidence in this framework is the uniformity of basic color term (BCT) systems across languages. One of the most influential lines of research known as the 'universal-foci' view, going back to Berlin & Kay (1999 [1969]), holds that color categories are organized around universal focal colors ultimately constrained by the biology of human vision. At the other end, several 'relativist' approaches emphasize the variability of color categorization systems across cultures, highlighting the effects of linguistic, cultural and contextual factors. In recent decades, multiple mechanisms and factors that constrain the observed patterns of variation have been revealed, including, for instance, cognitive mechanisms of attention to similarity and distinctiveness (MacLaury 1997), linguistic convention (Roberson *et al.* 2000, 2005), perceptual distances between colors (Jameson & D'Andrade 1997), distribution of colors in the environment and experience of the speakers (Webster *et al.* 2002; Yendrikhovskij 2001), cultural tradition and symbolism (Paramei 2007).

In theoretical and methodological terms, we can highlight a number of features characteristic of the cognitive framework that are relevant for this paper. First of all, the emphasis on the perceptual basis of color categories places the referential meaning of color terms at the very center of the analysis. In other words, color vocabulary is researched primarily as it serves to describe color percepts. Secondly, the analysis of color categorization principles relies on elicited experimental data following the traditions of anthropological, ethnographic and psychological research. The most common methods applied in the analysis of the color category structures involve mappings between color names and color samples (referents) associated with individual (usually basic) color terms across languages. Thirdly, the observed variation in the lexicalization of color categories tends to be used for formulating arguments for or against the universalist position and in many cases does not constitute the primary subject of the analysis. Finally, most of the evidence is drawn from cross-linguistic and cross-cultural data, while language-internal variation receives relatively little attention in this context. The latter is somewhat puzzling, given that already Berlin & Kay (1999 [1969]: 10) observe that language-internal or inter-speaker variation can be more pronounced than the variation between languages (see also Kay 1975).

In spite of the limited attention paid to language-internal variation, various factors have been shown to be relevant. Berlin and Kay report differences in the structure of color category systems among speakers of different age groups. Other relevant demographic variables include gender (e.g. Bimler *et al.* 2004) and variant types of color vision (e.g. Jameson & Komarova 2009). A series of studies by Athanasopoulos and colleagues (Athanasopoulos 2009; Athanasopoulos *et al*. 2010; Thierry *et al.* 2009) demonstrates that language contact in the form of bilingualism can have an effect on the structure of color categories. More specifically, the analysis of the terms for BLUE in Greek-English bilinguals reveals a shift in the location of the category foci towards that of monolingual speakers of the second language. Webster *et al*. (2002) point to the importance of expert usage; in their study of several Indian communities, they argue that the professional experience of silk merchants influences their location of focal colors.

Given the growing recognition of the linguistic mechanisms involved in color cognition, a comprehensive understanding of the contextualized lexicalization processes from a linguistic point of view can provide complementary evidence regarding language-internal variation.

2.2   Linguistic perspective on variation in color lexicons

Linguistic analyses of color vocabulary represent another research framework most directly involving color language and categorization. In this approach, traditionally based on written texts, usage-related and contextual variation in the linguistic expression of color concepts remains a major focus of investigation. Two related analytical approaches deserve our attention. The first approach focuses on the immediate linguistic context of color word usage including grammatical constructions and collocations with object names (Rakhilina 2007; Steinvall 2002). These analyses reveal variation in the combinatorial preferences of color words for specific taxonomical classes of objects including natural objects, humans, artifacts, etc. Thus, Steinvall finds that dictionary data (Oxford English Dictionary) show a higher combinability of BCTs with natural objects, while usage data from the Bank of English show a higher proportion of artifacts (cf. Rakhilina 2007). Furthermore, the analysis of non-basic (elaborate) color terms demonstrates significant effects of more fine-grained nominal fields on the usage of individual color terms. Comparing several groups of color terms with near-synonymous referential range, e.g. *crimson, plum, maroon, magenta,* and *puce*, Steinvall points out that "*crimson* and *magenta* are most often used in the field of PLANTS, whereas CLOTHING is the most salient nominal field in *plum* and *maroon*" (Steinvall 2002: 154).

These findings demonstrate the impact of context on the semasiological structure of color words and on the linguistic construal of the color property of an object. Depending on the type of object, similar colors can be lexicalized by different words (*plum* vs. *maroon*) and with different specificity (basic vs. elaborate color names). Furthermore, extending collocational analysis from descriptive (*blue dress*) to classifying (*red hair*) and figurative

(*blue mood*) usage reveals the semantic mechanisms that allow the use of color words beyond the color properties of objects (cf. Allan 2009; Verspoor & De Bie-Kerékjártó 2006).

The second approach relevant to this discussion involves the usage of color words in different types of texts, e.g. poetry (Bernhart 2001), advertising (Bergh 2007; Klaus 1989; Stoeva-Holm 2007; Wyler 2007), travelogues (Steinvall 2011), artists' discourse (Anishchanka 2010; Plümacher 2007). The broader context brought into the scope of these analyses gives access to an even wider range of factors involved in the usage of color names. For instance, Stoeva-Holm (2007) describes color naming strategies relative to the different types of referents common in fashion including single colors, color combinations and patterns. Bergh (2007) discusses marketing-related aspects of color naming in the automotive industry, e.g. status communication through evocative imagery in names like *tropic green metallic* or *diplomat blue*. Plümacher (2007) shows how the development of color theory in the fine arts has influenced the ways artists lexicalize color.

These examples of contextualized usage of color vocabulary reveal the multivariate nature of color lexicalization in specific communicative situations that can be accessed through the analysis of larger segments of discourse. However, one aspect of color naming that remains particularly challenging for text-based analyses is the referential meaning of color words. Verbal expression and written texts are known to be limited in conveying sensory information such as color (cf. Majid & Levinson 2011). In some cases this results in the preference of linguistic color studies for non-referential (figurative) meanings and structural characteristics of color words. In addition, the limited access to the referential aspects of color concepts hinders a comprehensive interdisciplinary understanding of the cognitive and linguistic mechanisms of color naming and categorization.

2.3   Referentially and lectally-enriched language description

Given the two traditions of color categorization research outlined in 1.1 and 1.2, the present paper attempts to link the linguistic analysis of usage-related variation in the semantics of color words with the methodological developments in cognitive color categorization research.

More specifically, we seek to enrich the text-based linguistic approach with the referent-related information as a basis for a multivariate analysis of the referential variation in color concepts. Our approach relies on the referentially and lectally-enriched model of language semantics proposed in Geeraerts *et al.* (1994) (see also Geeraerts 2006).

The studies of contextual variation in the semantics of color words discussed in the previous section are in line with a broader view of language-internal variation central to the cognitive usage-based approach to language. Focusing on usage events or usage situations, this view brings into the scope of analysis multiple perceptual, cognitive, linguistic, social, and cultural factors involved in the specific spatio-temporal context of language usage (cf. Geeraerts 2005).

