Shades of salience: multivariate analysis of prototypicality effects in color terms

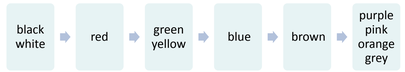
Alena Anishchanka, Natalia Levshina, Dirk Speelman, Dirk Geeraerts

I. Color in the lab and color in the 'real world'

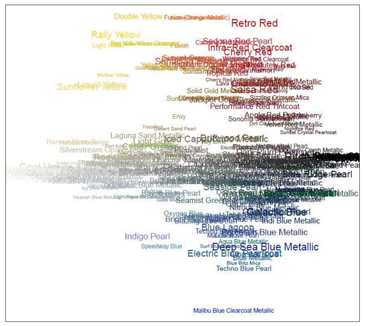
Can we use corpus-based parameters to measure salience? Can we measure salience beyond BCT?

Anthropological, psycholinguistic and linguistic studies of color categorization have developed a number of quantitative procedures for measuring the psychological and linguistic salience (basicness) of color terms.
However, these techniques designed for elicitation and color naming tasks remain limited to basic color term (BCT) analyses due to two constraints:

- practical considerations limit the number of color categories and observations
- experiments are designed to test the salience of BCT vs. non-BCT to corroborate Berlin and Kay's universal hierarchy of color categories

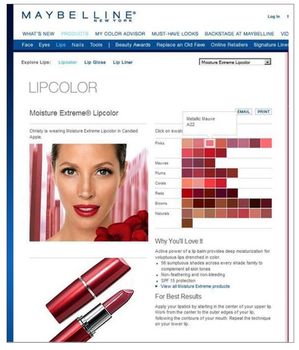


BCT evolutionary hierarchy (Berlin & Kay 1969)



Terms used to name colors of compact cars (based on MDS by Tom Ruette)

II. Data and variables

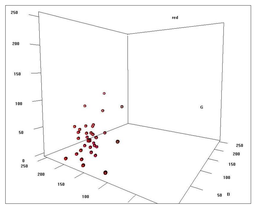


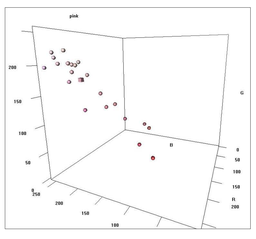
four product domains (USA market):

- cars: *GM, Chrysler Group, Toyota, Mercedes,* etc.
- clothing: *J.Crew, Banana republic, Saks Fifth Avenue, Covington (Sears), Columbia, Authentic Pigment*, etc.
- make up: *L'Orial, Trish McEvoy, MAC, Avon, Maybelline,* etc.
- house paints: *Benjamin Moore, Olympic, Glidden,* etc.

Formal	
Form.LenTok	Term length in constituents
Form.LenChar	Term length in characters
Form.LenPhon*	Term length in phonemes
Form.LenSyl*	Term length in syllables
Experimental	
Exp.RT*	reaction time in ms
Exp.Consist*	ratio of consistent to inconsistent naming
Exp.ElicitFreq*	relative frequency (%) of mention in elicitation experiment
Exp.ElicitRank*	sequence rank of mention in elicitation experiment
Evolutionary	
Evol.BK*	position of Term in BCT evolutionary hierarchy
Corpus-based	
Corp.Freq	frequency of Term used independently
Corp.CompFreq**	frequency of Term in compounds
Corp.CompTypes**	*n* of composites with Term (types)
Corp.TTR**	Term type-token ratio
Corp.Indep**	frequency of Term used independently / token frequency of Term
Corp.Head**	frequency of Term used as Head / token frequency of Term
Corp.HeadLL**	log-likelihood ratio of Term used as Head
Corp.CatNo	number of product categories where Term is used
Corp.MeanDist	average distance from Term exemplars to centroid in rgb color space

* only for a restricted set of Terms
** only for monolexemic Terms





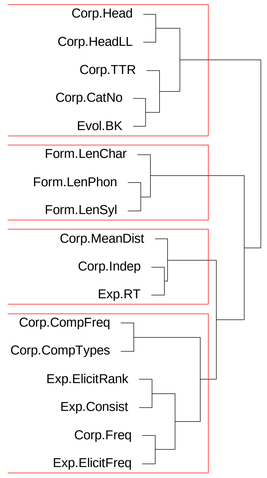
red and *pink* in rgb color space

III. Convergence and divergence in corpus-based and experimental measurements

The first series of analyses explores the relationships between experimental and corpus-derived salience parameters. Using the corpus-based parameters from our data and the available experimental data for 34 categories we address the following questions:

- do corpus-based parameters correlate with experimental ones?
- do experimental and corpus parameters reveal different patterns of salience?