Geeraerts *et al.* (1994) demonstrated that the usage-related variation in the semantic structure of linguistic forms can be systematically investigated along several dimensions including semasiological, onomasiological, formal and lectal.[1] Semasiological and onomasiological variation are two mechanisms central to linguistic categorization. The semasiological perspective takes the lexical form as a starting point and describes the variation in its semantic values or the internal category structure. The onomasiological perspective starts from the concept and describes the linguistic alternatives in its construal and naming. Using the lexical field of clothing terms as an example, Geeraerts *et al.* (1994: 190) demonstrate the intrinsic relation between the two types of variation. It is shown that referents are expressed more readily by the lexical category if they are the central members in its semasiological structure, and they are expressed more readily by a lexical item with a higher onomasiological salience. Furthermore, the semasiological and the onomasiological characteristics of lexical categories are linked with the formal expression of the concepts in polylexical expressions and are subject to lectal variation including speaker characteristics and pragmatic situational effects.

---

[1] The term 'lectal' is used as a generalization over sociolectal, dialectal, ideolectal, stylistic, etc. factors associated with the heterogeneity of linguistic communities (cf. Geeraerts 2005).

Crucially, the semasiological and the onomasiological perspective have their parallels in the tradition of cognitive color categorization research. At the methodological level, they correspond to the color mapping and color naming techniques used in the analysis of color categories. At the theoretical level, the semasiological perspective is represented by the prototype theory of color categories which assumes their fuzzy structure and the salience of focal colors. The onomasiological variation has traditionally received less attention in color categorization research, which might be attributed to the dominant role of the BCT hierarchy. Note, however, that the salience of BCTs implies an onomasiological perspective: they constitute the default level of color category selection (cf. Geeraerts *et al.* 1994: 134). Further, the ongoing debate about the number of BCTs suggests that there might be variation in their onomasiological salience. In addition, Jameson & Alvarado (2003) demonstrate the relevance of the onomasiological perspective showing that the categorization processes are affected by the available onomasiological choices in the specific languages and experimental designs.

Building on these convergent lines of argument, we propose a usage-based analysis of the referential structure of *blue* combining a semasiological and an onomasiological perspective. The following section discusses two methodological elements that are key to the present study. Firstly, it relies on referentially and lectally-enriched data representing authentic language usage in a marketing context, and, secondly, we apply referent mapping techniques that allow the modeling of the referential range of color words for a complementary semasiological and onomasiological analysis.

3.   Data and method

3.1   Database of color names and color samples

As shown in the previous sections, the linguistic studies of color semantics increasingly rely on the data representing authentic language usage in the form of corpora (Steinvall 2002; Verspoor & De Bie-Kerékjártó 2006). However, the existing text corpora give only an indirect access to color referents, a restriction that essentially limits the analyses to non-

referential aspects of color semantics. To address this issue, we compiled a database of color names and color samples (total number of observations in the database 43,880) intended to include referent-related information and to represent the usage of color vocabulary in real-life settings. The data were collected from the websites of manufacturing and retail companies in the US market in four product categories: cars (N=12,358), clothing (N=17,098), makeup (N=5,773), and house paints (N=8,651). Using automatic and manual data retrieval, we extracted all the instances of color names and color samples representing color options available for individual products, e.g. a specific car model or a sweater.[2]

Table 1 includes a selection of examples from the database illustrating three types of information relevant for the following analysis of the context-related referential variation in color concepts. Firstly, linguistic information was obtained in the form of color phrases (e.g. *dark blue, navy, timid blue*, etc.). Although limited in terms of linguistic context, these color names demonstrate the diversity of lexicalization patterns employed in color marketing. Secondly, we included the numeric specification of the color referents represented by RGB values in the html code of the web-sites. The RGB values stand for the amount of red, green and blue light in the graphical color representation on the screen, and are used as a format of digital color encoding. In the following case study, the RGB values converted to *CIELab* coordinates are used to operationalize the referential range of color phrases, applying the techniques that will be introduced in Section 3.2. Thirdly, the database includes the information about the usage context of the color name. In this paper we focus only on the most general context-related variable represented by the product categories.

Table 1. Examples of the data included in the database used in the following analysis

color phrase	*R*	*G*	*B*	*L*	*a*	*b*	*product category*
dark blue	23	46	70	18.25	-0.40	-17.74	cars

---

bimini blue	106	164	244	66.58	3.83	-45.77	clothes
timid blue	225	233	230	91.68	-3.22	0.52	paint
navy	58	69	73	28.47	-3.46	-3.92	clothes
marina	169	205	213	80.10	-10.24	-7.97	makeup

In order to control for the possible product bias, all the calculations reported below were made for a balanced dataset of 22000 observations including 5500 randomly selected observations from each product category.

3.2   Three-dimensional color modeling

The method applied in the present study relies on the analysis of the spatial distribution of the referents associated with a lexical form, e.g. *blue*, in a 3D color space (Figures 1 and 3). The mapping of color referents is based on their numeric representation encoded in the RGB values retrieved from the html code of the websites and converted to the coordinates of *CIELab* color space (Fairchild 1998; Hunt & Pointer 2011).[3] *CIELab* was selected as the color vision model since, in geometric terms, it currently provides the best approximation of perceptual distances between colors and with its three dimensions *L*, *a** and *b**, represents the whole gamut of colors perceived by humans. The vertical *L* (luminance, or brightness/lightness) dimension is scaled from 0 for black to 100 for white. Two perpendicular chromatic axes of the *CIELab* 'equator', *a** and *b**, each with the range [-100 – 100], describe hue by representing the two opponent perceptual systems. In the *a** axis, positive values correspond to the amount of the subjective red component while negative values to the amount of the green component. The *b** axis distinguishes between yellow (positive values) and blue (negative values). The values on the chromatic axes approaching zero correspond to desaturated colors, i.e. with whitish or greyish admixture.

---

[3] The *CIELab* coordinates were obtained with *convertColor* function implemented in R: *grDevices* package.
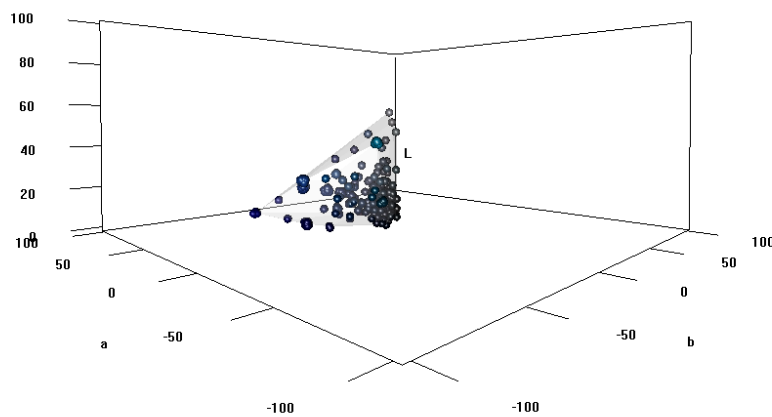
Figure 1. Referential range of *navy* in the *CIELab* color space

Given the *CIELab* coordinates, the referential range of the individual color phrases can be defined in this 3D color space in two geometrical ways illustrated in Figure 1: as a cloud of data points and as a convex hull.[4] The data cloud visualization shows the location of the individual referents of a color phrase in the form of a scatterplot and gives an idea of their density in the different sections of the color space. The convex hull, shown as a semi-transparent shape in Figures 1 and 3, delimits the boundary of the data cloud and the 3D section of the color space corresponding to the referential range of a color phrase. Mathematically, a convex hull for a set of points is defined as "the smallest convex set that contains the points" (Barber *et al.* 1996: 469) or, adapted to color referent distribution, it can be seen as the minimal (3D) form that encloses all the referents corresponding to individual lexical forms (e.g. *navy* or *blue*).[5] We use the convex hull representation to operationalize our intuition that color concepts make up 'chunks' of the color space. This operationalization is used for delimiting the referential range of individual color categories and will be discussed in detail in Section 4.1.

Representations of referent distributions in the color space give us a way to explore context-related variability in the asymmetrical mappings between color names and their

---

[4] The visualizations are plotted with the functions implemented in R: *rgl* package.
[5] Convex hull estimations were obtained with the functions available in R: *geometry* package.

referents from two alternative perspectives. From a semasiological perspective, we take the lexical form as a starting point and analyze the referent distribution it is associated with in the specific contexts. This perspective is comparable to experimental color mapping tasks, where subjects are asked to indicate the region of the color array that corresponds to the color name presented as a stimulus. An onomasiological analysis starts from the color concept represented by a region in the color space and investigates the lexical forms used to describe it in specific contexts. This approach is comparable to experimental color naming tasks, where subjects are asked to give names for the color chip stimuli. In the following sections, we combine these two perspectives to explore contextual variation in the referential range of the lexical form *blue*.

4.   Results and discussion

4.1   The semasiological range of monolexemic and schematic [*blue*][6]

Context-related variation in the referential range of monolexemic *blue* can be addressed in quantitative and in qualitative terms. If we compare the frequency of occurrence of monolexemic *blue* in the four product contexts, following the tradition of corpus-based analyses of color word usage (e.g. Steinvall 2002), we find the frequency distribution summarized in Table 2. The frequency of usage in the balanced dataset shows that most instances occur in the clothing subset, while in cars we find only 6 observations and in makeup only 1; in paints, not a single occurrence of monolexemic *blue* is found. This suggests a stronger association of monolexemic *blue* with the clothing domain compared to cars, makeup or paints.

Table 2. Frequency of monolexemic *blue* in the four product categories

Product category	*Frequency*

---

[6] Square brackets are added to indicate reference to schematic names, e.g. [*blue*], [*navy*] (operationalized as sets of all phrases with the respective form in the head position) and the referential range of the respective schematic names, e.g. [BLUE] (operationalized as the set of all referents corresponding to the schematic name [*blue*])

cars	6
clothes	38
makeup	1
paint	0

However, a more interesting question is whether this difference in frequency correlates with the semasiological variation in the spatial distribution of blue referents. In other words, does the difference in the usage of monolexemic *blue* boil down to the higher frequency of the name in specific contexts, e.g. clothing, or is the name also associated with different sections of the color space?

To answer this question, we analyze the geometrical distribution of *blue* referents in the *CIELab* color space for each of the four product categories (Figure 2). As pointed out in Section 3.2, the *L* axis models the lightness dimension and the *b** axis the yellow-blue dimension, which makes the *Lb** projection (Figure 2b) the most informative for the analysis of the referential range of *blue*. The visualization reveals noticeable variation in the spatial distribution of the *blue* referents across product contexts. The referents from the clothing subset, besides being more frequent, are dispersed over a larger section of the color space compared to the other product categories. The referents from the car subset and the few referents from makeup appear in the lower values of the *L* axis suggesting darker shades.
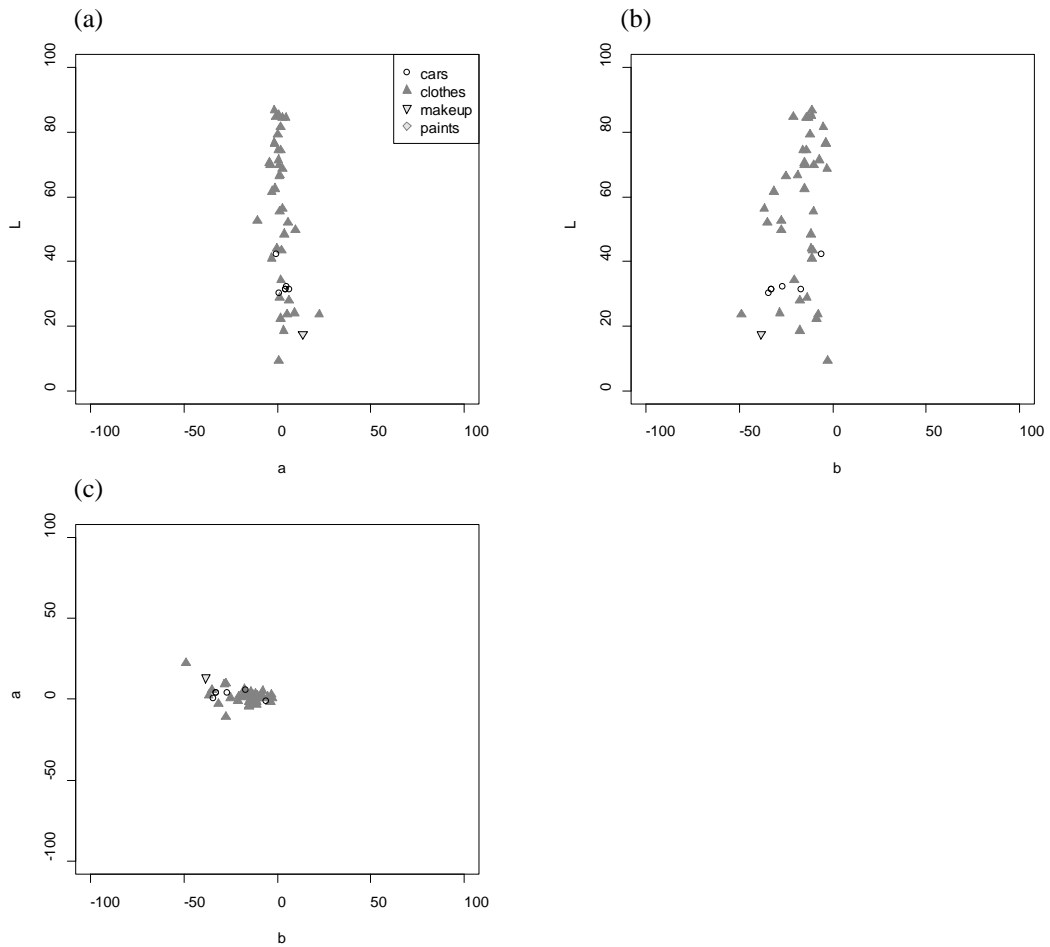
Figure 2. Referent distribution in the *CIELab* for monolexemic *blue* in the four product categories represented as projections of the 3D color space: (a) *La** projection; (b) *Lb** projection; (c) *a*b** projection

However, given the low frequencies of *blue* in cars and makeup, the question arises whether the observed distribution pattern for monolexemic *blue* is sufficient to draw reliable conclusions about the structure of the concept BLUE in these contexts. The question is even more relevant for the paint domain, where we have no information about the referent structure of *blue*, since the monolexemic form does not occur here.