Correlations between salience parameters

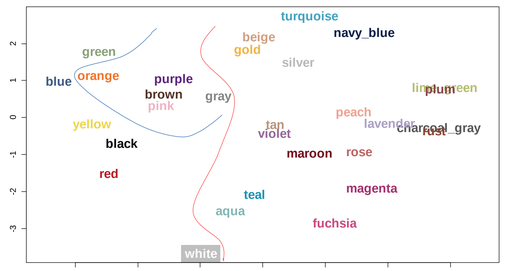


- frequency-based corpus measurements (type frequency, token frequency, frequency in compounds) group with experimental measurements
- such corpus-based characteristics as the usage of a color term in the head position, derivational productivity (measured as ratio), usage in different product categories correlate with Berlin and Kay evolutionary sequence
- purely formal measurements of color term length show the highest within cluster correlation and form a distinct group

Hierarchical cluster analysis of salience parameters for 34 categories based on squared Kendall's *tau* rank correlation coefficients

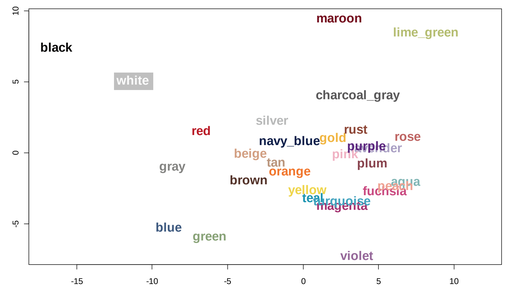
Patterns of salience

Kruskal's Non-metric Multidimensional Scaling, experimental measurements



stress: 9.3 %

Kruskal's Non-metric Multidimensional Scaling, corpus measurements

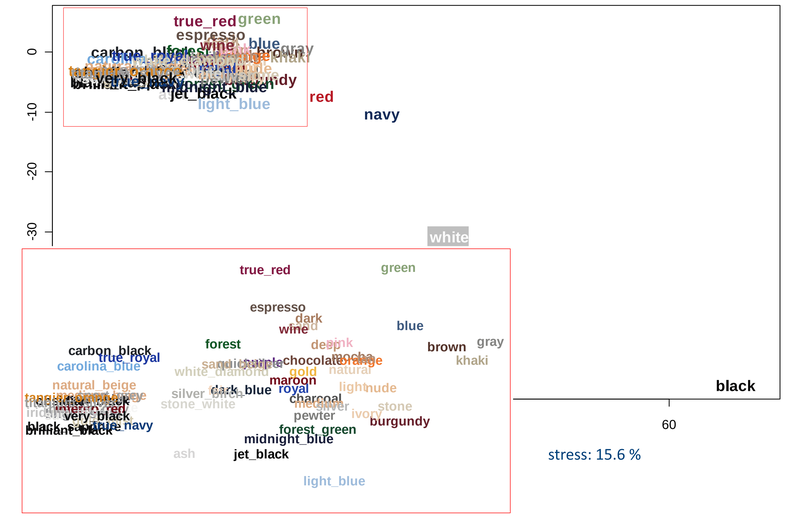


stress: 13.2 %

- experimental parameters highlight the distinction between basic and non-basic color terms
- corpus parameters partly reflect Berlin and Kay's hierarchy within primary BCT
- both experimental and corpus parameters distinguish between primary (*blue, black, red, white, green, yellow*) and secondary (*purple, brown, gray, pink*) basic color terms
- experimental parameters show stronger separation between secondary basic and non-basic color terms

IV. Salience beyond BCT

The second series of analyses explores the relative salience of 7700 color terms based on a sample of 16 400 observations representing four product categories.



stress: 15.6 %

Kruskal's Non-metric Multidimensional Scaling, frequency ≥ 20

Results

- salience of color terms is a continuous non-homogeneous parameter rather than a dichotomy between basic and non-basic color terms
- most of primary BCT (*black, blue, red, green*) are more distinct in their linguistic behavior compared to secondary BCT (*pink, orange, purple*) and especially to very densely clustered non-basic color terms
- certain non-basic color terms (*khaki, tan, stone*) come close in their linguistic characteristics to secondary BCT

V. Conclusions

- **color term usage in the "real world":**
by applying a bottom-up analysis to a specific context – online marketing materials – we propose an account of real-life color term usage in line with usage-based approach in Cognitive Linguistics
- **convergence of corpus-based and experimental measurements of salience:**
the analyses suggest both convergence of the two paradigms and a specific role of corpus-based measurements, which can be seen as evidence of multidimensional nature of linguistic salience and prototypicality effects
- **corpus-based measurements reveal a salience cline going beyond BCT:**
based on the corpus-based and formal parameters, we hypothesize an extended hierarchy of color terms

primary BCT *green, blue, red* → secondary BCT *orange, purple, pink* → non-basic monolexemic *wine, burgundy, ivory* → simple compounds *dark blue, forest green* → idiosyncratic compounds *deep sea blue, titanium silver*

- **granularity of the analyses and generalizations:**
the chosen granularity of the analyses specifically addresses the gap between the most salient BCT explored in categorization studies and idiosyncratic color terms hand-picked for the studies of color terms in advertising. This allows making generalizations on a larger scale than has been suggested in the previous research

Alena Anishchanka, alena.anishchanka@arts.kuleuven.be
Natalia Levshina, natalia.levshina@arts.kuleuven.be,
Univeristy of Leuven, Faculty of Arts,
Quantitative Lexicology and Variational Linguistics,
Blijde-Inkomstraat 21 (PO Box 3308) – 3000 Leuven - Belgium