To address this issue, we extend the analysis of referent distributions to a more schematic level of lexicalization of the concept BLUE. More specifically, we compare monolexemic *blue* with polylexemic names comprising *blue* (e.g. *dark blue, bimini blue,* etc.) as an additional means for analyzing this conceptual structure at less specific levels of lexicalization. The underlying logic of this argument lies in the diversity and specificity of

color naming in marketing. This notorious practice in color marketing is driven by the fine-grained distinctions that often need to be made between a large number of similar colors (Wyler 2007) and the tendency to create elaborate names that are intended to evoke appealing images 'boosting' the attractiveness of the product (Bergh 2007). To give an example, paint retailers often display dozens of blue shades next to each other presenting the complete palette to the consumer. In this situation, the distinctive function of color names makes monolexemic lexicalizations like *blue* not sufficiently informative. Instead, we commonly find names like *blue silk, capri blue, deep water blue,* etc. It stands to reason to analyze these names as composite expressions or phrases which specify color nuances of a more abstract or schematic [BLUE] category by adding a modifier to the BCT. From this perspective, the 'multiple-modifier + *blue*' names can be viewed as instantiations of the schematic name [*blue*] whose referential range can be mapped as an aggregated referential range of all the modifier + *blue* names (Figure 3).[7]
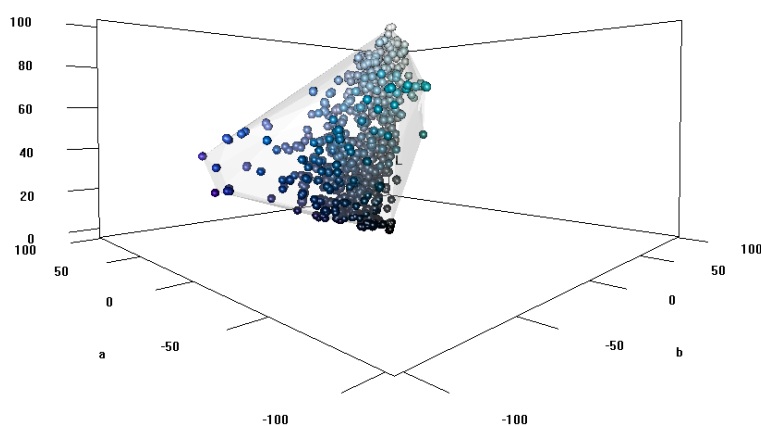

Figure 3. Referential range of schematic [*blue*] in the *CIELab*

To test the consistency of the variation pattern observed in the referential range of monolexemic *blue* (Figure 3)*,* we apply the same visualization technique to the referential range of schematic [*blue*] (Figure 4). In line with our expectation, the [BLUE] concept viewed

---

[7] This logic excludes the regular subordinate terms of *blue*, e.g. *navy*, *azure*, *ultramarine, etc.* which are viewed as lexicalizations of the more specific non-basic concepts.

at a more schematic level of lexicalization is much better represented not only in the total number of observations (see the frequencies in Figure 4d) but also in their spatial distribution in the color space. Crucially, the schematic lexicalizations give us access to the [BLUE] concept in the paints domain, where monolexemic *blue* is not used.
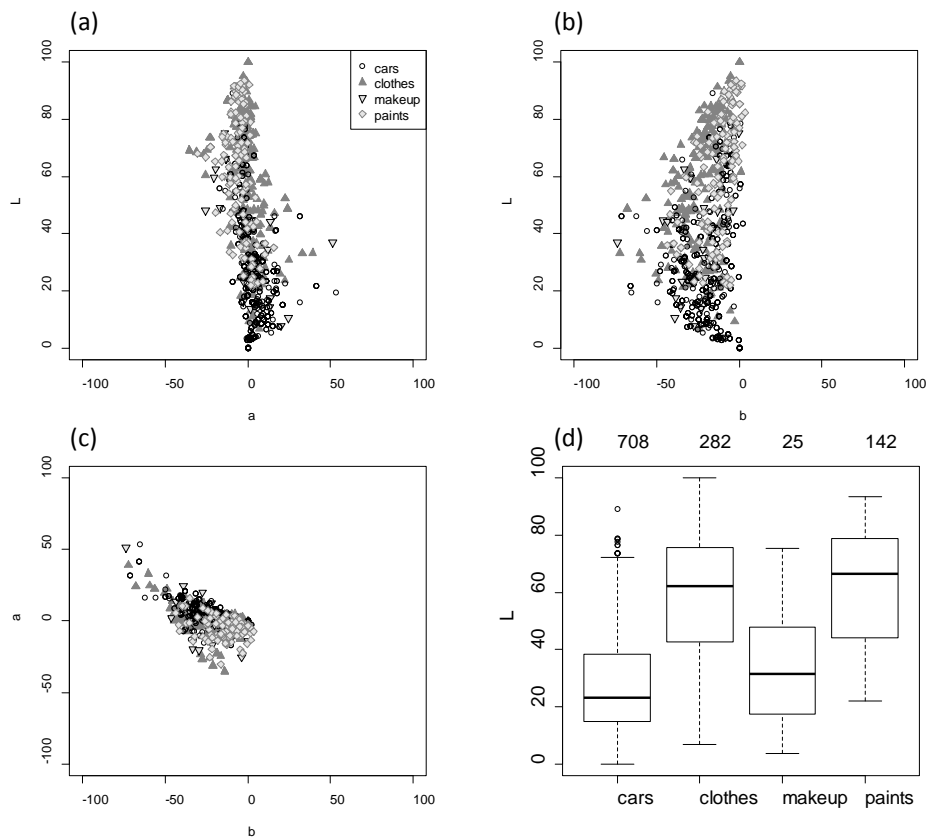

Figure 4. Referent distribution in the *CIELab* for schematic [*blue*] in the four product categories represented as projections of the 3D color space: (a) *La** projection; (b) *Lb** projection; (c) *a*b** projection; (d) distribution of the referents along *L* dimension in the four product categories

The scatterplots in Figure 4 reveal more clearly the variation in the referential range of schematic [*blue*] along the *L* axis in the four product contexts. Additionally, the distribution of *L* values for individual subsets is represented in the form of box-and-whisker plots (Figure 4d), where the whiskers show the range of the values for the *L* dimension for each subset, and the boxes represent the values of the 50% of the data and the medians in the respective subsets. Both the scatterplots and the boxplots indicate that [*blue*] referents in the

clothing domain are distributed over a larger section of the color space; however, they appear more densely concentrated in the higher values of the *L* axis suggesting lighter shades. Similarly, the referents from the paint subset appear more frequently in the lighter shades. On the other hand, [*blue*] referents in the car and makeup subsets tend to occur at the lower values of the *L* axis, indicating darker shades. In other words, the referent mapping reveals qualitative differences in the referential structure of [*blue*]. In clothing and in paints color names with *blue* in the head position tend to be associated with lighter shades, while in makeup and cars they map onto darker shades of blue.[8]

In the following section, we consider two interpretations for the observed lightness shifts in the semasiological structure of [*blue*]. Firstly, we consider an extralinguistic explanation and analyze to what extent the observed pattern depends on the 'availability' of lighter and darker blues in the different product contexts. In other words, it is possible that in the car domain [*blue*] happens to be associated with darker shades simply because there are more dark blue cars. Secondly, we analyze the linguistic competition between lexical forms and compare product-specific naming preferences for the same sections of the color space. For instance, the inspection of the clothing subset shows high frequency of *navy*. Our hypothesis is that the semasiological structure of [*blue*] in this domain 'suffers' from onomasiological competition between names. The remaining part of the analysis explores the viability of these two hypotheses.

---

[8] The statistical significance of the observed difference between the product categories was tested with a simple linear regression test. The model shows statistically significant difference in the lightness values between the car subset (reference level) and clothing and paints. The makeup subset is not significantly different from cars.

Coefficients	Estimate	Std. Error	Pr(>|t|)	
Intercept	27.342	0.721	<2e-16	***
clothes	31.578	1.352	<2e-16	***
makeup	6.594	3.906	0.092	
paints	34.760	1.765	<2e-16	***
Adjusted R2: 0.3994				

4.2 Referent distribution and onomasiological entrenchment in the BLUE category

In order to disentangle the effects of linguistic and extralinguistic factors in the referential variation of schematic [*blue*], we switch to the onomasiological perspective, which takes the conceptual structure of BLUE as a starting point and explores the patterns in its linguistic expression. As the first step in the analysis, we test the availability of the blue referents of different lightness across the product contexts. The main question we address is to what extent the observed association of [*blue*] with darker shades in cars and light blues in paints might be explained by the higher proportion of dark and light blue colors in these product categories. The major difficulty of this task is that dark and light blue referents can be named not only *blue* or *dark blue*, but can potentially appear under any name, e.g. *navy, marina, sky*, etc. (cf. Table 1). This means we need a method that allow us to identify blue referents as exemplars of BLUE independently of their actual name.

To address this technical issue, we rely on the convex hull operationalization of the referential range of [*blue*] (see Section 2.2.) and the assumed convexity of color categories. Gärdenfors (2004: 18) defines convexity as follows: "A convex region is characterized by the criterion that for every pair of points v1 and v2 in the region all points in between v1 and v2 are also in the region". In other words, if we know that samples v1 and v2 belong to BLUE, the points located between these referents potentially belong to the same category at the conceptual level irrespective of their linguistic expression which might contain the lexical form *blue* (*dark blue, timid blue*) or not (e.g. *navy* or *marina*). Starting from this assumption, we use *CIELab* coordinates to construct a three-dimensional convex hull for the set of referents corresponding to schematic [*blue*]. We can then identify all the referents in the balanced dataset, whose *CIELab* coordinates locate them within this convex hull, which serves as a geometric approximation of the boundary for the referential range of schematic [*blue*] and of BLUE. These referents will comprise a set of color samples geometrically

defined as {BLUE}, i.e. they are potential candidates to be named with schematic [*blue*] based on their location in the color space.[9]

As the next step, we explore the variability of {BLUE} set of referents along the lightness dimension to understand the availability of lighter and darker blues across product contexts. As a coarse-grained distinction between light and dark blues, four levels of lightness were identified for further analysis: L1 - dark blues ($0 < L \leq 25$), L2 - medium blues ($25 < L \leq 50$), L3 - light blues ($50 < L \leq 75$), and L4 - very light blues ($75 < L < 100$).[10] Based on this approximation, Figure 5 shows the proportion of lighter and darker blues for each product category.

Three product categories (cars, clothes, makeup) are marked by similar profiles, where dark shades make the largest proportion of the available blues and light blues are the smallest proportion. This pattern supports the extralinguistic explanation for the dominance of the darker shades in the referential range of [*blue*] in cars and makeup due to the higher availability of darker blue products in these categories. At the same time, the higher proportion of light blues in the paint subset explains the lighter referential range of [*blue*] in this domain. However, the extralinguistic explanation does not provide a sensible account for the dominance of the lighter shades in [*blue*] in the clothing subset (Figure 4). Figure 5 shows that the proportion of dark blues in clothing is similar to that in cars and makeup; however, they are not associated with [*blue*] names. One possible explanation that was pointed out above is the onomasiological competition between linguistic forms. The inspection of the color phrases in clothing shows a high proportion of *navy*, which happens to be a dark blue color. This might mean that dark blues in clothing are linguistically construed relative to the NAVY rather than BLUE category, which has an effect on the observed referential range of [*blue*].

---

[9] Curly brackets {BLUE} are added to indicate reference to the BLUE concept geometrically identified with a convex hull, i.e. the set of referents that are located within the convex hull corresponding to [*blue*] names irrespective of their actual name. This conceptual structure is used as a starting point for the onomasiological analysis of linguistic choices.

[10] The lightness levels were identified with the purpose of achieving equal ranges. The extreme values of the *L* axis were excluded to reduce the number of black and white referents that would otherwise be automatically included in the convex hull.
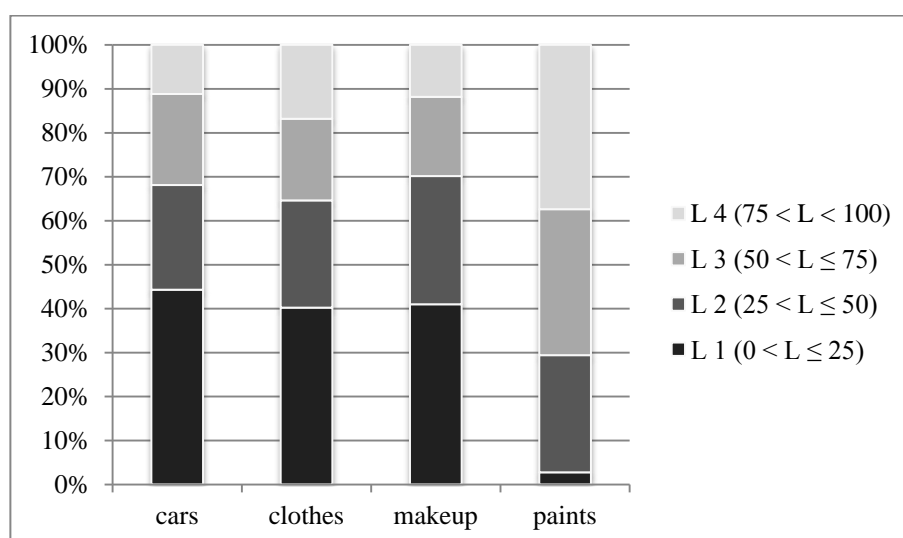
Figure 5. Proportion of the referents categorized into four lightness levels across the product categories

Testing this hypothesis at the final step in the analysis, we explore the onomasiological competition between the phrases used to name dark and light blue referents in the different product categories and its effect on the referential range of [*blue*]. In this case, we start from the question: Given a group of blue samples of the specified lightness level, what are the most commonly used names in each of the product contexts? Taking into account the diversity of color phrases in marketing discussed in Section 4.1, we focus on the schematic level of lexicalization operationalized as an aggregation over the forms used in the head position.

Table 3 lists the most frequent lexical forms used in the head position in each lightness group and their relative frequencies for the respective subset of referents. The higher relative frequency of the lexical form can be viewed as an indicator of its onomasiological salience or entrenchment (cf. Geeraerts *et al*. 1994: 138, 180). Table 3 reveals a number of onomasiological preferences for the referents of the different lightness levels in the {BLUE} category.[11] Most noticeably, schematic [*blue*] is the preferred onomasiological choice in 10

---

[11] A number of names in Table 3 do not belong to the BLUE category, e.g. [*black*], [*white*], [*gray*] [*charcoal*], [*silver*], [*green*], [*purple*], etc., which can be given two explanations. Firstly, this is due to the non-discreteness of color categories, which results in their referential

out of 16 groups of referents. The preference for [*blue*] is especially prominent in the mid-range blues (L2 and L3) where it is marked by higher entrenchment compared to the alternative lexical forms. However, there is one notable exception in the L1 group in clothes, where, in line with our expectation, [*navy*] is the most obvious onomasiological choice.

Table 3. Relative frequencies of the most frequent lexical forms used in the head position for {BLUE} referents per lightness group and product category

	Cars	*%*	*Clothes*	*%*	*Makeup*	*%*	*Paints*	*%*
L4	silver	46.21	white	31.05	clear	5.88	**blue**	9.51
	white	31.82	**blue**	25.63			sky	3.71
	blue	2.65	ash	9.75			lilac	1.86
	tungsten	2.65	birch	3.97			gray	1.62
			aqua	3.61			silver	1.62
L3	silver	43.47	**blue**	35.95	**blue**	5.77	**blue**	15.40
	blue	16.33	grey	15.03	silver	5.77	gray	3.13
	gray	6.33	gray	7.52	charcoal	3.85	bay	1.83
	quicksilver	3.47	heather	3.92	lagoon	3.85	green	1.57
	platinum	3.27	hydrangea	2.29	pool	3.85	purple	1.57
			turquoise	2.29			sky	1.57
L2	**blue**	39.89	**blue**	19.15	**blue**	11.90	**blue**	11.40
	gray	29.79	navy	14.18	black	9.52	night	3.58
	granite	3.55	royal	11.94	smoke	3.57	purple	3.58
	green	3.37	charcoal	11.19	steel	3.57	bay	2.28
	slate	3.19	black	6.47	teal	3.57	navy	2.28
L1	**blue**	37.02	navy	40.87	black	50.00	**blue**	21.88
	black	32.63	black	32.88	**blue**	9.32	black	18.75
	gray	5.92	charcoal	5.28	satin	3.39	navy	15.63
	sapphire	5.15	**blue**	3.17	marine	2.54	green	9.38
			indigo	1.51	navy	2.54		
			shadow	1.51				

---

overlap. In other words, very dark blue shades can be also categorized as BLACK, the desaturated blues as GREY and very light blues as WHITE. Secondly, the convex hull estimation for large referential ranges is a rather coarse-grained segmentation of the color space that does not allow fine-grained color distinction, especially given the fuzziness of the color categories. The presence of non-blue names does not contradict the observed patterns in the onomasiological entrenchment of blue names, and, in fact, opens another line of investigation into the onomasiological competition between BCCs (e.g. BLUE vs. BLACK, WHITE, or GREY).

In addition to the relative frequencies of the lexical forms summarized in Table 3, Figure 6 zooms in on the variation in the onomasiological salience of [*blue*] at the different levels of lightness across the product contexts. The most noticeable difference in the onomasiological salience of schematic [*blue*] is observed in the domains of cars and clothing. While we find the highest entrenchment values in the darker shades (L1 and L2), which go down in the lighter shades (L3 and L 4) for cars, we observe the opposite tendency with the highest salience of [*blue*] in the lighter shades (L3 and L4) and lower salience in the darker shades (L1 and L2) for clothing. For paints and makeup, Figure 6 and Table 3 reveal the relatively lower and more consistent entrenchment values of [*blue*], which might be related to higher diversity of naming in these two product categories. In other words, there are more competing names for BLUE in these contexts. Interestingly, Table 3 suggests a relatively high entrenchment of schematic [*blue*] in dark blue paints (L1 group), while Figure 5 shows that this group of referents is the smallest part of the [*blue*] referential range. This means that even if dark blue exemplars are not very common in this product context they are still most likely to be named with a schematic [*blue*] name, indicating the onomasiological salience of the schematic form in this context despite the absence of monolexemic *blue*.
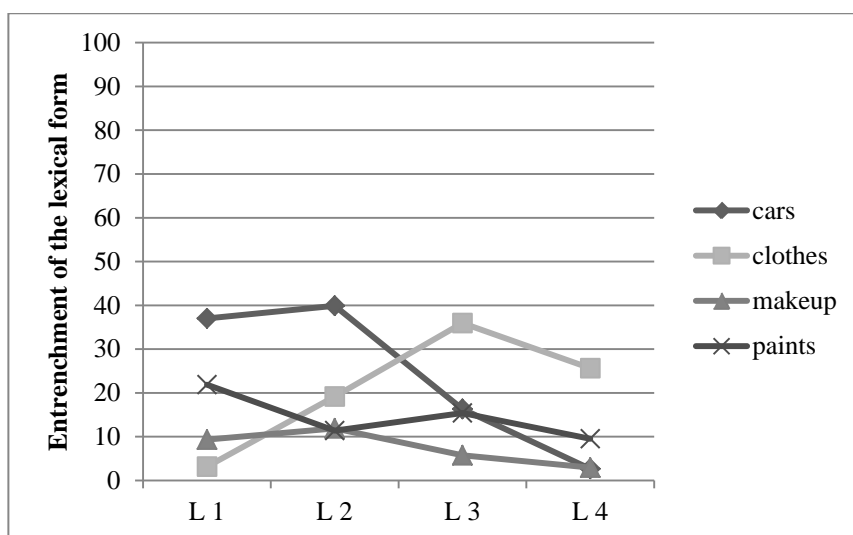

Figure 6. The onomasiological entrenchment of the schematic [*blue*] across lightness levels and product contexts

Figure 6 shows a different pattern in clothing, where the onomasiological entrenchment of schematic [*blue*] decreases in the darker shades in contrast to the other product categories. At the same time, Table 3 reveals a higher onomasiological entrenchment of [*navy*] for the darker blue shades. Figure 7 compares the entrenchment values of the two lexical forms for darker and lighter blue referents in this product category. Very clearly [*navy*] is the preferred choice for the darker blue shades and does not extend into the lighter shades as far as [*blue*] does. As a result [*blue*] is ousted from this segment of the color space, which leads to the contraction of its semasiological range towards lighter shades that was observed in Figure 4. What is also interesting from a linguistic point of view is that in the L2 group of medium blues, the entrenchment of the two forms is very close, meaning that the two names are almost equally likely naming choices.[12]
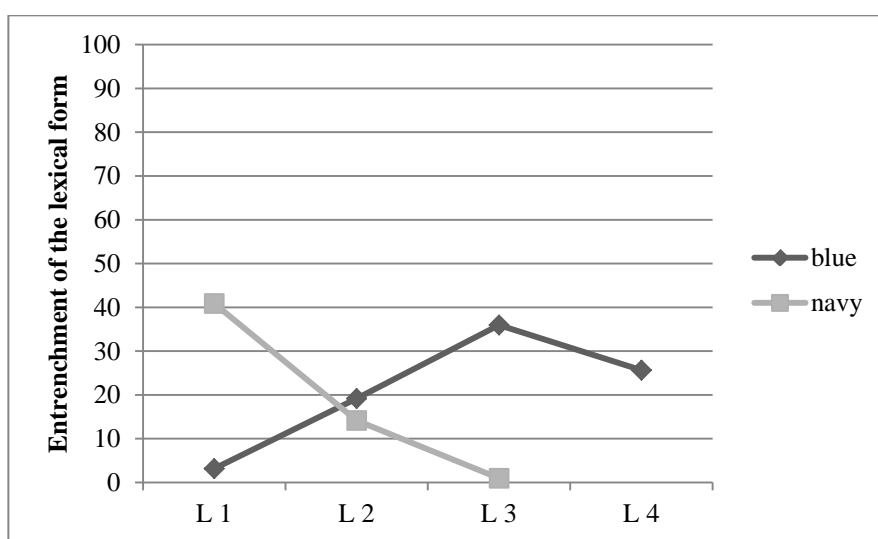

Figure 7. The onomasiological entrenchment of the schematic [*blue*] and [*navy*] across lightness levels in the clothing domain

The entrenchment of [*navy*] for dark blue colors is most obvious in the clothing subset; however, there are traces of this trend in paints, where we find a similar proportion of

---

[12] It is worth adding that the compound form *navy blue* is infrequent in our data (cars N=10, clothing N=1, makeup N=1, paints N=0). Given these low frequencies, no separate analysis was carried out for the compound form.

[*blue*] and [*navy*] in the L1 group. Interestingly, in cars we find no instances of [*navy*] at any of the lightness levels. This might seem surprising; however, one possible explanation can be related to the etymology of *navy*. *Navy* as a name for dark blue color is assumed to have originated as a reference to the dark blue uniform of navy officers. The origin of the name in the clothing domain and its traditionally wider circulation in this context might have resulted in its higher entrenchment. Furthermore, the original association of *navy* with heavy (woolen) fabrics might explain its absence in the context of automotive colors. *Navy* might be seen as less suitable for the shiny polished surfaces of cars for which references to metals and gems are more appropriate.

5. Conclusion

The present study addresses one of the underexplored issues in color semantic research involving usage-related variation in the referential structure of color concepts. Focusing on the lexical form *blue* as an example, we explore linguistic contextual factors that define its referential structure in four marketing contexts. The study combines a frequency analysis of the lexical form usage commonly applied in linguistic color term research with 3D modeling of the referential range of *blue* in the *CIELab* color space. By combining these two methods, the paper brings together two strands of color categorization research that have so far largely existed alongside each other, i.e. the cognitively oriented tradition surveyed in Section 2.1, and the linguistic tradition surveyed in Section 2.2. At the same time, the study continues the line of referentially-enriched semantic research introduced in Geeraerts *et al*. (1994).

Starting from a semasiological perspective, we map the referential range of *blue* in four product subsets: cars, clothing, makeup, and house paints. The comparison of the referent distributions reveals product-specific patterns of variation along the *L* axis. Thus, in the car and makeup subsets the referential range of *blue* is dominated by darker shades, whereas in clothing and makeup the name is associated with lighter sections of the BLUE category. The pattern is observed in monolexemic *blue*, but it is much more pronounced in

the schematic [*blue*], operationalized as an aggregated representation of modifier + *blue* names.

In order to explore the factors involved in the lightness shifts in the referential range of [*blue*], we apply an onomasiological analysis, which relies on language-independent identification of the {BLUE} concept operationalized as a convex hull in the *CIELab* color space. Firstly, the analysis reveals an uneven distribution of lighter and darker blues in the individual product categories. Thus, in cars, clothing, and makeup {BLUE} is represented by darker shades, while in paints the potential referents of [*blue*] are dominated by lighter shades. Although the availability of the lighter and darker blue referents in cars and paints has a likely effect on the semasiological structure of [*blue*] in those product categories, this referential effect cannot be the entire story. In fact, in the clothing domain we observe a linguistic effect related to higher onomasiological entrenchment of another name for dark blue shades — [*navy*]. In other words, while in the other three product categories [*blue*] is the dominant linguistic choice, dark blue clothing is most commonly named with [*navy*]. It ousts [*blue*] names out of this section of the color space leading to the contraction of its referential range towards lighter shades. This result might also indicate that color samples in clothing are categorized and linguistically construed at a more specific level, that of non-basic categories like NAVY; however, further analysis of a wider range of categories would be required to test this intuitive observation.

The proposed analysis is intended to contribute to the interdisciplinary research in color semantics in two ways. From a theoretical point of view, we argue for a quantitative usage-based approach to language-internal variation in the lexicalization of color concepts. This study reveals meaningful patterns of contextual variation in the referential range of *blue*, specifically related to its usage in marketing different product categories. The results of the analysis support the findings regarding semasiological and onomasiological variation in the color domain (Steinvall 2002) and other semantic domains, e.g. clothes (Geeraerts *et al*. 1994; Geeraerts 2006). First, the semasiological range of color categories depends on the type of product they are applied to: the prototypical core of color categories is not stable across

different types of entities. Second, this semasiological effect can be partly explained on referential grounds (the objective frequency of different colors and shades is different in the various products), but this does not exhaust the explanation. In addition to the referential effect, there is an onomasiological effect at work, in the sense that the entrenchment of color names (the likelihood with which they are used to refer to a given color or shade) is not stable across different types of entities. Both findings can be summarized in one sentence: typically blue cars may have a different color than typically blue house paints, and the same color may be more readily called *blue* in cars and *navy* in clothes.

From a methodological perspective, we hope to have demonstrated how the use of referentially enriched data (in the present case, as derived from visual information incorporated in html pages) provides an opportunity to bridge the gap between linguistic, corpus-based work on color terms and cognitive, experimental work on color categorization. Given the huge amounts of internet data currently available, this combination holds considerable potential for further research.

References

Allan, Keith. 2009. The connotations of English colour terms: Colour-based X-phemisms. *Journal of Pragmatics* 41(3). 626–637.

Anishchanka, Alena. 2010. Vantage construal in the attributive use of basic color terms: The AcN and N of Nc constructions. *Language Sciences* 32(2). 170–183.

Athanasopoulos, Panos. 2009. Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition* 12(01). 83–95.

Athanasopoulos, Panos, Ljubica Damjanovic, Andrea Krajciova & Miho Sasaki. 2010. Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition* 14. 9–17.

Barber C. Bradford, David P. Dobkin & Hannu Huhdanpaa. 1996. The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* 22. 469–483.

Bergh, Gunnar. 2007. The semiosis of Swedish car color names: Descriptive and amplifying functions. In Robert E. MacLaury, Galina. V. Paramei & Don Dedrick (eds.), 337–345.

Berlin, Brent & Paul Kay. 1999[1969]. *Basic color terms*. Stanford, CA: CSLI.

Bernhart, Toni. 2001. *Farbe und text – quantitative sondierungen am beispiel von Hans Henny Jahnn*. In Proceedings Der GLDV-Frühjahrstagung 2001, 181–190. Gießen: Henning Lobin.

Bimler, David, John Kirkland & Kimberley Jameson. 2004. Quantifying variations in personal color spaces: Are there sex differences in color vision? *Color Research & Application* 29(2). 128–134.

Dedrick, Don. 1998. *Naming the rainbow: Colour language, colour science, and culture*. Dordrecht: Kluwer.

Dubois, Danièle. 2007. From psychophysics to semiophysics: Categories as acts of meaning. In Martina Plümacher & Peter Holz (eds.), 167–184.

Fairchild, Mark D. 1998. *Color appearance models*. Reading, Mass: Addison-Wesley.

Gärdenfors, Peter. 2004. Conceptual spaces as a framework for knowledge representation. *Mind and Matter* 2(2). 9–27.

Geeraerts, Dirk. 2005. Lectal variation and empirical data in Cognitive Linguistics. In Francisco J. Ruiz de Mendoza Ibáñez & Sandra Peña Cervel (eds.), *Cognitive Linguistics. Internal dynamics and interdisciplinary interactions*, 163–189. Berlin: Mouton.

Geeraerts, Dirk. 2006. *Words and other wonders. Papers on lexical and semantic topics*. Berlin: Mouton.

Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The structure of lexical variation. Meaning, naming, and context*. Berlin: Mouton.

Hunt, Robert W. G. & Michael R Pointer. 2011. *Measuring colour,* 4th edn. New York, NY: Wiley.

Jameson, Kimberly A. & Nancy Alvarado. 2003. The relational correspondence between category exemplars and names. *Philosophical Psychology* 16(1). 25–49.

Jameson, Kimberly A. & Roy G. D'Andrade. 1997. It's not really red, green, yellow, blue: An inquiry into perceptual color space. In Clyde L. Hardin & Luisa Maffi (eds.), *Color categories in thought and language*, 295–319. Cambridge: CUP.

Jameson, Kimberly A. & Natalia L. Komarova. 2009. Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *Journal of the Optical Society of America* 26(6). 1414–1423.

Kay, Paul. 1975. Synchronic variability and diachronic change in basic color terms. *Language in Society* 4(3). 257–270.

Klaus, Hilde. 1989. Beobachtungen zu den modefarbenwörtern in der deutschen gegenwartssprache. *Zeitschrift für Germanistische Linguistik* 17. 22–57.

MacLaury, Robert E. 1997. *Color and cognition in Mesoamerica: Constructing categories as vantages*. Austin, TX: University of Texas Press.

MacLaury, Robert E., Galina V. Paramei & Don Dedrick (eds.). 2007. *Anthropology of color: Interdisciplinary multilevel modeling*. Amsterdam: Benjamins.

Majid, Asifa & Stephen C. Levinson. 2011. The senses in language and culture. *Senses & Society* 6(1). 5–18.

Paramei, Galina. V. 2007. Russian 'blues': Controversies of basicness. In Robert E. MacLaury, Galina. V. Paramei & Don Dedrick (eds.), 75–105.

Plümacher, Martina. 2007. Color perception, color description and metaphor. In Martina Plümacher & Peter Holz (eds.), 61–84.

Plümacher, Martina & Peter Holz (eds.). 2007. *Speaking of colors and odors*. Amsterdam: Benjamins.

Rakhilina, Ekaterina V. 2007. Linguistic construal of colors: The case of Russian. In Robert E. MacLaury, Galina. V. Paramei & Don Dedrick (eds.), 363–378.

Roberson, Debi, Jules Davidoff, Ian Davies & Laura Shapiro. 2005. Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology* 50(4). 378–411.

Roberson, Debi, Ian R.L. Davies & Jules Davidoff. 2000. Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology General* 129(3). 369–398.

Steinvall, Anders. 2002. *English colour terms in context*. Umeå: Umeå UniversityPh.D thesis.

Steinvall, Anders. 2011. The power of colour term precision. In Carole P. Biggam, Carole A. Hough, Christian J. Kay & David R. Simmons (eds.), *New directions in colour studies*, 219–231. Amsterdam: Benjamins.

Stoeva-Holm, Dessislava. 2007. Color terms in fashion. In Robert E. MacLaury, Galina. V. Paramei & Don Dedrick (eds.), 421–440.

Thierry, Guillaume, Panos Athanasopoulos, Alison Wiggett, Benjamin Dering, & Jan-Rouke Kuipers. 2009. Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America* 106(11). 4567–4570.

Verspoor, Marjolijn H. and Ágnes De Bie-Kerékjártó. 2006. Colorful bits of experience: From bluestocking to blue movie. *English Studies* 87(1). 78–98.

Webster, Michael A., Shernaaz M Webster, Shrikant Bharadwaj, Richa Verma, Jaikishan Jaikumar, Gitanjali Madan & E. Vaithilingham. 2002. Variations in normal color vision. III. Unique hues in Indian and United States observers. *Journal of the Optical Society of America.* 19(10). 1951–62.

Wyler, Siegfried. 2007. Color terms between elegance and beauty: The verbalization of color with textiles and cosmetics. In Martina Plümacher & Peter Holz (eds.), 113–128.

Yendrikhovskij, Sergej N. 2001. A computational model of color categorization. *Color Research and Application* 26. 235–238.